ALONSO/FINN

FUNDAMENTAL UNIVERSITY PHYSICS

Second Edition

Volume II

Fields and Waves

The atomic masses, based on the exact number 12.00000 as the assigned atomic mass of the principal isotope of carbon, ^{12}C, are the most recent (1961) values adopted by the International Union of Pure and Applied Chemistry. The unit of mass used in this table is called *atomic mass*

Period	Series	Group→ I	II	III	IV
1	1	1 H 1.00797			
2	2	3 Li 6.939	4 Be 9.0122	5 B 10.811	6 C 12.01115
3	3	11 Na 22.9898	12 Mg 24.312	13 Al 26.9815	14 Si 28.086
4	4	19 K 39.102	20 Ca 40.08	21 Sc 44.956	22 Ti 47.90
	5	29 Cu 63.54	30 Zn 65.37	31 Ga 69.72	32 Ge 72.59
5	6	37 Rb 85.47	38 Sr 87.62	39 Y 88.905	40 Zr 91.22
	7	47 Ag 107.870	48 Cd 112.40	49 In 114.82	50 Sn 118.69
6	8	55 Cs 132.905	56 Ba 137.34	57–71 Lanthanide series*	72 Hf 178.49
	9	79 Au 196.967	80 Hg 200.59	81 Tl 204.37	82 Pb 207.19
7	10	87 Fr [223]	88 Ra [226.05]	89–Actinide series**	

* Lanthanide series:	57 La 138.91	58 Ce 140.12	59 Pr 140.907	60 Nd 144.24	61 Pm [147]	62 Sm 150.35
** Actinide series:	89 Ac [227]	90 Th 232.038	91 Pa [231]	92 U 238.03	93 Np [237]	94 Pu [242]

Table A–2 Fundamental Constants

Constant	Symbol	Value
Velocity of light	c	2.9979×10^{8} m s^{-1}
Elementary charge	e	1.6021×10^{-19} C
Electron rest mass	m_e	9.1091×10^{-31} kg
Proton rest mass	m_p	1.6725×10^{-27} kg
Neutron rest mass	m_n	1.6748×10^{-27} kg
Planck constant	h	6.6256×10^{-34} J s
	$\hbar = h/2\pi$	1.0545×10^{-34} J s
Charge-to-mass ratio for electron	e/m_e	1.7588×10^{11} kg^{-1} C
Quantum charge ratio	h/e	4.1356×10^{-15} J s C^{-1}
Bohr radius	a_0	5.2917×10^{-11} m
Compton wavelength:		
of electron	$\lambda_{C,e}$	2.4262×10^{-12} m
of proton	$\lambda_{C,p}$	1.3214×10^{-15} m
Rydberg constant	R	1.0974×10^{7} m^{-1}

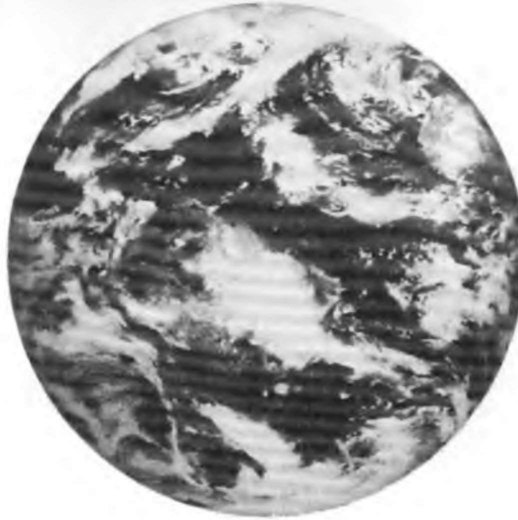
unit (amu): 1 amu = 1.6604×10^{-27} kg. The atomic mass of carbon is 12.01115 on this scale because it is the average of the different isotopes naturally present in carbon. (For artificially produced elements, the approximate atomic mass of the most stable isotope is given in brackets.)

V	VI	VII	VIII			0
						2 He 4.0026
7 N 14.0067	8 O 15.9994	9 F 18.9984				10 Ne 20.183
15 P 30.9738	16 S 32.064	17 Cl 35.453				18 A 39.948
23 V 50.942	24 Cr 51.996	25 Mn 54.9380	26 Fe 55.847	27 Co 58.9332	28 Ni 58.71	
33 As 74.9216	34 Se 78.96	35 Br 79.909				36 Kr 83.80
41 Nb 92.906	42 Mo 95.94	43 Tc [99]	44 Ru 101.07	45 Rh 102.905	46 Pd 106.4	
51 Sb 121.75	52 Te 127.60	53 I 126.9044				54 Xe 131.30
73 Ta 180.948	74 W 183.85	75 Re 186.2	76 Os 190.2	77 Ir 192.2	78 Pt 195.09	
83 Bi 208.980	84 Po [210]	85 At [210]				86 Rn [222]

63 Eu 151.96	64 Gd 157.25	65 Tb 158.924	66 Dy 162.50	67 Ho 164.930	68 Er 167.26	69 Tm 168.934	70 Yb 173.04	71 Lu 174.97
95 Am [243]	96 Cm [245]	97 Bk [249]	98 Cf [249]	99 Es [253]	100 Fm [255]	101 Md [256]	102 No	103

Constant	Symbol	Value
Bohr magneton	μ_B	9.2732×10^{-24} J T^{-1}
Avogadro constant	N_A	6.0225×10^{23} mol^{-1}
Boltzmann constant	k	1.3805×10^{-23} J °K^{-1}
Gas constant	R	8.3143 J °K^{-1} mol^{-1}
Ideal gas normal volume (STP)	V_0	2.2414×10^{-2} m^3 mol^{-1}
Faraday constant	F	9.6487×10^4 C mol^{-1}
Coulomb constant	K_e	8.9874×10^9 N m^2 C^{-2}
Vacuum permittivity	ϵ_0	8.8544×10^{-12} N^{-1} m^{-2} C^2
Magnetic constant	K_m	1.0000×10^{-7} m kg C^{-2}
Vacuum permeability	μ_0	1.3566×10^{-6} m kg C^{-2}
Gravitational constant	γ	6.670×10^{-11} N m^2 kg^{-2}
Acceleration of gravity at sea level and at equator	g	9.7805 m s^{-2}

Numerical constants: $\pi = 3.1416$; $e = 2.7183$; $\sqrt{2} = 1.4142$; $\sqrt{3} = 1.7320$



INTERNATIONAL ADVISORY BOARD
(for the Second Edition)

FUNDAMENTAL UNIVERSITY PHYSICS 2nd Edition

VOLUME TWO
FIELDS AND WAVES

MARCELO ALONSO
Florida Institute of Technology
Melbourne, Florida

Former Director
Department of Scientific Affairs
Organization of American States
Washington, D. C.

EDWARD J. FINN
Department of Physics
Georgetown University

ADDISON-WESLEY PUBLISHING COMPANY

Reading, Massachusetts • Menlo Park, California
Don Mills, Ontario • Wokingham, England • Amsterdam • Bonn • Sydney
Singapore • Tokyo • Madrid • Bogota • Santiago • San Juan

Library of Congress Cataloging in Publication Data
(Revised for Vol. 2)

Alonso, Marcelo.
Fundamental University Physics.

CONTENTS: v. 1. Mechanics and Thermodynamics. –v. 2. Fields and Waves. –[etc.]
Includes index.
1. Physics. I. Finn, Edward J., joint author.
II. Title.
QC21.A4 1980 530 79-51802
ISBN 0-201-00162-4 Paperback edition
ISBN 0-201-00077-6 Hardbound edition

FGHIJKLMN-MA-99876543210

PREFACE TO THE SECOND EDITION

Physics is a fundamental science that has a profound influence on all other sciences. Since this is so, not only physics majors and engineering students, but anyone who plans a career in science (including students majoring in biology, chemistry and mathematics), must have a thorough understanding of the fundamental ideas of physics.

The primary purpose of a general physics course (and perhaps the only reason it is allowed a place in the curriculum) is to give the student a unified view of physics. This should be done without bringing in too many details. A unified view of physics is attained by analyzing the basic principles, developing their implications, and discussing their limitations. The student will learn specific applications of the basic principles in the more specialized courses that follow. Consequently, this text presents what we believe are the fundamental ideas that constitute the core of today's physics. We have given careful consideration to the recommendations and suggestions of previous users of the text and the International Advisory Board of Editors in selecting the subject matter and the order and method of its presentation.

In many courses physics is taught as if it were a conglomeration of several sciences, more or less related, but without any real unifying view. The traditional division of physics into (the "science" of) mechanics, heat and kinetic theory, sound, optics, electricity and magnetism, and modern physics no longer has any justification. We have departed from this traditional approach. Instead, we follow a logical and unified presentation, emphasizing the conservation principles, the concepts of fields and waves, and the atomic view of matter. The special theory of relativity is used extensively throughout the text as one of the guiding principles that must be met by any physical theory. Many ideas of quantum physics are introduced rather early.

For convenience, the text appears in three volumes and the subject matter has been divided into five parts: (1) Mechanics, (2) Interactions and Fields, (3) Waves, (4) Quantum Physics, (5) Statistical Physics.

In Volume I we present mechanics in order to establish the fundamental principles needed to describe the motions we observe around us. Included in this volume, in order to adapt to the requirements of many schools, we have incorporated an elementary introduction to thermodynamics and statistical mechanics.

All phenomena in nature are the result of interactions, and interactions are analyzed in terms of fields. Part 2, in Volume II, considers not only those kinds of interactions we understand best (the gravitational and electromagnetic interactions, which are the interactions responsible for most of the macroscopic phenomena we observe), but also includes a discussion of the nuclear interaction. For the sake of convenience, the discussion of the gravitational interaction has been placed in Volume I; in Volume II we discuss electromagnetism in considerable detail, concluding with the formulation of Maxwell's equations. Part 3, which deals with wave phenomena as a consequence of the field concept, is also included in Volume II. It is here that we have included much of the material usually covered under the headings of acoustics and optics. The emphasis, however, has been placed on electromagnetic waves as a natural extension of Maxwell's equations. Part 3 and Volume II conclude with a discussion of Matter Waves as an introduction to the mathematical formulation of quantum mechanics. Thus, Volumes I and II cover the usual material in most introductory general physics courses.

Volume III includes the final two Parts of the text. In Part 4 we analyze the structure of matter—that is, atoms, molecules, nuclei and fundamental particles—an analysis preceded by the necessary background in quantum mechanics. This part constitutes an elementary introduction to the quantum theory of matter. Finally, in Part 5 we talk about the properties of matter in bulk. The principles of statistical mechanics are first presented and then applied to some simple, but fundamental, cases. We discuss thermodynamics from the point of view of statistical mechanics. Part 5 concludes with a study of the thermal properties of matter which explains how the principles of statistical mechanics and thermodynamics may be applied. Therefore, Volume III covers the subject matter included in most introductory Modern Physics courses, with the advantage that it constitutes a logical extension of Volumes I and II.

This text differs from standard university-level physics texts not only in its approach, but also in its content. We have included a number of fundamental topics not found in many books and we have deleted other topics that are traditional. The level of mathematics used in the text assumes that the student has had a minimal introduction to calculus and is currently enrolled in the introductory course of that subject. Also, it is highly desirable that the student have had a physics course in high school. Many applications of fundamental principles, as well as a few more advanced topics, appear in the form of worked-out examples; these may be discussed at the instructor's convenience or proposed to individual students on a selective basis. The material in the examples thus allows for flexibility in designing the course in accordance with both the wishes of the instructor and the background of the students. The problems at the end of each chapter are divided into two groups: basic problems and challenging problems. The basic problems are designed to drill the student and assist him in mastering the matter. The majority of these problems should be solved without too much effort. The challenging problems, on the other hand, should serve to stimulate the student, testing his understanding and initiative. A number of the challenging problems have been taken from the free-response section of the Advanced Placement Physics Examination with the permission of the College Entrance Examination Board and the Educational Testing Service. These are identified at the end of the problem; e.g., (AP-B; 1975) identifies a problem from the 1975 B (non-calculus) Exam, while (AP-C; 1975) is a problem from the calculus-based examination of the same year.

Universities have been under great pressure to incorporate into the curricula for all sciences new subjects that are more relevant than the traditional topics. We expect that this text will relieve some of this pressure by raising the students' level of understanding of physical concepts and increasing their ability to manipulate the corresponding mathematical relations. This will permit an upgrading of intermediate courses presently offered in the undergraduate curriculum, from which the traditional courses in mechanics, electromagnetism and modern physics will benefit most. Thus the physics student will finish undergraduate education at a higher level of knowledge than formerly possible—an important benefit for those who terminate their formal training at this point. Also, there will now be room for newer (and perhaps more exciting) courses at the graduate level. It is gratifying to encounter this same trend of upgrading in the more recent basic textbooks in other sciences.

The text is designed for a three-semester or four-quarter general physics course. It may also be used in those curricula in which the general physics course, using Volume I and II, is followed by a one- or two-semester course in modern physics, which would use Volume III. In either case, the sequence would offer a unified presentation to the student.

We hope that this text will be of assistance to those progressive physics instructors who are constantly struggling to improve the courses they teach. We also earnestly hope that it will stimulate the many students who deserve a presentation of physics that is more mature than that of most traditional courses.

We want to express our gratitude to all those whose assistance has made the completion of this work and its revision possible. We recognize our distinguished colleagues, Professors D. Lazarus and H. S. Robertson, who read the original manuscript, and Dr. R. G. Hughes, who solved all the problems. We also wish to express our deep appreciation to the many users throughout the world of the first edition of this text. Their helpful comments, which reached us in any one of the ten languages in which that edition has been published, were responsible for a number of corrections and revisions. In particular, the encouragement and suggestions offered by the International Advisory Board of Editors, whose membership is listed opposite the title page, have greatly assisted us in improving the clarity of presentation. Their help has been invaluable. However, we remain solely responsible for the deficiencies in the text. We are also grateful for the ability and dedication of the staff of the International Division of Addison-Wesley. Last, but certainly not least, we sincerely thank our wives, who have been so patient with us.

Washington, D. C.
March, 1979

M. A.
E. J. F.

NOTE TO THE STUDENT

This is a book about the fundamentals of physics, written for students majoring in science or engineering. The concepts and ideas you learn from it will, in all probability, become part of your professional life and your way of thinking. The better you understand them, the easier the rest of your undergraduate and graduate education will be.

The course in physics that you are about to begin is naturally more advanced than your high-school physics course. You must be prepared to tackle numerous difficult puzzles. To grasp the laws and techniques of physics may be, at times, a slow and painful process. Before you enter those regions of physics that appeal to your imagination, you must master other, less appealing, but very fundamental ones, without which you cannot use or understand physics properly.

You should keep two main objectives before you while taking this course. First: become thoroughly familiar with the handful of basic laws and principles that constitute the core of physics. Second: develop the ability to manipulate these ideas and apply them to concrete situations; in other words, to think and act as a physicist. You can achieve the first objective mainly by reading and re-reading those sections in large print in the text. To help you attain the second objective, there are many worked-out examples, in small print, throughout the text, and there are the homework problems at the end of each chapter. We strongly recommend that you first read the main text and, once you are acquainted with it, proceed with those examples and problems assigned by the instructor. The examples either illustrate an application of the theory to a concrete situation, or extend the theory by considering new aspects of the problem discussed. Sometimes they provide some justification for the theory.

The problems at the end of each chapter vary in degree of difficulty and have been broken into two categories: basic problems and challenging problems. The basic problems are mostly of the type that should be solvable after reading the text material; they are made available so that you may apply what you have read to a given particular situation. The challenging problems, on the other hand, should force you to perform a series of steps before the answer can be obtained; in other words, you may be required to return to material previously introduced in order to work a problem. Those challenging problems followed by, for example, (AP-B; 1970) or (AP-C; 1970) are taken from the free-response section of the Advanced Placement Physics Examination; the B-exams are non-calculus based and the C-exams are calculus-based; the year stated is the year the question appeared on the given test. In general, it is a good idea to try to solve a problem in a symbolic or algebraic form first, and insert numerical values only at the end. If you cannot solve an assigned problem in a reasonable time, lay the problem aside and make a second attempt later. For those few problems that refuse to yield a solution, you should seek help. One source of self-help that will teach you the *method* of problem-solving is the book *How to Solve It* (second edition), by G. Polya (Garden City, N. Y.: Doubleday, 1957).

Physics is a quantitative science that requires mathematics for the expression of its ideas. All the mathematics used in this book can be found in a standard calculus text, and you should consult such a text whenever you do not understand a mathematical derivation. But by no means should you feel discouraged by a mathematical difficulty; in case of mathematical trouble, consult your instructor or a more advanced student. For the physical scientist and engineer, mathematics is a tool, and is second in importance to understanding the physical ideas. For your convenience, some of the most useful mathematical relations are listed in an appendix at the end of the book.

All physical calculations must be carried out using a consistent set of units. In this book the SI system is used. You may find it unfamiliar at first; however, it requires very little effort to become acquainted with it. Also, it is the system that is used in all major government laboratories throughout the world and is becoming standard in all the major scientific publications. It is a good idea to use a mechanical or electronic slide rule from the start; the accuracy of these instruments and their ability to hold intermediate results will save you many hours of computation. Mechanical slide rules, even the simplest, have three-place accuracy and this is almost always sufficient for problems in this text. The electronic slide rule/calculator has considerably greater accuracy and appears to be the indispensable tool for the scientist of the future.

The text does not stress the historical aspects of physics. For those students interested in the evolution of ideas in physics in a historical context, there are a number of informative texts available. In particular, we would recommend the fine book by Holton and Roller, *Foundations of Modern Physical Science*, second edition, (Reading, Mass.: Addison-Wesley, 1973).

CONTENTS

PART TWO

ELECTRO-MAGNETIC FIELDS

Once the general rules governing motion have been grasped, the next step in developing an understanding of physics is to investigate the interactions responsible for such motions. There are several kinds of interactions.

The *gravitational interaction* manifests itself in planetary motion and in the motion of matter in bulk. Gravitation, although the weakest of all known interactions, was the first interaction to be studied carefully because agricultural and other forecasting purposes provoked an early interest in astronomy and because many phenomena caused by gravitation affect people's lives directly.

The *electromagnetic interaction* is the best-understood interaction and perhaps the most important for daily life. Most of the phenomena observed every day, including chemical and biological processes, are the result of electromagnetic interactions between atoms and molecules.

The *strong*, or *nuclear*, *interaction* is responsible for holding protons and neutrons (known as nucleons) within the atomic nucleus, and for other related phenomena. In spite of intensive research, knowledge of this interaction is still incomplete.

The *weak interaction* is responsible for certain processes, such as beta decay, among the fundamental particles. Our understanding of this interaction also is still very meager.

The relative strengths of these above interactions, measured against the strong interaction as 1, are electromagnetic $\sim 10^{-2}$, weak $\sim 10^{-5}$, and gravitational $\sim 10^{-38}$. Among the as-yet-unsolved problems of physics are why there appear to be only four interactions and why there is such a wide difference in their strengths. Alternatively one might ask why there is not *only* one interaction that, in various limits, gives the appearance of the four interactions experimentally identified.

It is interesting to see what Isaac Newton said about interactions 200 years ago:

> Have not the small Particles of Bodies certain Powers, or Forces, by which they act . . . upon one another for producing a great Part of the Phenomena of Nature? For it's well known, that Bodies act one upon another by the Attractions of Gravity, Magnetism, and Electricity . . . and make it not improbable but that there may be more attractive Powers than these . . . . How these attractions may be performed, I do not here consider . . . . The Attractions of Gravity, Magnetism, and Electricity, reach to very sensible distances . . . and there may be others which reach to so small distances as hitherto escape observation . . . . (*Opticks*, Book III, Query 31)

To describe these interactions, the concept of a field is introduced. By *field* we mean a physical entity that extends over a region of space and is described by a function of both position and time. Introducing this concept to describe the interaction between two particles is appropriate since the interaction between them depends on their relative positions and motions. Accordingly, for each interaction a particle is assumed to produce around it a corresponding field. This field, in turn, acts on a second particle to produce the required interaction. The second particle, of course, has its own field, which acts on the first particle and results in a mutual interaction.

The electromagnetic interaction is customarily described in terms of two fields: the electric field and the magnetic field. However, it should be emphasized that these two fields are not independent entities but rather are intimately related to each other, and the separation of the electromagnetic field into these two components is dictated by the relative motion of the electric charges and the observer.

On the other hand, the concept of a field is not used exclusively to describe interactions but also to describe other physical phenomena. For example a meteorologist may express both the atmospheric pressure and the temperature as functions of the latitude and the longitude on the earth's surface as well as the height above it.

We have come to recognize that the key features of a field, in order that the field properly describe an interaction between particles, are that the field, itself, must possess energy and momentum and that the field must be able to transport both of these properties from one particle to another.

Gravitational interaction and the gravitational field were discussed in Chapter 13 of Volume I. In this volume the electromagnetic interaction will be considered in depth. The remaining two interactions, the weak and the strong (nuclear) interactions, will be discussed descriptively in this volume; a detailed investigation into their properties is reserved for Volume III.

CHAPTER ONE

ELECTRIC INTERACTION

1.1 Introduction

Consider a very simple experiment: Run a comb through a person's hair on a very dry day: when the comb is brought close to tiny pieces of paper. the paper scraps are swiftly attracted by the comb. Similar phenomena occur if a glass rod is rubbed with a silk cloth or an amber rod is rubbed with a piece of fur. We describe these experiments by saying that as a result of rubbing. materials may acquire a new property, which is called *electricity* (from the Greek word *elektron*, meaning amber), and that this electrical property gives rise to an interaction much stronger than gravitation.

Several fundamental differences exist between electrical and gravitational interactions. First there is only one kind of gravitational interaction, resulting in a universal attraction between any two masses. However, experiment shows that there are two kinds of electrical interactions. When an electrified glass rod is placed near a small cork ball hanging from a string. the rod attracts the ball (Fig. 1-1a). If the experiment is repeated with an electrified amber rod. the same attractive effect on the ball is observed (Fig. 1-1b). However. if both rods approach the ball simultaneously as shown in Fig. 1-1(c), instead of a larger attraction, a smaller attraction on the ball or no attraction at all is observed. These simple experiments indicate that although both the electrified glass rod and the amber rod attract the cork ball, they do it by opposite physical processes. When both rods are present. they counteract each other to produce a smaller or even null effect. Therefore there are two kinds of electrified states; one glasslike and the other amberlike. The first is called *positive* and the other *negative*.

Suppose now that two cork balls are touched by an electrified glass rod. It may be assumed that the two balls also become positively electrified. When the balls are brought together. they *repel* each other (Fig. 1-2a). The same result occurs after the balls are touched by an electrified amber rod and acquire negative electrification (Fig. 1-2b). However if one ball is touched by the glass rod and the other by the

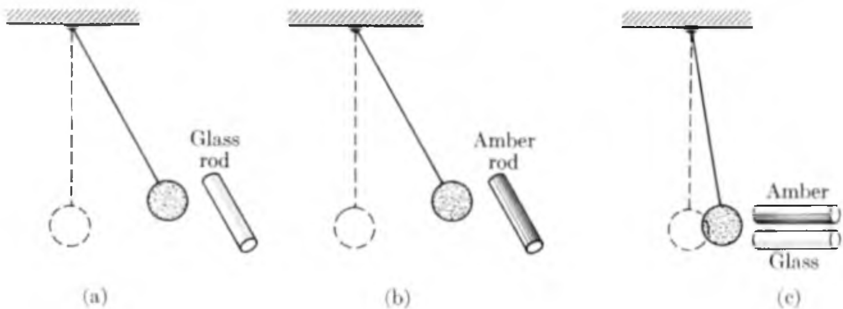


Fig. 1-1. Experiments with electrified glass and amber rods.

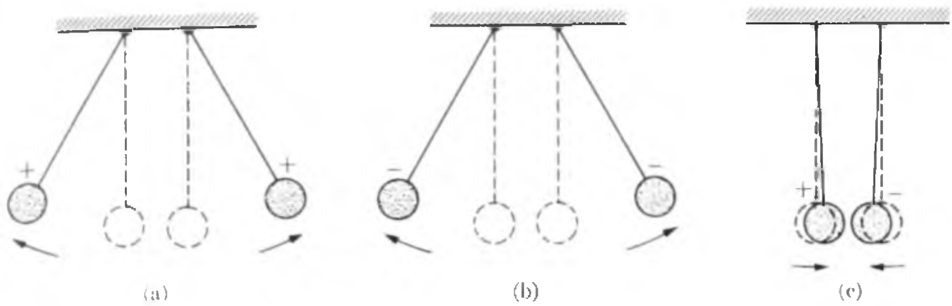


Fig. 1-2. Electric interactions between like and unlike charges.

amber rod, the balls attract each other (Fig. 1-2c). Thus, although the gravitational interaction is always attractive, the electrical interaction may be either attractive or repulsive.

> *Two bodies with the same kind of electrification (either positive or negative) repel each other, but if they have different kinds of electrification (one positive and the other negative), they attract each other.*

This statement is indicated schematically in Fig. 1-3. Had the electrical interaction been only attractive or only repulsive, the existence of gravitation might never have been noticed because the electrical interaction is much stronger. However since most bodies seem to be composed of equal amounts of positive and negative electricity, the net electrical interaction between any two macroscopic bodies is very small or zero. Thus as a result of a cumulative mass effect, the dominant macroscopic interaction appears to be the much weaker gravitational interaction.

1.2 Electric Charge

In the same way that the strength of the gravitational interaction is characterized by attaching to each body a gravitational mass, the state of electrification of a body is characterized by defining an *electric mass*, more commonly called *electric charge* (or simply charge) and represented by the symbol q. Thus any piece of matter or any particle is characterized by two independent but fundamental properties: mass and charge.

Since there are two kinds of electrification, there are also two kinds of electric charge: positive and negative. A body exhibiting positive electrification has a positive

Fig. 1-3. Forces between like and unlike charges.

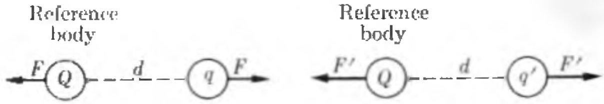


Fig. 1-4. Comparison of electric charges q and q', and their electric interactions with a third charge Q.

electric charge, and a body with negative electrification has a negative electric charge. The net charge of a body is the algebraic sum of its positive and negative charges. A body having equal amounts of positive and negative charges (i.e., zero net charge) is called electrically *neutral*. On the other hand, a particle having a non-zero net charge is often called an *ion*. Since it does not exhibit gross electrical forces, matter in bulk is assumed to be composed of equal amounts of positive and negative charges.

For an operational definition of the charge of an electrified body, we adopt the following procedure. Choose an arbitrary charged body Q (Fig. 1-4) and, place the charge q at a distance d from Q. Measure the force F exerted on q. Next, place another charge q' at the same distance d from Q and measure the force F'. Then define the values of the charges q and q' proportional to the forces F and F'. That is,

$$\frac{q}{q'}=\frac{F}{F'}. \tag{1.1}$$

If a value of unity is arbitrarily assigned to the charge q', the value of q can be obtained. This method of comparing charges is very similar to the one for comparing the masses of two bodies. This definition of charge implies that, all geometrical factors being equal, the force of electrical interaction is proportional to the charges of the particles.

It has been found that in all physical processes thus far observed in nature, the net charge of an isolated system remains constant. In other words

> *the net or total charge does not change for any process occurring within an isolated system.*

No exception has been found to this rule, known as the *principle of conservation of charge*. We shall have occasion to discuss this rule later when we deal with processes involving fundamental particles.

1.3 Coulomb's Law

Consider the electric interaction between two charged particles *at rest* in the observer's inertial frame of reference or, at most, in motion with a very small velocity; the results of such an interaction constitute *electrostatics*. The electrostatic interaction for two charged particles is given by *Coulomb's law*, named after the French engineer Charles A. de Coulomb (1736–1806), who was the first to state this law in the following manner:

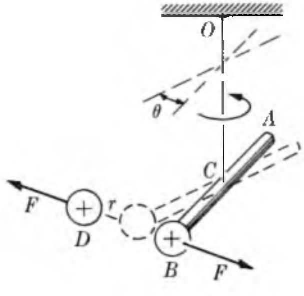


Fig. 1-5. Cavendish torsion balance for verifying the law of electric interaction between two charges.

The electrostatic interaction between two charged particles is proportional to their charges and to the inverse of the square of the distance between them, and its direction is along the line joining the two charges.

This law may be expressed mathematically by

$$F=K_e\frac{qq'}{r^2} \tag{1.2}$$

where r is the distance between the two charges q and q', F is the force acting on either charge, and K_e is a constant to be determined by our choice of units. This law is very similar to the law for gravitational interaction. Thus many mathematical results proved in Chapter 13 of Volume I are applicable here simply by replacing $\gamma mm'$ by $K_e qq'$.

We can experimentally verify the inverse-square law (1.2) by measuring the force between two given charges placed at several distances. A possible experimental arrangement, indicated in Fig. 1-5, is similar to the Cavendish torsion balance used to verify the law of gravitation. The rod AB with a charged sphere B at the end is suspended from the fiber OC. Then another charged body D is brought near. As a result of the forces between spheres D and B, the rod AB experiences a torque and twists the fiber OC. The force between the charge at B and the charge at D is found by measuring the angle θ through which the fiber OC is rotated to restore equilibrium.

The constant K_e in Eq. (1.2) is similar to the constant γ in the law of gravitation except that in the gravitational case the units of mass, distance, and force were already defined and the value of γ was determined experimentally. In the present case although the units of force and distance have already been defined, the unit of charge is as yet undefined. If a definite statement about the unit of charge is made, then the value of K_e may be determined experimentally. Alternatively if the value of K_e is given, Eq. (1.2) may be used to define the unit of charge. In SI units the value of K_e is assigned the numerical value of $10^{-7}c^2=8.9874\times10^9$ where c is the velocity of light in vacuum.* For practical purposes, we may say that K_e is equal to 9×10^9.

*The choice of a particular value for K_e will be explained in Section 5.7.

Then when the distance is measured in meters and the force in newtons, Eq. (1.2) becomes

$$F = 9 \times 10^9 \frac{qq'}{r^2}. \tag{1.3}$$

The unit of charge is called a *coulomb* and, by application of Eq. (1.3), is defined in terms of the force experienced: *when placed one meter from an equal charge in vacuum, the coulomb of charge experiences a repulsive force of* 8.9874×10^9 *newtons*. Formula (1.3) holds only for two charged particles at rest relative to the observer and in vacuum; that is, for two charged particles in the absence of any other charge or matter (see Section 2.6). According to Eq. (1.2), K_e is expressed in N m^2 C^{-2} or m^3 kg s^{-2} C^{-2}. For practical and computational reasons, it is more convenient to express K_e in the form

$$K_e = \frac{1}{4\pi\epsilon_0}, \tag{1.4}$$

where the new physical constant ϵ_0 is called the *vacuum permittivity*. According to the value assigned to K_e, ϵ_0 has the value

$$\epsilon_0 = \frac{10^7}{4\pi c^2} = 8.854 \times 10^{-12}\ \mathrm{N^{-1}\ m^{-2}\ C^2}. \tag{1.5}$$

Accordingly, Eq. (1.3) is normally written in the form

$$F = \frac{qq'}{4\pi\epsilon_0 r^2}. \tag{1.6}$$

The signs of the charges q and q' must be included when Eq. (1.6) is used. A negative value of F corresponds to attraction and a positive value corresponds to repulsion.

Example 1.1. Given the charge arrangement of Fig. 1-6, in which $q_1 = +1.5 \times 10^{-3}$ C, $q_2 = -0.5 \times 10^{-3}$ C, $q_3 = +0.2 \times 10^{-3}$ C, and $AC = r_1 = 1.2$ m, $BC = r_2 = 0.5$ m, find the resultant force on charge q_3.

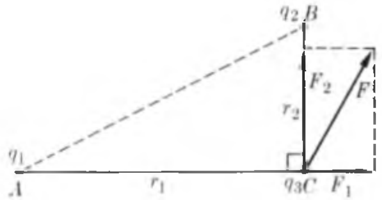


Fig. 1-6. Resultant electric force produced by q_1 and q_2 on q_3.

▼ The force F_1 between q_1 and q_3 is repulsive while the force F_2 between q_2 and q_3 is attractive. From Eq. (1.6) their values are

$$F_1=\frac{q_1q_3}{4\pi\epsilon_0 r_1^2}=1.875\times10^3\ \mathrm{N}. \qquad F_2=\frac{q_2q_3}{4\pi\epsilon_0 r_2^2}=-3.6\times10^3\ \mathrm{N}.$$

Therefore the magnitude of the resultant force is

$$F=\sqrt{F_1^2+F_2^2}=4.06\times10^3\ \mathrm{N}. \ ▲$$

1.4 Electric Field

Any region in space in which an electric charge experiences a force is called an *electric field*. The force is due to the presence of other charges in that region. For example, a charge q placed in a region where there are other charges q_1, q_2, q_3, etc. (Fig. 1-7) experiences a force $\boldsymbol{F}=\boldsymbol{F}_1+\boldsymbol{F}_2+\boldsymbol{F}_3+\cdots$, and is in an electric field produced by the charges $q_1, q_2, q_3, \ldots$. (The charge q of course also exerts forces on $q_1, q_2, q_3, \ldots$, but we ignore these forces for the present.) Since the force that each charge $q_1, q_2, q_3, \ldots$ produces on the charge q is proportional to q, the resultant force $\boldsymbol{F}$ is also proportional to q; that is, the force on a particle placed in an electric field is proportional to the charge of the particle. Therefore to determine whether an electric field is present in a certain region, a small test charge must be brought into the region and the force experienced by the test charge must be analyzed.

By definition the *intensity of the electric field*, $\boldsymbol{\mathscr{E}}$, at a point equals the force per unit charge experienced by the test charge placed at that point. Thus

$$\boldsymbol{\mathscr{E}}=\frac{\boldsymbol{F}}{q} \qquad \text{or} \qquad \boldsymbol{F}=q\boldsymbol{\mathscr{E}}. \tag{1.7}$$

The electric field intensity $\mathscr{E}$ is expressed in newtons/coulomb or N C^{-1} or m kg s^{-2} C^{-1}, in fundamental units.

Note that in view of the definition (1.7) if q is positive, the force $\boldsymbol{F}$ acting on the charge has the same direction as the field $\boldsymbol{\mathscr{E}}$; but if q is negative, the force $\boldsymbol{F}$ has the direction opposite to $\boldsymbol{\mathscr{E}}$ (Fig. 1-8). Therefore if there is an electric field in a region

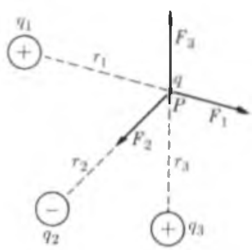


Fig. 1-7. The electric forces acting on a positive charge at P. The resultant force on q is the vector sum of all the forces.

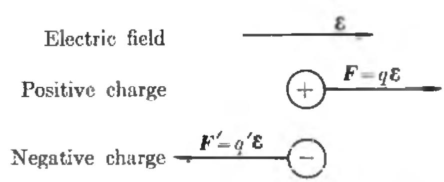
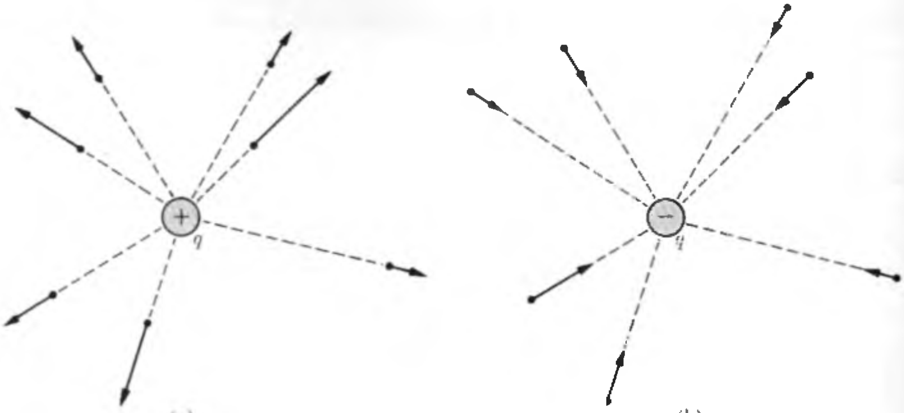


Fig. 1-8. Direction of the force produced by an electric field on a positive and a negative charge.



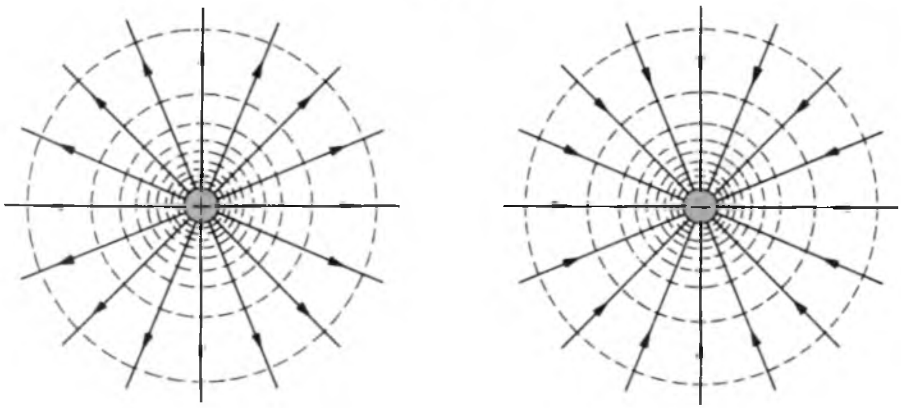
(a) (b)

Fig. 1-9. Electric field produced by a positive and a negative charge.

where positive and negative particles or ions are present, the field will tend to move the positively and negatively charged bodies in opposite directions, and produce a separation of charge. This effect is sometimes called *polarization*.

Writing Eq. (1.6) in the form $F = q'(q/4\pi\epsilon_0 r^2)$ gives the force produced by the charge q on the charge q' placed a distance r from q. From Eq. (1.7) we may also say that the electric field $\mathscr{E}$ at the point where q' is placed is such that $F = q'\mathscr{E}$. Therefore comparing both expressions of F, we conclude that the electric field at a distance r from a point charge q is $\mathscr{E} = q/4\pi\epsilon_0 r^2$, or in vector form

$$\mathscr{E} = \frac{q}{4\pi\epsilon_0 r^2}\boldsymbol{u}_r \tag{1.8}$$



(a) (b)

Fig. 1-10. Lines of force (solid lines) and equipotential surfaces (dotted lines) of the electric field of a positive and a negative charge.

where $\boldsymbol{u}_r$ is the unit vector in the radial direction, away from the charge q. Expression (1.8) is valid for both positive and negative charges; the direction of $\mathscr{E}$ relative to $\boldsymbol{u}_r$ is given by the sign of q. Thus $\mathscr{E}$ is directed away from a positive charge and toward a negative charge. In the corresponding formula for the gravitational field, the negative sign was written explicitly because the gravitational interaction is always attractive. Figure 1-9(a) indicates the electric field at points near a positive charge, and Fig. 1-9(b) shows the electric field at points near a negative charge.

Just as in the case of a gravitational field, an electric field may be represented by lines (called *field lines* or *lines of force*) that at each point are tangent to the direction of the electric field at the point. The lines of force in Fig. 1-10(a) depict the electric field of a positive charge, and those in Fig. 1-10(b) show the electric field of a negative charge. In each case the lines of force are radial lines passing through the charge. These electric field lines are called lines of force because they define the direction in which a positive test charge q' would tend to move when placed at that point in the field.

When several charges are present as in Fig. 1-7, the resultant electric field at any point is the vector sum of the electric fields produced at the point by each charge. That is,

$$\mathscr{E} = \mathscr{E}_1 + \mathscr{E}_2 + \mathscr{E}_3 + \cdots = \sum_i \mathscr{E}_i.$$

Figure 1-11 shows the resultant electric field in the case of a positive and a negative

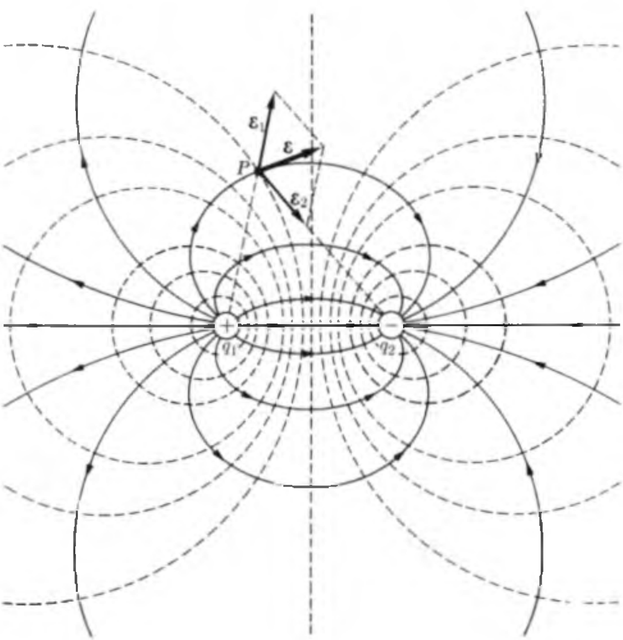


Fig. 1-11. Lines of force and equipotential surfaces of the electric field of two equal but opposite charges.

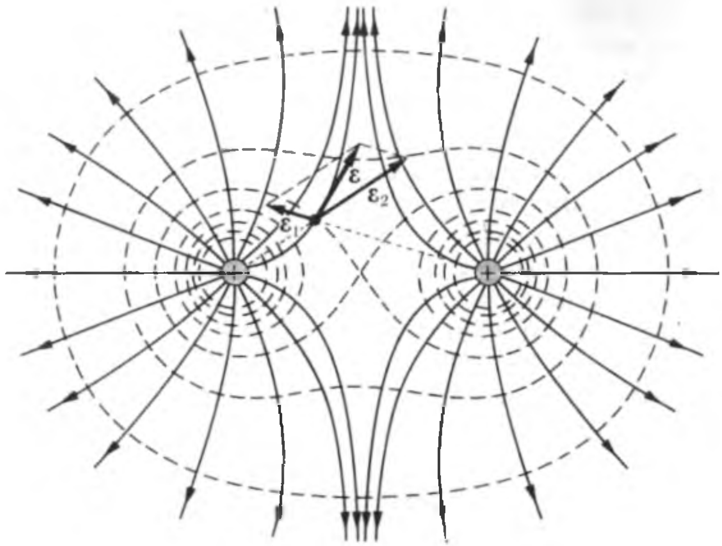


Fig. 1-12. Lines of force and equipotential surfaces of the electric field of two identical charges.

charge of the same magnitude, such as a proton and an electron in a hydrogen atom. Figure 1-12 shows the lines of force for two equal positive charges, such as the two protons in a hydrogen molecule. In both figures the lines of force of the resultant electric field produced by the two charges have also been represented.

A *uniform* electric field has the same intensity and direction everywhere and is represented by parallel lines of force (Fig. 1-13). The best way of producing a uniform electric field is by charging, with equal and opposite charges, two parallel metal plates. Symmetry indicates that the field is uniform, but this assertion is later verified in Section 2.3.

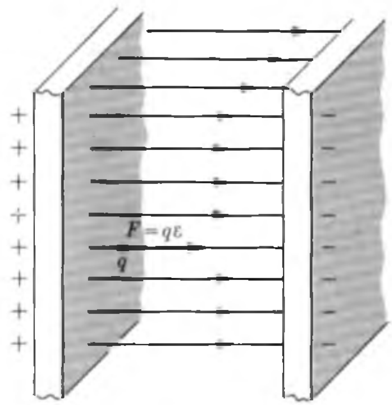


Fig. 1-13. Uniform electric field.

Fig. 1-14. Resultant electric field produced by q_1 and q_2 at C.

Example 1.2. Determine the electric field produced at C (in Fig. 1-6) by charges q_1 and q_2, which are defined in Example 1.1.

▼ There are two methods of solution. Because the solution to Example 1.1 gave the force F on charge q_3 at C, using Eq. (1.7) gives

$$\mathscr{E}=\frac{F}{q_3}=2.03\times10^7\ \mathrm{N\,C^{-1}}.$$

Another procedure is first to use Eq. (1.8) to compute the electric field produced at C (Fig. 1-14) by each of the charges. Equation (1.8) gives

$$\mathscr{E}_1=\frac{q_1}{4\pi\epsilon_0 r_1^2}=9.37\times10^6\ \mathrm{N\,C^{-1}},$$

and

$$\mathscr{E}_2=\frac{q_2}{4\pi\epsilon_0 r_2^2}=18.0\times10^6\ \mathrm{N\,C^{-1}}.$$

Since the electric fields are perpendicular, the magnitude of the resultant field is

$$\mathscr{E}=\sqrt{\mathscr{E}_1^2+\mathscr{E}_2^2}=2.03\times10^7\ \mathrm{NC^{-1}}.$$

The two results are identical. ▲

Example 1.3 The motion of an electric charge in a uniform field.

▼ The equation of motion of an electric charge in an electric field is found from the equation

$$\boldsymbol{F}=m\boldsymbol{a}=q\boldsymbol{\mathscr{E}} \qquad \text{or} \qquad \boldsymbol{a}=\frac{q}{m}\boldsymbol{\mathscr{E}}.$$

The acceleration of a body in an electric field depends therefore on the ratio q/m. Since this ratio is generally different for different charged particles or ions, their accelerations in the same electric field are also different. If the field $\boldsymbol{\mathscr{E}}$ is uniform, the acceleration $\boldsymbol{a}$ is constant and the path of the electric charge is a parabola just as in the case of projectile motion in a uniform gravitational field.

An interesting case is that of a charged particle passing through an electric field occupying a limited region in space (Fig. 1-15). For simplicity assume that the initial velocity v_0 of the particle when it enters the field is perpendicular to the direction of the electric field. The X-axis is placed parallel to the initial velocity of the particle and the Y-axis is placed parallel to the field. After

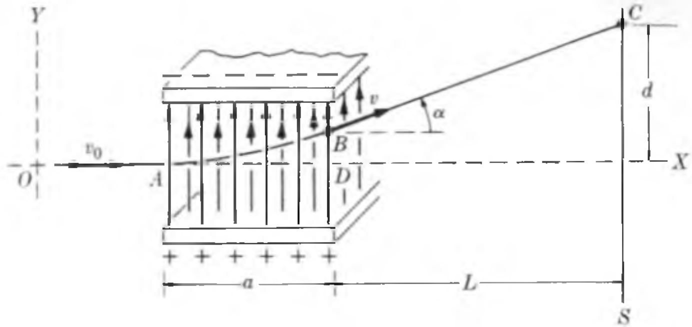


Fig. 1-15. Deflection of a positive charge by a uniform electric field.

crossing the field, the particle resumes rectilinear motion but with a different velocity v and in a different direction. Therefore, the electric field has produced a deflection, measured by the angle α.

While the particle is moving through the field with an acceleration $a_y=(q/m)\mathscr{E}$, the coordinates of the particle are given by

$$x=v_0t, \qquad y=\tfrac{1}{2}(q/m)\mathscr{E}t^2.$$

With the time t eliminated the equation of the path is

$$y=\frac{1}{2}\left(\frac{q}{m}\right)\left(\frac{\mathscr{E}}{v_0^2}\right)x^2,$$

verifying that the path is a parabola. The deflection α is found by calculating the slope dy/dx of the path at $x=a$. The result is

$$\tan\alpha=(dy/dx)_{x=a}=q\mathscr{E}a/mv_0^2.$$

If a screen S is placed at L, the particle with given q/m and velocity v_0 will reach a point C on the screen. Noting that $\tan\alpha$ is approximately equal to d/L because the vertical displacement BD is small compared with d when $L$is large, we have

$$\frac{q\mathscr{E}a}{mv_0^2}\cong\frac{d}{L}. \tag{1.9}$$

By measuring d, L, a, and $\mathscr{E}$, we may obtain the velocity v_0 (or the kinetic energy) given the ratio q/m; conversely, the ratio q/m may be found given v_0. Therefore when a stream of particles, all having the same ratio q/m, passes through the electric field, they are deflected by an amount inversely proportional to their entering kinetic energies.

A device such as the one illustrated in Fig. 1-15 may be used as an *energy analyzer*, separating identical charged particles moving with different energies. For example β-rays are electrons emitted by some radioactive materials; if a beta emitter is placed at O, all the electrons will concentrate at the same spot on the screen only if they have the same energy. If they are emitted with different energies, the electrons will be spread over a region of the screen. This second situation is found experimentally and has great importance from the point of view of explaining nuclear structure.

By using two sets of parallel charged plates, we can produce two mutually perpendicular fields, one horizontal along HH' and another vertical along VV' as shown in Fig. 1-16. By adjusting the relative intensity of the fields, we can obtain an arbitrary deflection of the electron beam to any

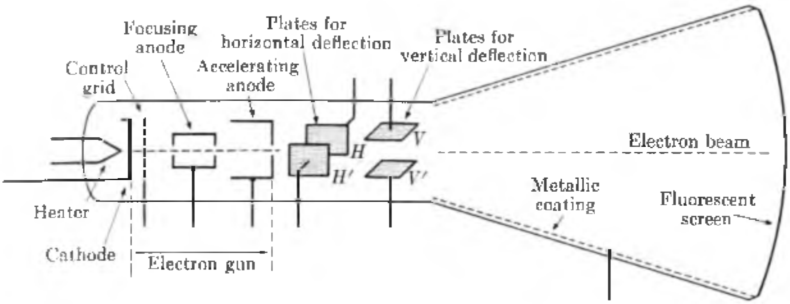


Fig. 1-16. Motion of a charge under crossed electric fields. Electrons are emitted from the cathode and accelerated by a large electric field. A hole in the accelerating anode allows the electrons to pass out of the electron gun and between the two sets of deflection plates. The metallic coating inside the tube keeps the right end free of electric fields by shielding the external sources and by conducting away the electrons of the beam.

spot on the screen. If either or both of the two fields are variable, the spot on the screen can be made to describe various curves or patterns. Practical applications of this effect occur in *television tubes* and in *oscilloscopes.* ▲

1.5 The Quantization of Electric Charge

Many experiments have been devised to resolve the question of whether the electric charge on a body is an integral multiple of a finite quantity or whether the charge may be subdivided continuously. The classical experiment to show that electric charge appears not just in any amount, but as a multiple of a fundamental unit, or *quantum*, is that of the American physicist Robert A. Millikan (1869–1953). For several years during the early part of this century he repeatedly performed what is now known as the *oil-drop experiment.* Millikan set up, between two parallel horizontal plates A and B (Fig. 1-17), a vertical electric field $\mathscr{E}$ that could be switched on and off.

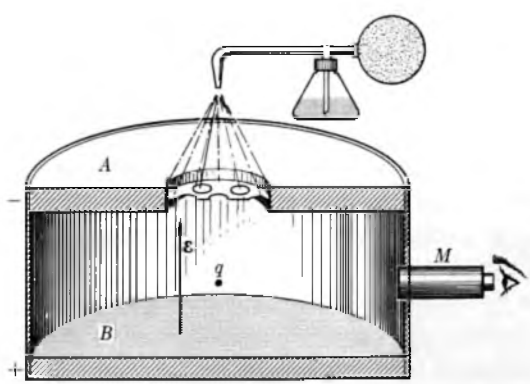


Fig. 1-17. Millikan oil-drop experiment. The motion of the charged oil drop q is observed through the microscope M.

At its center the upper plate had a few small perforations through which oil drops, produced by an atomizer, could pass. Most of the oil drops were charged by friction with the nozzle of the atomizer.

This experiment will be first analyzed from a theoretical standpoint. Call m the mass and r the radius of one oil drop. For this drop the equation of motion for free fall (i.e., with the electric field switched off) when its downward velocity is v is

$$ma = mg - 6\pi\eta r v$$

where the second term on the right is due to the viscous friction of air. [See Eqs. (7.19) and (7.20) in Volume I.] The terminal velocity v_1 of the drop, when $a=0$, is

$$v_1 = \frac{mg}{6\pi\eta r} = \frac{2\rho r^2 g}{9\eta} \tag{1.10}$$

where ρ represents the oil density and the relation $m = (\frac{4}{3}\pi r^3)\rho$ has been used. (To be precise, the buoyancy of the air must also be taken into account by writing $\rho - \rho_a$ instead of ρ where ρ_a is the air density.)

If the drop has a positive charge q, when the electric field is applied, the equation of motion of the oil drop in the upward direction when its velocity is v is

$$ma = q\mathscr{E} - mg - 6\pi\eta r v;$$

and its terminal velocity in the upward direction v_2, when $a=0$, is

$$v_2 = \frac{q\mathscr{E} - mg}{6\pi\eta r}.$$

Or solving for q and using Eq. (1.10) to eliminate mg gives the charge on the drop:

$$q = \frac{6\pi\eta r(v_1 + v_2)}{\mathscr{E}}. \tag{1.11}$$

The radius of the drop may be found by measuring v_1 and solving Eq. (1.10) for r. The charge q is obtained by measuring v_2 and applying Eq. (1.11). (If the charge is negative, the upward motion is produced by applying a downward electric field.)

A different procedure is followed in actual practice. The upward and downward motion of the drop is observed several times by successively switching the field on and off. The velocity v_1 remains the same; but the velocity v_2 occasionally changes, suggesting a change in the charge of the drop. These changes are due to the occasional ionization of the surrounding air by cosmic rays. While moving through the air, the drop may pick up some of these ions. Changes in charge can also be induced by placing near the plates a source of x- or γ-rays that increase the ionization of the air.

According to Eq. (1.11), the changes Δq and Δv_2 of charge and upward velocity are related by

$$\Delta q = \frac{6\pi\eta r}{\mathscr{E}}\Delta v_2. \tag{1.12}$$

Sometimes Δq is positive and at other times negative, depending on the nature of the charge modification. Repeating the oil-drop experiment many times with different

Table 1-1. Mass and Charge of the Electron, Proton and Neutron.

Particle	Mass	Charge
Electron	$m_e = 9.1091 \times 10^{-31}$ kg	$-e$
Proton	$m_p = 1.6725 \times 10^{-27}$ kg	$+e$
Neutron	$m_n = 1.6748 \times 10^{-27}$ kg	0

drops allows us to conclude that the changes Δq are always multiples of a fundamental charge e (that is, $\Delta q = ne$); the value of e is

$$e = 1.6021 \times 10^{-19} \text{ C}. \tag{1.13}$$

The quantity e is called the *elementary charge*. *All charges observed in nature are equivalents or multiples of the elementary charge e*; so far no exception has been found to this rule. It therefore seems to be a fundamental law of nature that electric charge is quantized. Until the present time, no one has found an explanation for this fact in terms of more fundamental concepts.

A second important aspect of electric charge is that the elementary charge is always associated with some fixed mass, an association that gives rise to what may be called a *fundamental particle*. In a later chapter (Section 4.4), some methods for measuring the ratio q/m will be discussed; then if q is known, m can be obtained. In this way several fundamental particles have been identified. For the present, we assume that the building blocks of all atoms are only three particles: the *electron*, the *proton*, and the *neutron*. Their characteristics are outlined in Table 1.1.

Note that the neutron carries no electric charge; however, the neutron does have other electrical properties, which will be discussed later. That the proton mass is about 1840 times larger than the electron mass has a profound influence in many physical phenomena.

At this point it may be noticed that the number of electrons or protons necessary to make up a negative or positive charge equal to one coulomb is $1/1.6021 \times 10^{-19} = 6.2418 \times 10^{18}$.

1.6 Electric Potential

A charged particle placed in an electric field has potential energy because the field exerts force on the charge, and therefore work must be done to bring the charge to a particular place in the field; that is, work is done by the electric field whenever a charge moves from one place to another. The *electric potential* at a point in an electric field is defined as the potential energy per unit charge placed at that point; therefore electric potential is a scalar quantity. Designating the electric potential at a particular

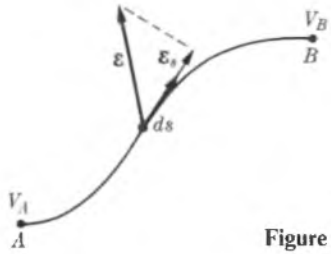


Figure 1-18.

point by V and the potential energy of a charge q placed at the same point by E_p, we may write

$$V=\frac{E_p}{q} \qquad \text{or} \qquad E_p=qV. \tag{1.14}$$

The zero of electric potential is chosen to coincide with the zero of potential energy; in most cases the zero of electric potential energy has been chosen at infinity. The SI unit for electric potential is joules/coulomb (JC^{-1}), a unit called the *volt* (V) in honor of the Italian scientist Alessandro Volta (1745–1827). In terms of the fundamental units, $V=m^2kg\,s^{-2}\,C^{-1}$.

The relationship between the electric potential and the electric field is important enough to be considered in some detail. Suppose a charge q moves from point A to point B in an electric field (Fig. 1-18). Applying Eq. (1.14) yields the change in potential energy of the charge:

$$E_{p,A}-E_{p,B}=q(V_A-V_B).$$

However according to the definition of potential energy, the left-hand side of this equation gives the work done on the charge when it moves from point A to point B. This work we designate by $W_{A\to B}$. Thus

$$W_{A\to B}=q(V_A-V_B). \tag{1.15}$$

Expression (1.15) permits the *electric potential difference* between two points to be defined as the work done by the electric field to move a *unit* charge from the first point to the second.

Since the electric force on the charge is $F=q\mathcal{E}$ where $\mathcal{E}$ is the electric field, again from fundamental definitions we may write

$$W_{A\to B}=\int_A^B F\cdot dr=\int_A^B q\mathcal{E}\cdot dr$$

where the integral is calculated along a path joining A and B. Combining this equation with Eq. (1.15) and cancelling the common factor of q give

$$\int_A^B \mathcal{E}\cdot dr=V_A-V_b. \tag{1.16}$$

which gives the specific relation between the electric field and the potential difference. Note that if points A and B coincide so that $A=B$, then the path of integration is a closed curve; and

$$\oint \mathscr{E}\cdot d\mathbf{r}=0 \tag{1.17}$$

indicates that the work done by a static electric field when a charge traverses a closed path is zero. Thus the static electric field corresponds to a conservative force.

Referring to Fig. 1-18, Eq. (1.16) can be written in the alternate form

$$\int_A^B \mathscr{E}_s\, ds=V_A-V_B$$

where $\mathscr{E}_s$ is the component of $\mathscr{E}$ along the path. One may also write

$$\int_A^B \mathscr{E}_s\, ds=-(V_B-V_A)=-\int_A^B dV.$$

Suppose now that points A and B are so close that each integral in the equation above practically reduces to a single term. Then

$$\mathscr{E}_s\, ds=-dV \qquad \text{or} \qquad \mathscr{E}_s=-\frac{\partial V}{\partial s}. \tag{1.18}$$

This result shows that the component of the electric field in a certain direction is equal to the negative of the change of electric potential per unit length in the same direction. For example the rectangular components of the electric field are given by

$$\mathscr{E}_x=-\frac{\partial V}{\partial x}, \qquad \mathscr{E}_y=-\frac{\partial V}{\partial y}, \qquad \mathscr{E}_z=-\frac{\partial V}{\partial z}. \tag{1.19}$$

When a relation such as Eq. (1.18) exists, we say that there is a field such that the strength of the field is the negative of the *gradient* of the potential. For this case the electric field is the negative of the gradient of the electric potential. A more compact form of Eq. (1.19) is

$$\mathscr{E}=-\text{grad } V. \tag{1.20}$$

Equation (1.18) or (1.19) is used to find the electric potential when the electric field is known, and conversely.

Consider the simple case of a uniform electric field (Fig. 1-19). The first of Eq. (1.19) gives $\mathscr{E}\, dx=-dV$ when the X-axis is parallel to the field. Since $\mathscr{E}$ is constant, and assuming $V=0$ at $x=0$, integration yields

$$\int_0^V dV=-\int_0^x \mathscr{E}\, dx=-\mathscr{E}\int_0^x dx$$

or

$$V=-\mathscr{E}x. \tag{1.21}$$

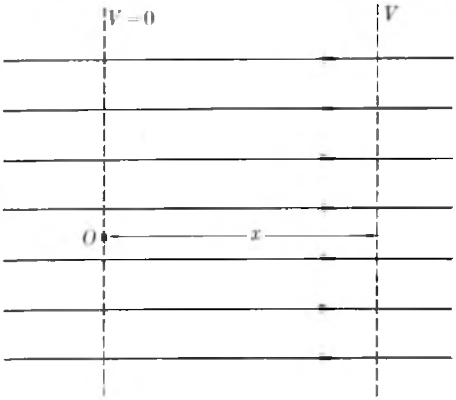


Fig. 1-19. Uniform electric field.

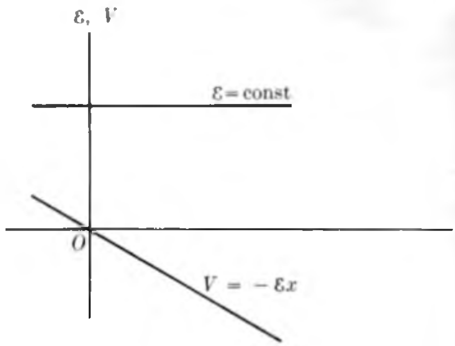


Fig. 1-20. Variations of $\mathscr{E}$ and V for a uniform electric field.

This very useful relation has been represented graphically in Fig. 1-20. Note that because of the negative sign in Eq. (1.20) or Eq. (1.21), the electric field points in the direction in which the electric potential decreases. When the two points x_1 and x_2 are considered, Eq. (1.21) gives $V_1 = -\mathscr{E}x_1$ and $V_2 = -\mathscr{E}x_2$. Subtracting one from the other gives $V_2 - V_1 = -\mathscr{E}(x_2 - x_1)$; or calling $d = x_2 - x_1$, the electric field may be written as

$$\mathscr{E} = -\frac{V_2 - V_1}{d} = \frac{V_1 - V_2}{d}. \tag{1.22}$$

Although this relation is valid only for uniform electric fields, it can be used to *estimate* the electric field between two points separated by a distance d if the potential difference $V_1 - V_2$ between them is known. When the potential difference $V_1 - V_2$ is positive, the field points in the direction from x_1 to x_2; and when the difference is negative, the field points in the opposite direction. Equation (1.22) [or in fact Eq. (1.18) or Eq. (1.19)] indicates that the electric field can also be expressed in volts/meter, a unit which is equivalent to the SI unit newtons/coulomb given before. This equivalence can be seen in the following way:

$$\frac{\text{volt}}{\text{meter}} = \frac{\text{joule}}{\text{coulomb-meter}} = \frac{\text{newton-meter}}{\text{coulomb-meter}} = \frac{\text{newton}}{\text{coulomb}}.$$

By common usage the term volt/meter (V m^{-1}) is preferred to N C^{-1}.

1.7 Electric Potential of a Point Charge

To obtain the electric potential of a point charge, replace s in Eq. (1.18) by the distance r since the electric field is radial; that is, $\mathscr{E} = -\partial V/\partial r$. Remembering Eq. (1.8), this equation may be written as

$$\frac{1}{4\pi\epsilon_0}\frac{q}{r^2}=-\frac{\partial V}{\partial r}$$

or

$$\int dV=-\frac{q}{4\pi\epsilon_0}\int\frac{dr}{r^2}.$$

Integrating and assuming that $V=0$ for $r=\infty$, as in the gravitational case, the potential is

$$V=\frac{q}{4\pi\epsilon_0 r}. \tag{1.23}$$

The electric potential V is positive or negative depending on the sign of the charge q producing it.

If there are several charges q_1, q_2, q_3, ..., the electric potential at a point P (Fig. 1-7) is the scalar sum of their individual potentials. That is.

$$V=\frac{q_1}{4\pi\epsilon_0 r_1}+\frac{q_2}{4\pi\epsilon_0 r_2}+\frac{q_3}{4\pi\epsilon_0 r_3}+\cdots=\frac{1}{4\pi\epsilon_0}\sum_i\frac{q_i}{r_i}. \tag{1.24}$$

Therefore it is in general easier to compute the potential resulting from a distribution of charges and from that potential to calculate the resultant electric field by using Eq. (1.17), than to calculate in the reverse order.

Surfaces having the same electric potential at all points—that is, $V=$const—are called *equipotential surfaces*. At each point of an equipotential surface the direction of the electric field is perpendicular to the surface; that is, the lines of force are orthogonal to the equipotential surfaces. This is so because no work is done between two points when the two points are at the same potential. For a uniform field it may be seen from Eq. (1.21) that $V=$const implies $x=$const. and therefore that the equipotential surfaces are planes as indicated by the dashed lines in Fig. 1-19. For a point charge, Eq. (1.23) indicates that the equipotential surfaces are the spheres $r=$const, indicated by dashed lines in Fig. 1-10(a) and (b). For several charges the equipotential surfaces are given by $\Sigma_i(q_i/r_i)=$const according to Eq. (1.24). For two charges the equipotential surfaces have been indicated by dashed lines in Figs. 1-11 and 1-12.

Example 1.4. Electric potential energy of charge q_3 in Example 1.1.

▼ Refer back to Fig. 1-6 and use Eq. (1.23). The electric potentials produced at C by charges q_1 and q_2 (located at A and B respectively) are

$$V_1=\frac{q_1}{4\pi\epsilon_0 r_1}=11.25\times10^6\text{ V},\qquad V_2=\frac{q_2}{4\pi\epsilon_0 r_2}=-9\times10^6\text{ V}.$$

Thus the total electric potential at C is

$$V = V_1 + V_2 = 2.25 \times 10^6 \text{ V}.$$

The potential energy of charge q_3 is then

$$E_p = q_3 V = (0.2 \times 10^{-3} \text{ C})(2.25 \times 10^6 \text{ V}) = 4.5 \times 10^2 \text{ J}.$$

When this example is compared with Example 1.2, the difference between handling the electric field and the electric potential may be seen. ▲

Example 1.5. Electric field and electric potential produced by a very long, straight filament carrying a charge λ per unit length.

▼ Divide the filament into small portions, each of length ds (Fig. 1-21) and therefore with a charge $dq = \lambda\, ds$. The magnitude of the electric field that the element ds produces at P is

$$d\mathcal{E} = \frac{\lambda\, ds}{4\pi\epsilon_0 r^2}$$

and is directed along the line AP. Because of the symmetry of the problem, for every element ds at distance s above O, there is another element at the same distance below O. Therefore, when the electric fields produced by all elements are added, their components parallel to the filament give a total value of zero. Thus the resultant electric field is along OP and is found by summing all the components parallel to OP given by $d\mathcal{E} \cos\alpha$ due to each element ds:

$$\mathcal{E} = \int d\mathcal{E} \cos\alpha = \frac{\lambda}{4\pi\epsilon_0} \int \frac{ds}{r^2} \cos\alpha.$$

From the figure note that $r = R \sec\alpha$ and $s = R \tan\alpha$ so that $ds = R \sec^2\alpha\, d\alpha$. Making these substitutions, integrating from $\alpha = 0$ to $\alpha = \pi/2$, and multiplying by two (since the two halves of the filament make the same contribution), we may write the electric field as

$$\mathcal{E} = \frac{2\lambda}{4\pi\epsilon_0 R} \int_0^{\pi/2} \cos\alpha\, d\alpha = \frac{\lambda}{2\pi\epsilon_0 R}.$$

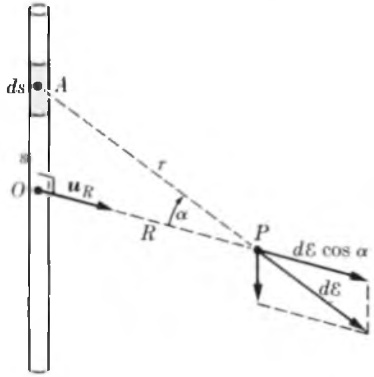


Fig. 1-21. Electric field of a very long, charged filament.

Therefore the electric field of the filament varies as R^{-1}. In vector form

$$\mathscr{E}=\frac{\lambda}{2\pi\epsilon_0 R}u_R.$$

To find the electric potential, the relation $\mathscr{E}=-\text{grad } V$ allows us to write $\mathscr{E}=-u_r(\partial V/\partial R)$ because the electric field is the only radical. Therefore

$$\frac{dV}{dR}=-\frac{\lambda}{2\pi\epsilon_0 R}.$$

Integration yields

$$V=-\frac{\lambda}{2\pi\epsilon_0}\ln R+C.$$

The value of the constant C is fixed by assigning a convenient value to the potential at a certain point in space. It is customary in this case to assign the zero of the potential to the point where $R=1$, giving $C=0$.

Therefore the electric potential is

$$V=-\frac{\lambda}{2\pi\epsilon_0}\ln R.$$

The student should attempt to solve this problem in the reverse order by first finding the potential and afterward the field. ▲

1.8 Energy Relations in an Electric Field

The total energy of a charged particle or ion that has mass m and charge q and moves in an electric field is

$$E=E_k+E_p=\tfrac{1}{2}mv^2+qV. \tag{1.25}$$

When the ion moves from position P_1 (where the electric potential is V_1) to position P_2 (where the potential is V_2), Eq. (1.25), combined with the principle of conservation of energy, gives

$$\tfrac{1}{2}mv_1^2+qV_1=\tfrac{1}{2}mv_2^2+qV_2. \tag{1.26}$$

Recalling the definition of kinetic energy, $W=\frac{1}{2}mv_2^2-\frac{1}{2}mv_1^2$ is the work done on the charged particle when it moves from P_1 to P_2. Rearrangement of Eq. (1.26) gives

$$W=\tfrac{1}{2}mv_2^2-\tfrac{1}{2}mv_1^2=q(V_1-V_2). \tag{1.27}$$

Equation (1.27) allows a precise definition of the volt as equal to the electric potential difference through which a charge of one coulomb has to move to gain an amount of energy equal to one joule.

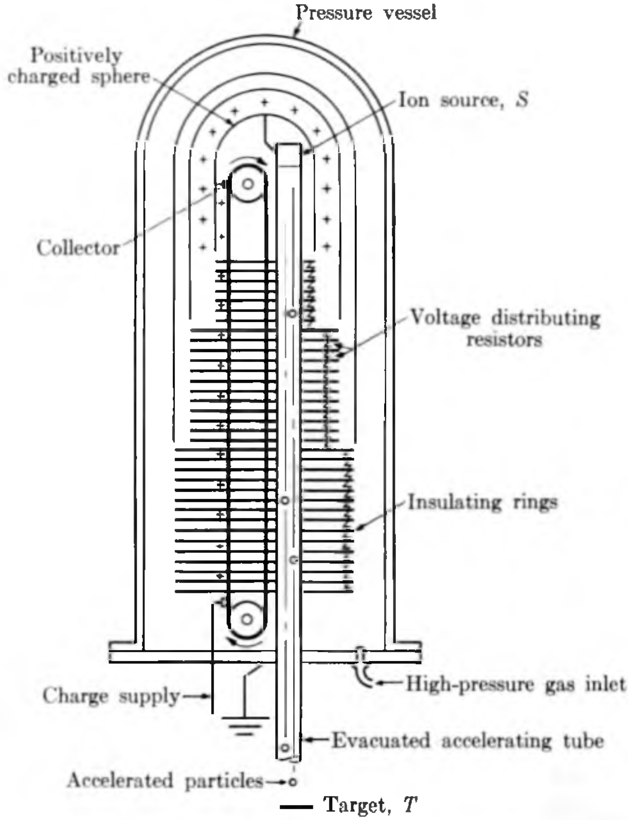


Fig. 1-22. Simplified cross section of a Van de Graaff electrostatic accelerator. A high-speed motor runs a belt, made of an insulating material, over two pulleys. Electric charge from a voltage source is picked up by the belt at the lower end and conveyed upward. A collector draws the charge off onto the metal sphere at the top, and builds up a high electric potential on the sphere. Positive ions are produced at this high voltage end and are accelerated downward by the potential difference between the charged sphere and the ground potential at the other end.

Note from Eq. (1.27) that a positively charged particle ($q>0$) gains kinetic energy when moving from a larger to a smaller potential ($V_1>V_2$); a negatively charged particle ($q<0$) must move from a lower to a higher potential ($V_1<V_2$) to gain kinetic energy.

If the zero of electric potential is chosen at $P_2(V_2=0)$ and the experiment is arranged so that at P_1 the ions have zero velocity ($v_1=0$), Eq. (1.27) becomes (without the subscripts)

$$\tfrac{1}{2}mv^2=qV, \tag{1.28}$$

an expression that gives the kinetic energy acquired by a charged particle when it starts from rest and moves through an electric potential difference V. This principle is applied in *electrostatic accelerators*, for example.

A typical accelerator (Fig. 1-22) consists of an evacuated tube through which an electric potential difference V is applied. At one end there is an ion source S injecting charged particles into the tube. The particles arrive at the other end with an energy given by Eq. (1.28). These fast ions impinge on a target T, made of a material chosen according to the nature of the experiment to be performed. The result of this collision is some kind of nuclear reaction. The energy of the impinging ions is transferred to the target, which therefore must be constantly cooled since otherwise it would melt or vaporize.

There are several types of electrostatic accelerators (Cockroft-Walton, Van de Graaff, etc.). Each depends on a different method for producing the potential difference V. In any case, electrostatic accelerators are limited in energy by the maximum voltage difference that can be applied without producing an electrical breakdown of the materials used. This potential difference cannot exceed a few million volts.

Because the fundamental particles and nuclei have a charge that is equal to the fundamental charge e or is a multiple of it, Eq. (1.27) suggests defining a new unit of energy, the *electron volt*, (eV). An electron volt is equal to the energy gained by a particle of charge e when it moves through a potential difference of one volt. Thus using the value of e from Eq. (1.13) gives

$$\mathrm{eV} = (1.6021 \times 10^{-19}\ \mathrm{C})\,(1\ \mathrm{V}) = 1.6021 \times 10^{-19}\ \mathrm{J}.$$

In moving through a potential difference ΔV, a particle of charge ve, gains an energy of $v\,\Delta V$ eV. Convenient multiples of the electron volt are the *kiloelectron volt* (keV) and the *megaelectron volt* (MeV), commonly called the million electron volt.

It is very useful to express the rest mass energy of the fundamental particles in this unit. The results are

$$\begin{aligned} E_e &= m_e c^2 = 8.1867 \times 10^{-14}\ \mathrm{J} = 0.5110\ \mathrm{MeV}. \\ E_p &= m_p c^2 = 1.5032 \times 10^{-10}\ \mathrm{J} = 938.26\ \mathrm{MeV}, \\ E_n &= m_n c^2 = 1.5053 \times 10^{-10}\ \mathrm{J} = 939.55\ \mathrm{MeV}. \end{aligned} \tag{1.29}$$

1.9 Electric Current

The example of an electrostatic accelerator (Section 1.8) with a stream of charged particles swiftly accelerated along its tube suggests introducing at this time the very important concept of *electric current*. An electric current consists of a stream of charged particles or ions. This definition applies to ions in an accelerator, in an electrolytic solution, in an ionized gas or plasma, or to electrons in a metallic conductor. In order for an electric current to be produced, an electric field must be applied to move the charged particles in a well-defined direction.

The *intensity* of an electric current is defined as the electric charge passing per unit time through a section of the region where the charge flows, such as a section of

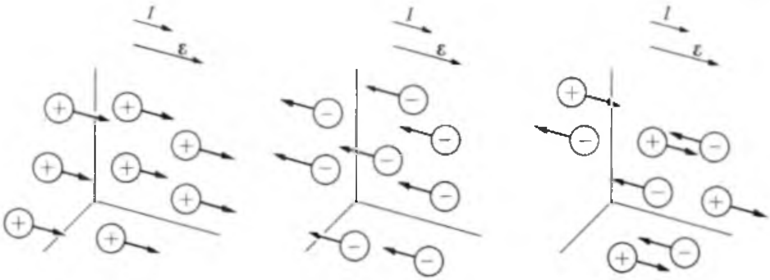


Fig. 1-23. Motion of positive and negative ions resulting in an electric current I produced by an electric field $\mathscr{E}$.

an accelerator tube or of a metallic wire. Therefore if in time t, N charged particles, each carrying a charge q, pass through a section of the conducting medium, the total charge passing is $Q = Nq$; and the intensity of the current is

$$I = \frac{Nq}{t} = \frac{Q}{t}. \tag{1.30}$$

Actually this expression gives the average current in time t; the instantaneous current is

$$I = \frac{dQ}{dt}. \tag{1.31}$$

Electric current is expressed in coulombs/second or s^{-1} C, a unit called the *ampere* (A) in honor of the French physicist André M. Ampère (1775–1836). An ampere is the intensity of an electric current corresponding to a charge of one coulomb passing through a section of the material every second.

The *direction* of an electric current is assumed to be the same as that of moving positively charged particles. Also, the direction is the same as that of the applied electric field or of the potential drop that produces the motion of the charged particles (Fig. 1-23a). Therefore if a current is due to the motion of negatively charged particles, such as electrons, by definition the direction of the current is opposite to the actual motion of the electrons (Fig. 1-23b).

Maintaining an electric current requires the expenditure of energy because the moving ions are accelerated by the electric field. Suppose that N ions, each with charge q, move through a potential difference ΔV in time t. Each ion gains an energy $q\Delta V$, and the total energy gained is $Nq\Delta V = Q\Delta V$. The energy per unit time, or the *power* required to maintain the current, is then

$$P = \frac{Q\Delta V}{t} = I\Delta V. \tag{1.32}$$

This equation gives, for example, the power required to drive the accelerator discussed in the previous section. The equation also gives the rate at which energy is transferred to the accelerator's target, and therefore the rate at which energy must be removed by the target's coolant. Expression (1.32) is thus of general validity and gives the power required to maintain an electric current I through a potential difference V applied to two points of any conducting media. Note from Eq. (1.32) that

$$\text{volts} \times \text{amperes} = \frac{\text{joules}}{\text{coulomb}} \times \frac{\text{coulombs}}{\text{second}} = \frac{\text{joules}}{\text{second}} = \text{watts}$$

so that the units are consistent.

1.10 Electric Dipole

An interesting and important arrangement of charges is an *electric dipole*. The arrangement consists of two equal and opposite charges $+q$ and $-q$, separated a very small distance a (Fig. 1-24).

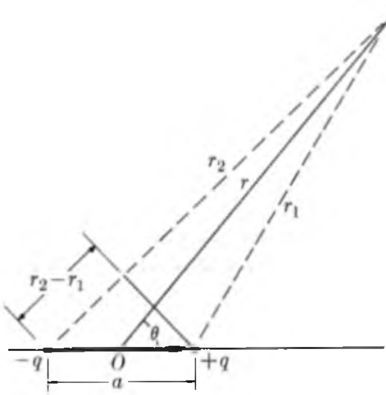


Figure 1-24

The electric dipole moment $\boldsymbol{p}$* is defined by

$$\boldsymbol{p} = q\boldsymbol{a} \tag{1.33}$$

where $\boldsymbol{a}$ is the displacement from the negative to the positive charge. At a point P the electric potential due to the electric dipole is found from Eq. (1.24):

$$V = \frac{1}{4\pi\epsilon_0}\left(\frac{q}{r_1} - \frac{q}{r_2}\right) = \frac{1}{4\pi\epsilon_0}\,\frac{q(r_2 - r_1)}{r_1 r_2}.$$

*Note that by custom the symbols for momentum and for electric dipole moment are the same.

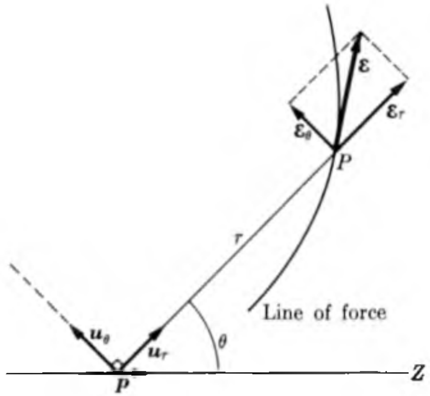


Figure 1-25

If the distance a is very small compared with r, the approximations

$$r_2 - r_1 = a \cos\theta \qquad \text{and} \qquad r_1 r_2 = r^2$$

may be made and result in

$$V \cong \frac{qa\cos\theta}{4\pi\epsilon_0 r^2}$$

or (1.34)

$$V = \frac{p\cos\theta}{4\pi\epsilon_0 r^2}.$$

To obtain the electric field intensity of the electric dipole. Eq. (1.18) is used. Since Eq. (1.18) gives the electric potential in polar coordinates, the components of $\mathscr{E}$ will be given in polar coordinates. To obtain the radial component $\mathscr{E}_r$, note that $ds = dr$. Then from Eq. (1.18)

$$\mathscr{E}_r = -\frac{\partial V}{\partial r} = \frac{2p\cos\theta}{4\pi\epsilon_0 r^3}. \tag{1.35}$$

For the transverse component $\mathscr{E}_\theta$ the appropriate polar coordinate element is the arc length $ds = r\,d\theta$, resulting in

$$\mathscr{E}_\theta = -\frac{1}{r}\frac{\partial V}{\partial \theta} = \frac{p\sin\theta}{4\pi\epsilon_0 r^3}. \tag{1.36}$$

These two components are shown in Fig. 1-25. The lines of force for a small electric dipole are indicated in Fig. 1-26. Although in an electric dipole the two charges are equal and opposite. and result in a zero net charge, their slight displacement is enough to produce a nonvanishing electric field.

In atoms the center of mass of the electrons coincides with the nucleus, and therefore the average electric dipole moment of the atom is zero (Fig. 1-27a). If an external field is applied, the electronic motion is distorted and the center of mass of

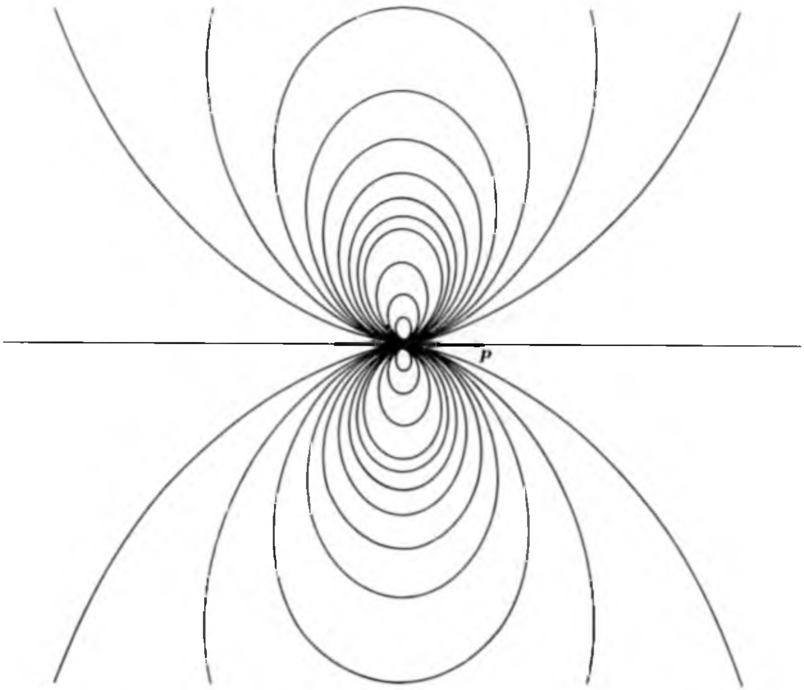


Fig. 1-26. Lines of force of the electric field of an electric dipole.

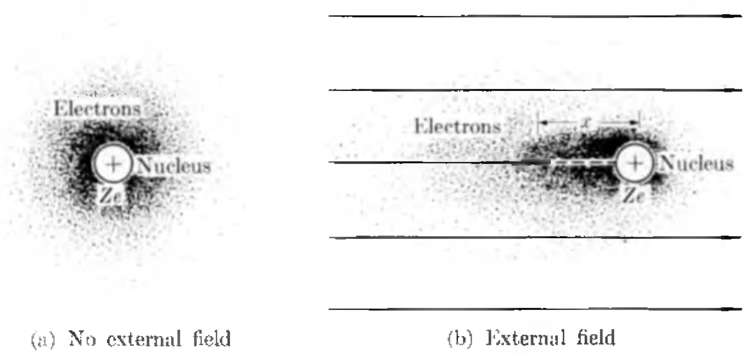


Fig. 1-27. Polarization of an atom under an external electric field.

the electrons is displaced a distance x relative to the nucleus (Fig. 1-27b). The atom is thus *polarized* and becomes an electric dipole of moment $\boldsymbol{p}$. This moment is proportional to the external field $\mathscr{E}$.

Molecules, on the other hand, may have a permanent electric dipole moment. Such molecules are called *polar*. For example in the HCl molecule (Fig. 1-28) the

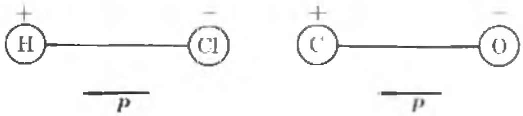


Fig. 1-28. Polar diatomic molecules.

electron of the H atom spends more time moving around the Cl atom than around the H atom. Therefore the center of negative charges does not coincide with that of the positive charges, and the molecule has a dipole moment directed from the Cl atom to the H atom. That is, we may write H^+Cl^-. The electric dipole of the HCl molecule is $p = 3.43 \times 10^{-30}$ C m. In the CO molecule, the charge distribution is only slightly asymmetric and the electric dipole moment is relatively small, about 0.4×10^{-30} C m, with the carbon atom corresponding to the positive end and the oxygen atom to the negative end of the molecule.

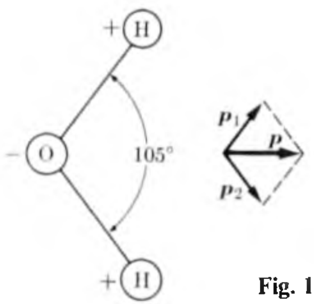


Fig. 1-29. Electric dipole of an H_2O molecule.

In a molecule such as H_2O, in which the two H—O bonds are at an angle slightly over 90° (Fig. 1-29), the electrons try to crowd around the oxygen atom, which thereupon becomes slightly negative relative to the H atoms. Each H—O bond thus contributes to the electric dipole moment, whose resultant, because of symmetry, lies along the axis of the molecule and has a value equal to 6.2×10^{-30} C m. In the CO_2 molecule all the atoms are in a straight line (Fig. 1-30), and the resultant electric dipole moment is zero because of the symmetry. Thus electric dipole moments can provide useful information about the structure of molecules. The values of p for several polar molecules are given in Table 1-2.

When an electric dipole is placed in an electric field, a force is produced on each charge of the dipole (Fig. 1-31). The resultant force is zero unless the electric field is

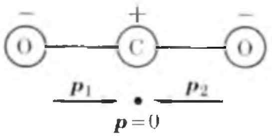


Fig. 1-30. The CO_2 molecule has no electric dipole.

Table 1-2 Electric Dipole Moments for Selected Polar Molecules*

Molecule	p, Cm
HCl	3.43×10^{-30}
HBr	2.60×10^{-30}
HI	1.26×10^{-30}
CO	0.40×10^{-30}
H_2O	6.2×10^{-30}
H_2S	5.3×10^{-30}
SO_2	5.3×10^{-30}
NH_3	5.0×10^{-30}
C_2H_5OH	3.66×10^{-30}

* Molecules with zero dipole moment include CO_2, H_2, CH_4 (methane), C_2H_6 (ethane), and CCl_4 (carbon tetrachloride).

not uniform. The net force is

$$\boldsymbol{F}_{\text{net}}=q\mathscr{E}_{+}-q\mathscr{E}_{-}=q(\mathscr{E}_{+}-\mathscr{E}_{-}).$$

Consider the special case in which the electric field is along the X-axis and the dipole is oriented parallel to the field. Then if we consider magnitudes only, $\mathscr{E}_{+}-\mathscr{E}_{-}=(d\mathscr{E}/dx)a$, and therefore $\boldsymbol{F}=p(d\mathscr{E}/dx)$. This result shows that *an electric dipole oriented parallel to a nonuniform electric field tends to move in the direction in which the field increases*. The opposite result is obtained if the dipole is oriented anti-parallel to the field. Of course when the electric field is uniform, the resultant force on the electric dipole is zero irrespective of its orientation.

The potential energy of the dipole is

$$E_p=qV_{+}-qV_{-}=q(V_{+}-V_{-})=-qa\left(-\frac{V_{+}-V_{-}}{a}\right).$$

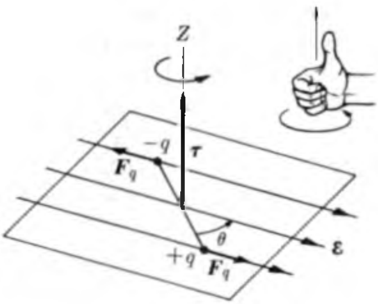


Fig. 1-31. Electric dipole in an external electric field.

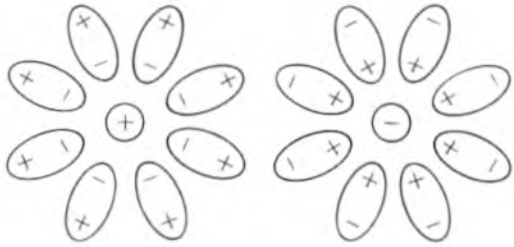


Fig. 1-32. Polarization effects of an ion in solution.

If Eq. (1.19) is used to describe the uniform electric field and θ is the angle between the dipole and the electric field, the last factor is just the component $\mathscr{E}_a = \mathscr{E} \cos\theta$ of the field $\boldsymbol{\mathscr{E}}$ parallel to $\boldsymbol{a}$. Therefore $E_p = -qa\mathscr{E}_a$, or

$$E_p = -p\mathscr{E}\cos\theta = -\boldsymbol{p}\cdot\boldsymbol{\mathscr{E}}. \tag{1.37}$$

The potential energy is a minimum when $\theta = 0$, which indicates that *the electric dipole is in equilibrium when it is oriented parallel to the field.* If the slight difference between $\boldsymbol{\mathscr{E}}_+$ and $\boldsymbol{\mathscr{E}}_-$ is neglected (or when the field is uniform), the forces $q\boldsymbol{\mathscr{E}}_+$ and $-q\boldsymbol{\mathscr{E}}_-$ on the charges comprising the dipole form a couple whose torque is

$$\boldsymbol{\tau} = \boldsymbol{a} \times (q\boldsymbol{\mathscr{E}}) = (q\boldsymbol{a}) \times \boldsymbol{\mathscr{E}} = \boldsymbol{p} \times \boldsymbol{\mathscr{E}}. \tag{1.38}$$

From this expression, as well as from Fig. 1-31, it may be seen that *the torque of the electric field tends to align the dipole parallel to the field.* The magnitude of the torque is $\tau = p\mathscr{E}\sin\theta$, and its direction is as indicated in Fig. 1-31.

These properties of an electric dipole when placed in an electric field have very important applications. For example, the electric field of an ion in solution polarizes the molecules of the solvent that surrounds the ions and they become oriented in the form indicated in Fig. 1-32. These oriented molecules become more or less attached to the ion, increase its effective mass, and decrease its effective charge, which is partially screened by the molecules. The net effect is that the mobility of the ion in an external field is decreased. Also when a gas or a liquid whose molecules are permanent dipoles is placed in an electric field, the molecules tend to align with their dipoles parallel as a result of the torques due to the electric field. We say then that the substance has been *polarized*. This matter will be taken up in the next chapter.

Example 1.6 Electric field of an electric dipole expressed in vector form.

▼ From Fig. 1-25 the electric field of a dipole may be written as

$$\boldsymbol{\mathscr{E}} = \boldsymbol{u}_r\mathscr{E}_r + \boldsymbol{u}_\theta\mathscr{E}_\theta = \frac{1}{4\pi\epsilon_0 r^3}(\boldsymbol{u}_r 2p\cos\theta + \boldsymbol{u}_\theta p\sin\theta).$$

From the same figure it is possible to write the electric dipole as

$$p = p(u_r \cos\theta - u_\theta \sin\theta).$$

Using this expression to eliminate the term $p \sin\theta$ in the expression for $\mathscr{E}$ yields

$$\mathscr{E} = \frac{1}{4\pi\epsilon_0 r^3}(3u_r p \cos\theta - p).$$

Also $p\cos\theta = u_r \cdot p$. Therefore

$$\mathscr{E} = \frac{3u_r(u_r \cdot p) - p}{4\pi\epsilon_0 r^3},$$

which gives the electric dipole field in vector form. ▲

Example 1.7. Calculation of the interaction energy between two electric dipoles. The result will be used to estimate the interaction energy between two water molecules. Discussion of relative orientation effects.

▼ In Example 1.6 the electric field produced by one dipole at distance r was derived in a special form that relied on the dipole orientation. With its electric dipole moment called p_1, that result may be rewritten as

$$\mathscr{E}_1 = \frac{3u_r(u_r \cdot p_1) - p_1}{4\pi\epsilon_0 r^3}.$$

If the moment of the second dipole is designated as p_2, the interaction energy between the two dipoles, using Eq. (1.37), is

$$E_{p,12} = -p_2 \cdot \mathscr{E}_1 = -\frac{3(u_r \cdot p_1)(u_r \cdot p_2) - p_1 \cdot p_2}{4\pi\epsilon_0 r^3}. \tag{1.39}$$

Several important conclusions can be derived from this result. (1) The interaction energy $E_{p,12}$ is *symmetric* with respect to the two dipoles because everything remains the same if p_1 and p_2 are interchanged. This is a result to be expected. (2) The interaction between two dipoles is *noncentral* because it depends on the angles of the position vector r (or of the unit vector u_r) with p_1 and p_2. Consequently, motion under a dipole–dipole interaction does not conserve the orbital angular momentum of the dipoles. (3) The force between two dipoles is not along the line joining them (except in certain specific positions). (4) Since the potential energy between electric dipoles varies as r^{-3} with the distance, the force, which is the negative gradient of the potential energy, decreases as r^{-4}; therefore the interaction between two electric dipoles diminishes with the distance much faster than the interaction between two charges.

The geometry corresponding to Eq. (1.39) is illustrated in Fig. 1-33, in which (a) corresponds to the general case. In (b) the two dipoles are aligned along the line joining them. Thus $p_1 \cdot p_2 = p_1 p_2$, $u_r \cdot p_1 = p_1$, and $u_r \cdot p_2 = p_2$ so that

$$E_{p,12} = -\frac{2p_1p_2}{4\pi\epsilon_0 r^3};$$

and an attraction results between the dipoles because of the negative sign. In (c) again $p_1 \cdot p_2 = p_1 p_2$;

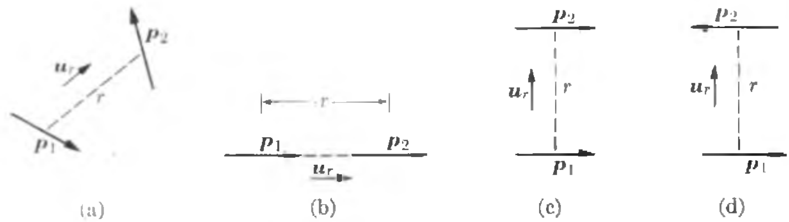


Fig. 1-33. Interaction between two electric dipoles.

but $u_r \cdot p_1 = 0$ and $u_r \cdot p_2 = 0$ so that

$$E_{p,12} = +\frac{p_1 p_2}{4\pi\epsilon_0 r^3}.$$

Since it is positive, this value indicates a repulsion between the dipoles. Finally in (d) we have $p_1 \cdot p_2 = -p_1 p_2$, which gives

$$E_{p,12} = -\frac{p_1 p_2}{4\pi\epsilon_0 r^3},$$

meaning that there will be attraction between the dipoles.

An understanding of the interaction between two electric dipoles is of great importance because molecular forces are due in large part to this type of interaction. Consider two water molecules in the relative position of Fig. 1-33(b) at their normal separation in the liquid phase, about 3.1×10^{-10} m. Their electric dipole moment is 6.1×10^{-30} m C. Therefore their interaction potential energy is calculated as

$$E_{p,12} = -\frac{9 \times 10^9 \times 2 \times (6.1 \times 10^{-30})^2}{(3.1 \times 10^{-10})^3} = -2.22 \times 10^{-20}\ \text{J}.$$

This result is larger by a factor of ten than the interaction energy mentioned in Section 13.9 of Volume I, which was estimated using the value of the heat of vaporization. The student, however, must realize that the present result corresponds to the *instantaneous* interaction energy between two water molecules in the relative position of Fig. 1-33(b). Since water molecules are in continuous motion, their relative orientation is continuously changing. Thus to obtain $E_{p,12}$ Eq. (1.37) must be averaged over all possible relative orientations. When this averaging is done, there is better agreement.

The student should compare the result above for the electric interaction $E_{p,12}$ between two water molecules with the corresponding gravitational interaction for the same relative position. ▲

1.11 Higher Electric Multipoles

It is possible to define higher-order or multipole electric moments. For example, charge distribution such as that in Fig. 1-34 constitutes an *electric quadrupole*. Its total charge is zero and we may show that its electric dipole moment is also zero. It

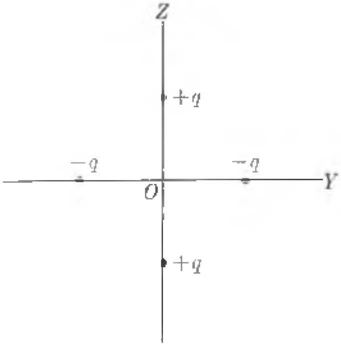


Fig. 1-34. Electric quadrupole.

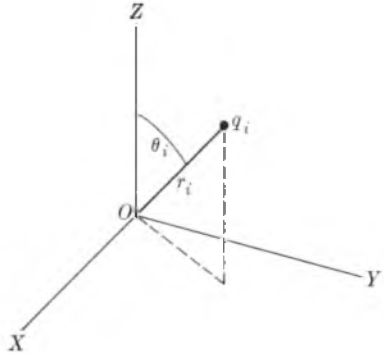


Figure 1-35

is not easy to give a general definition of the *electric quadrupole moment* in an elementary way. However, it may be shown that the electric quadrupole moment of a charge distribution relative to a symmetry axis, such as the Z-axis, is defined by

$$Q = \tfrac{1}{2} \sum_i q_i r_i^2 (3 \cos^2 \theta_i - 1) \tag{1.40}$$

where r_i is the distance of charge i from the center and θ_i is the angle that r_i makes with the axis (Fig. 1-35). Note that $z_i = r_i \cos \theta_i$. Then Eq. (1.40) may be written as

$$Q = \tfrac{1}{2} \sum_i q_i (3z_i^2 - r_i^2). \tag{1.41}$$

The electric quadrupole moment is zero for a spherical distribution of charge, positive for an elongated or prolate charge distribution, and negative for an oblate or flattened charge distribution (Fig. 1-36). Therefore the electric quadrupole moment gives an indication of the degree of departure from the spherical form of a charge distribution. For example, atomic nuclei are usually considered spherical. However, careful measurements indicate that certain nuclei have relatively large electric quadrupole moments. This has been interpreted as indicating that such

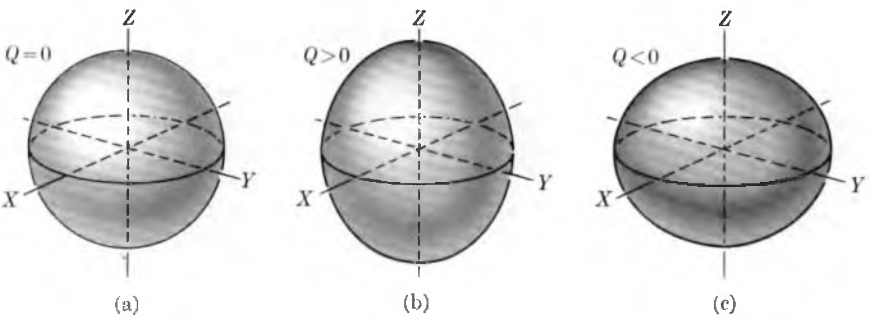


Fig. 1-36. Electric quadrupole of ellipsoidal charge distributions.

nuclei are greatly deformed, and thus the electric field they produce departs from that of a point charge. This in turn affects the energy of the electronic motion.

It should be noted that the potential of a point charge decreases as r^{-1} and the field as r^{-2} while for an electric dipole the potential decreases as r^{-2} and the field as r^{-3}. In an analogous way it can be proved that for an electric quadrupole the potential varies as r^{-3} and the field as r^{-4}. Similar results are obtained for higher-order multipoles. That is, the higher the multipole order, the smaller the range within which its electric field has any noticeable effect.

Example 1.8. Electric potential for the charge arrangement of a *linear electric quadrupole* (Fig. 1-37).

▼ The total charge of the system is zero. Also the electric dipole moment is zero since $p=+q(+a)-2q(0)+q(-a)=0$. However, the electric quadrupole moment, using Eq. (1.41), is

$$Q=\tfrac{1}{2}\{q(3a^2-a^2)-2q(0)+q[3(-a)^2-a^2]\}=2qa^2.$$

The electric potential produced by the system of charges at point P is

$$V=\frac{1}{4\pi\epsilon_0}\left(\frac{q}{r_1}-\frac{2q}{r}+\frac{q}{r_2}\right)=\frac{q}{4\pi\epsilon_0}\left(\frac{1}{r_1}-\frac{2}{r}+\frac{1}{r_2}\right). \qquad (1.42)$$

From the figure we see that

$$r_1=(r^2-2ar\cos\theta+a^2)^{1/2}.$$

or

$$r_1=r\left(1-\frac{2a\cos\theta}{r}+\frac{a^2}{r^2}\right)^{1/2}$$

and

$$\frac{1}{r_1}=\frac{1}{r}\left(1-\frac{2a\cos\theta}{r}+\frac{a^2}{r^2}\right)^{-1/2}. \qquad (1.43)$$

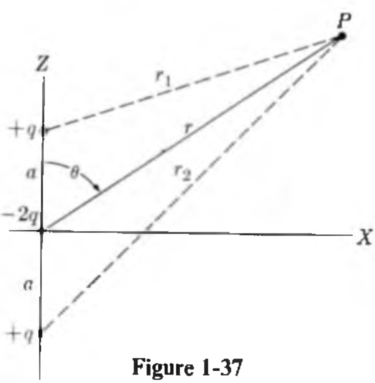


Figure 1-37

Next assume a is very small compared with r. From the binomial expansion given by Eq. (M.22) $(1+x)^{-1/2}=1-\frac{1}{2}x+\frac{3}{8}x^2+\cdots$ with $x=-2a\cos\theta/r+a^2/r^2$. Then we may rewrite Eq. (1.43) as

$$\frac{1}{r_1}=\frac{1}{r}\left[1-\frac{1}{2}\left(-\frac{2a\cos\theta}{r}+\frac{a^2}{r^2}\right)+\frac{3}{8}\left(-\frac{2a\cos\theta}{r}+\frac{a^2}{r^2}\right)^2+\cdots\right].$$

Expanding the bracket and keeping only terms having r^3 or less in the denominator gives

$$\frac{1}{r_1}=\frac{1}{r}+\frac{a\cos\theta}{r^2}+\frac{a^2}{2r^3}(3\cos^2\theta-1)+\cdots. \tag{1.44}$$

Similarly $r_2=(r^2+2ar\cos\theta+a^2)^{1/2}$, and therefore

$$\frac{1}{r_2}=\frac{1}{r}-\frac{a\cos\theta}{r^2}+\frac{a^2}{2r^3}(3\cos^2\theta-1)+\cdots. \tag{1.45}$$

If we substitute both results (1.44) and (1.45) in Eq. (1.42) and simplify, the potential is given by

$$V=\frac{qa^2(3\cos^2\theta-1)}{4\pi\epsilon_0 r^3}.$$

Therefore with $Q=2qa^2$,

$$V=\frac{Q(3\cos^2\theta-1)}{2(4\pi\epsilon_0)r^3}, \tag{1.46}$$

which gives the electric potential of a linear electric quadrupole. The electric field may be obtained by applying Eq. (1.20), as previously done in Section 1.10 for the electric dipole. ▲

Problems

1.1 Find the electric force of repulsion between the two protons in a hydrogen molecule if their separation is 0.74×10^{-10} m. Compare your answer with their gravitational attraction at the same separation.

1.2 Find the electric force of attraction between the proton and the electron in a hydrogen atom if the electron describes a circular orbit of 0.53×10^{-10} m radius. Compare your answer with their gravitational attraction at the same separation.

1.3 Compare the electrostatic repulsion between two electrons with their gravitational attraction at the same distance. Repeat for two protons. (Recall Problem 1.1.)

1.4 Two identical cork balls of mass m have equal charges q (Fig. 1-38). The balls are attached to two strings, each having length l and hanging from the same point. Find the angle θ the strings will make with the vertical when equilibrium is reached.

1.5 What must be the charge on a particle of mass 2×10^{-3} kg for it to remain stationary in the laboratory when the particle is placed in a downward-directed uniform electric field of intensity 500 N C^{-1}?

1.6 The electric field in the region between the deflecting plates of a certain cathode-ray oscilloscope is 3×10^4 N C^{-1}. (a) What is the

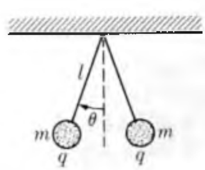


Figure 1-38

force on an electron in this region? (b) What is the acceleration of an electron when it is acted on by this force? Compare your answer with the acceleration of gravity.

1.7 A charge of $+2.5 \times 10^{-8}$ C is placed in an upward-directed uniform electric field whose intensity is 5×10^4 N C^{-1}. What is the work of the electrical force on the charge when the charge moves (a) 4.5 m horizontally? (b) 0.8 m downward? (c) 2.6 m at an angle of 45° upward from the horizontal?

1.8 A uniform electric field exists in the region between two oppositely charged plane parallel plates. An electron is released from rest at the surface of the negatively charged plate and strikes the surface of the opposite plate, 2×10^{-2} m distant from the first, in a time interval of 1.5×10^{-8} sec. (a) Calculate the electric field. (b) Calculate the velocity of the electron when it strikes the second plate.

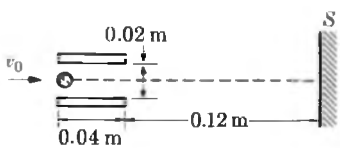


Figure 1-39

1.9 In Fig. 1-39 an electron, with an initial horizontal velocity of 2×10^7 ms^{-1}, is projected along the axis midway between a pair of plates having a uniform electric field with an intensity of 2×10^4 N C^{-1} in the upward direction. Assume that the field begins and ends sharply at the edges of the plates. (a) How far below the axis has the electron moved when it reaches the end of the plates? (b) At what angle with the axis is the electron moving as it leaves the plates? (c) How far below the axis will the electron strike the fluorescent screen S?

1.10 An electron is projected into a uniform electric field of intensity 5×10^3 N C^{-1}. The direction of the field is vertically upward. The initial velocity of the electron is 10^7 m s^{-1}, at an angle of 30° above the horizontal. (a) Calculate the time required for the electron to reach its maximum height. (b) Calculate the maximum distance the electron rises vertically above its initial elevation. (c) After what horizontal distance does the electron return to its original elevation? (d) Sketch the trajectory of the electron.

1.11 An oil droplet of mass 3×10^{-14} kg and of radius 2×10^{-6} m carries 10 excess electrons. What is the terminal velocity of the droplet (a) when it falls in a region in which there is no electric field? (b) when the droplet falls in an electric field whose intensity is 3×10^5 N C^{-1} directed downward? The viscosity of air is 1.80×10^{-5} Pa s. Neglect the buoyant force of the air.

1.12 A charged oil drop in a Millikan oil-drop apparatus is observed to fall through a distance of 1 mm in 27.4 sec, in the absence of any external field. The same drop can be held stationary in a field of 2.37×10^4 N C^{-1}. How many excess electrons has the drop acquired? The viscosity of air is 1.80×10^{-5} Pa s. The density of the oil is 800 kg m^{-3}, and the density of air is 1.30 kg m^{-3}.

1.13 A charged oil drop falls 4.0 mm in 16.0 sec at constant speed in air in the absence of an electric field. The density of the oil is 800 kg m^{-3}, that of the air is 1.30 kg m^{-3}, and the coefficient of viscosity of the air is 1.80×10^{-5} Pa s. (a) Calculate the radius and the mass of the drop. (b) If the drop carries one electronic unit of charge and is in an electric field of 2×10^5 N C^{-1}, what is the ratio of the force of the electric field on the drop to its weight?

1.14 Two point charges, 5 μC and -10 μC, are spaced 1 m apart. (a) Find the magnitude and the direction of the electric field at a point 0.6 m from the first charge and 0.8 from the second charge. (b) Where is the electric field zero because of these two charges?

1.15 In an apparatus for measuring the electronic charge e by Millikan's method, an electric field of 6.34×10^4 N C^{-1} is required to maintain a charged oil drop at rest. If the plates are 1.5×10^{-2} m apart, what potential difference between them is required?

1.16 Three positive charges of 2×10^{-7} C, 1×10^{-7} C, and 3×10^{-7} C are in a straight line with the second charge in the center so that the separation between two adjacent charges is 0.10 m. Calculate (a) the resultant force on each charge caused by the others, (b) the potential energy of each charge caused by the others, and (c) the internal potential energy of the system. (d) Compare (c) with the sum of the results obtained in (b), and explain.

1.17 Solve Problem 1.16 for a case in which the second charge is negative.

1.18 The electric potential at a certain distance from a point charge is 600 V and the electric field is 200 N C^{-1}. (a) What is the distance to the point charge? (b) What is the magnitude of the charge?

1.19 The maximum charge that can be retained by one of the spherical terminals of a large Van de Graaff generator is about 10^{-3} C. Assume that a positive charge of this magnitude is distributed uniformly over the surface of a sphere (of radius 2.7 m) in otherwise empty space. (a) Compute the magnitude of the electric intensity at a point that is outside the sphere and is 5 m from its center. (b) If an electron were released at this point, what would the magnitude and the direction of the electron's initial acceleration be? (c) What would be the velocity of the electron when it reached the sphere?

1.20 A small sphere of mass 2×10^{-4} kg hangs by a thread between two parallel vertical plates 5×10^{-2} m apart. The charge on the sphere is 6×10^{-9} C. What is the electric potential difference between the plates if the thread assumes an angle of 10° with the vertical?

1.21 Two positive point charges of 2×10^{-7} C and 3×10^{-7} C are separated by a distance of 0.10 m. Compute the resultant electric field and electric potential (a) at the midpoint between them, (b) at a point 0.04 m from the first and on the line between them, (c) at a point 0.04 m from the first and on the line joining the charges but outside them, and (d) at a point 0.10 m from each charge.

1.22 Solve Problem 1.21 for a case in which the second charge is negative.

1.23 Referring again to Problem 1.21, calculate the work required to move a charge of 4×10^{-7} C from the point in (c) to the point in (d). Is it necessary to specify the path?

1.24 Two positive point charges, each of magnitude q, are fixed on the Y-axis at the points $y=+a$ and $y=-a$. (a) Draw a diagram showing the positions of the charges. (b) What is the electric potential at the origin? (c) Show that the electric potential at any point on the X-axis is

$$V=\frac{q}{2\pi\epsilon_0\sqrt{a^2+x^2}}.$$

(d) Sketch a graph of the electric potential on the X-axis as a function of x over the range from $x=+5a$ to $x=-5a$. (e) At what value of x is the electric potential one-half that at the origin? (f) From (c), obtain the electric field on the X-axis.

1.25 For the charges in Problem 1.24, suppose that a positively charged particle of charge q' and mass m is displaced slightly from the origin in the direction of the X-axis and is then released. (a) What is the particle's velocity at infinity? (b) Sketch a graph of the velocity of the particle as a function of x. (c) If the particle is projected toward the left along the X-axis from a point at a large distance to the right of the origin, and the velocity of the particle is half that acquired in part (a), at what distance from the origin will the particle come to rest? (d) If a negatively charged particle at rest on the X-axis were released at a very large distance from the origin, what would be the velocity of the particle as it passed the origin?

1.26 (a) Referring again to the charges described in Problem 1.24, make a plot of the electric potential along the Y-axis from $y=+5a$ to $y=-5a$. (b) Compare this plot with the plot in Problem 1.24(d). (c) Is the potential a minimum at the origin?

1.27 In a rectangular coordinate system a charge of 25×10^{-9} C is placed at the origin of coordinates, and a charge of -25×10^{-9} C is placed at the point $x=6$ m, $y=0$. What is the electric field (a) at $x = 3$ m, $y=0$? (b) at $x=3$ m, $y=4$ m?

1.28 Equal electric charges of 1 μC each are placed at the vertices of an equilateral triangle whose sides are 0.1 m in length. Calculate (a) the force and the potential energy of each charge as a result of the interactions with the others, (b) the resultant electric field and electric potential at the center of the triangle, and (c) the internal potential energy of the system.

1.29 (a) Referring to Problem 1.28, make a plot of the lines of force of the electric field produced by the three charges. (b) Also on the same diagram plot the equipotential surfaces.

1.30 What is the final velocity of an electron accelerated through a potential difference of 12,000 V if the electron has an initial velocity of 10^7 m s^{-1}?

1.31 A potential difference of 1600 V is established between two parallel plates 4×10^{-2} m apart. An electron is released from the negative plate at the same instant that a proton is released from the positive plate. (a) How far from the positive plate will the electron and the proton be when they pass each other? (b) How do the velocities of the electron and the proton compare when they strike the opposite plates? (c) How do the energies of the electron and the proton compare when they strike the opposite plates?

1.32 A linear accelerator having a voltage difference of 800 kV produces a proton beam having a current of 1 mA. Calculate (a) the number of protons that strike the target per second, (b) the power required to accelerate the protons, and (c) the velocity of the protons when they hit the target. (d) Given that the protons lose 80 per cent of their energy in the target, calculate the rate, expressed in cal s^{-1}, at which energy in the form of heat must be removed from the target.

1.33 An electron, after being accelerated by a potential difference of 565 V, enters a uniform electric field of 3500 N C^{-1} at an angle of 60° with the direction of the field. After 5×10^{-8} s, what are (a) the components of the electron's velocity parallel and perpendicular to the field, (b) the magnitude and direction of the velocity of the electron, and (c) its coordinates relative to the point of entry? (d) What is the electron's total energy?

1.34 Two large plane metal plates are mounted vertically 4×10^{-2} m apart and charged to a potential difference of 200 V. (a) With what velocity must an electron be projected horizontally from the positive plate so that the electron will arrive at the negative plate with a velocity of 10^7 m s^{-1}? (b) With what velocity must the electron be projected from the positive plate in a direction at an angle of 37° above the horizontal so that the horizontal component of the velocity of the electron when it arrives at the negative plate is 10^7 m s^{-1}? (c) What is the magnitude of the vertical component of the velocity when the electron arrives at the negative plate? (d) What is the electron's time of transit from one plate to the other in each case? (e) With what velocity will the electron arrive at the negative plate if the electron is projected horizontally from the positive plate with a speed of 10^6 m s^{-1}?

1.35 (a) Estimate the average electric force of attraction between two water molecules in the gaseous phase at STP because of their electric dipole moments. Consider several possible relative orientations of their electric dipoles. (b) Compare this attraction with their gravitational attraction. The electric dipole moment of H_2O is 6.2×10^{-30} Cm. [Recall Ex. 1.7.]

1.36 (a) Relative to the Z-axis, what are the electric dipole and quadrupole moments of the charge distribution shown in Fig. 1-40? (b) Find the electric potential and the electric field at points along the Z-axis if z is very large compared with a. (c) Repeat the calculation for the Y-axis.

1.37 Assuming that all charges are positive, repeat Problem 1.36.

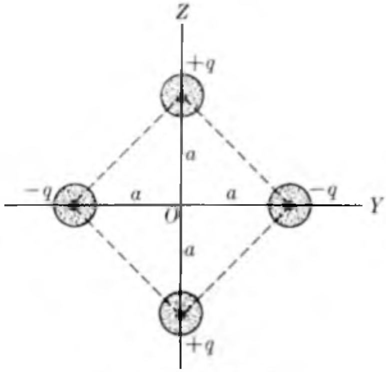


Figure 1-40

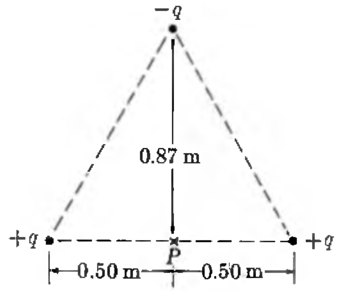


Figure 1-41

CHALLENGING PROBLEMS

1.38 Three charges of equal magnitude with signs as indicated in Fig. 1-41 are located at the corners of an equilateral triangle of side equal to 1.0 meter. The magnitude of each charge is 1.0×10^{-6} coulomb. Point P is midway between the two positive charges. (a) Determine the magnitude and direction of the electric field $\mathscr{E}$ at point P. (d) Determine the electric potential at point P. (AP-B; 1970)

1.39 A particle of charge $-q$ and mass m is released from rest at a distance $3R$ from a fixed charge $+Q$ (Fig. 1-42). (a) Determine the change in potential energy as the separation between the charges changes from $3R$ to $2R$. (b) Determine the speed of the particle when the separation is $2R$. (AP-B; 1970)

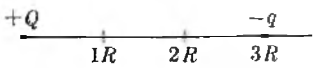


Figure 1-42

1.40 An electric car of mass 800 kilograms can climb a hill 1,000 meters long and 60 meters high in 100 seconds at constant speed (Fig. 1-43). The car is operated by a 48-volt battery. Neglecting friction and assuming 100 percent use of electrical energy, find the current delivered by the battery. (AP-B; 1970)

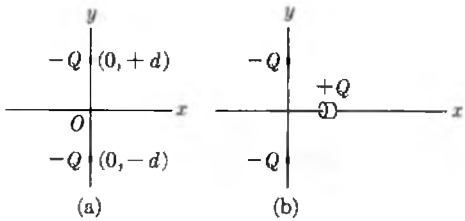


Figure 1-44

1.41 Two equal negative point charges $-Q$ are fixed at coordinates $(0, +d)$ and $(0, -d)$ on the x, y coordinate system shown in Fig. 1-44a. (a) Derive an expression for the x component $\mathscr{E}_x$ of the resultant electric field produced by these two charges at any point on the x-axis. (b) In terms of the quantity d, find the x coordinates of the points on the x-axis where the magnitude of $\mathscr{E}_x$ is a maximum. A small positively charged bead (dimensions much less

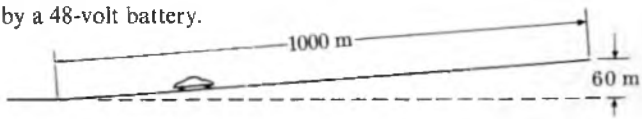


Figure 1-43

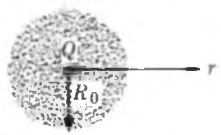


Figure 1-45

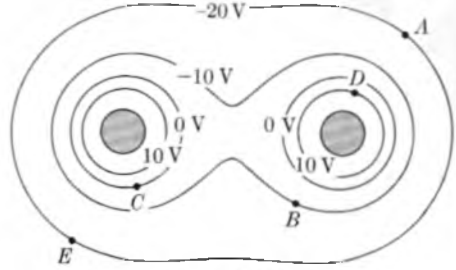


Figure 1-46

than d) is placed on a thin thread stretched along the x-axis as shown in Fig. 1-44b. (c) Show that the force acting on the bead obeys the approximate relation $F_x = -Kx$ for small displacements x from its equilibrium position, where K is a constant. Evaluate K in terms of Q and d. (AP-C: 1970)

1.42 Consider a uniform spherical positive charge distribution of radius R_0 and total charge Q as shown in Fig. 1-45. (a) Determine the electric field $\mathscr{E}$ of this charge distribution as a function of the distance r from the center of the sphere for $r > R_0$ and for $r < R_0$. (b) Determine the electric potential $V(r)$ of this charge distribution for $r > R_0$ and for $r < R_0$. Choose $V(r) = 0$ when r is infinite. (c) A positive point charge q is fired from a great distance toward the center of the distribution. When q is far from the distribution its kinetic energy is K_0. Depending on the value of K_0 the charge q will either pass through the center or be reflected by the distribution. Determine the minimum value of K_0 such that the charge q will pass through the distribution. (AP-C: 1970)

1.43 The diagram in Fig. 1-46 shows some of the equipotentials in a plane perpendicular to two parallel charged metal cylinders. The potential of each line is labeled. (a) The left cylinder is charged positively. What is the sign of the charge on the other cylinder? (b) On a copy of the diagram, sketch lines to describe the electric field produced by the charged cylinders. (c) Determine the potential difference, $V_A - V_B$, between points A and B. (d) How much work is done by the field if a charge of 0.50 coulomb is moved along a path from point A to point E and then to point D? (AP-B: 1974)

1.44 Two identical electric charges $+Q$ are located at two corners A and B of an isosceles triangle as shown in Fig. 1-47. (a) How much work does the electric field do on a small test charge $+q$ as the charge moves from point C to infinity? (b) In terms of the given quantities, determine where a third charge $+2Q$ should be placed so that the electric field at point C is zero. Indicate the location of this charge on the diagram. (AP-B: 1975)

1.45 Two stationary point charges $+q$ are located on the y-axis as shown in Fig. 1-48. A third charge $+q$ is brought in from infinity along the x-axis. (a) Express the potential energy of the movable charge as a function of

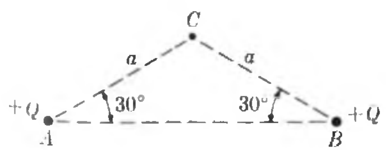


Figure 1-47

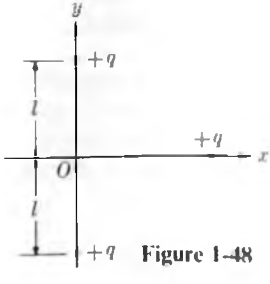


Figure 1-48

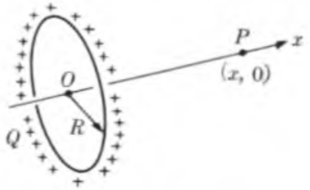


Figure 1-49

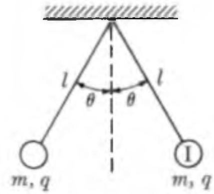


Figure 1-50

its position on the x-axis. (b) Determine the magnitude and direction of the force acting on the movable charge when it is located at the position $x=l$. (c) Determine the work done by the electric field as the charge moves from infinity to the origin. (AP-C: 1975)

1.46 A charge $+Q$ is uniformly distributed around a wire ring of radius R as shown in Fig. 1-49. Assume that the electric potential is zero at $x=$infinity with the origin O of the x-axis at the center of the ring. (a) What is the electric potential at a point P on the x-axis? (b) Where along the x-axis is the electric potential the greatest? Justify your answer. (c) What is the magnitude and direction of the electric field $\mathscr{E}$ at point P? (d) On a set of axes, make a sketch of $|\mathscr{E}|$ as a function of the distance along the x-axis showing significant features. (AP-C; 1977)

1.47 Two small spheres, each of mass m and positive charge q, hang from light threads of lengths l. Each thread makes an angle θ with the vertical as shown in Fig. 1-50. (a) On a diagram draw and label all forces on sphere I. (b) Develop an expression for the charge q in terms of m, l, θ, g, and the Coulomb's law constant. (AP-B: 1979)

1.48 An unstable nucleus, initially at rest, decays into two spherical fragments. Fragment A has radius r_A, charge q_A, and rest mass m_A; fragment B has radius r_B, charge q_B, and rest mass m_B. (a) Suppose that after the decay process fragment A has velocity v_A that is small compared with the velocity of light. Develop an expression for the kinetic energy of fragment B in terms of v_A, m_A, and m_B. At one stage in the decay process, the two fragments may be regarded as touching spheres, both at rest as shown in Fig. 1-51. Write expressions for each of the following. (b) The electrostatic force between these two fragments at the instant they are touching. (c) The electrostatic potential energy U of these two fragments at the instant they are touching. (d) The rest mass m of the nucleus before it decays. (AP-B: 1980)

1.49 Assuming that the two strings hang from points separated a distance d (Fig. 1-52), repeat Problem 1.4. How could this arrangement be used to verify experimentally the inverse-square law by varying the distance d and observing the angle θ?

1.50 In a particular fission of a uranium nucleus, the two fragments are ^{95}Y and ^{141}I, having masses practically equal to 95 amu and 141 amu, respectively. Their radii can be computed using the expression

$$R=1.2\times10^{-15}\,A^{1/3}\ \mathrm{m}$$

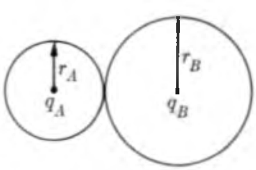


Figure 1-51

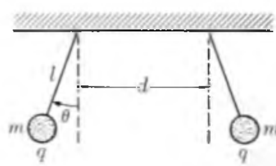


Figure 1-52

where A is the mass number. Assuming that the two fragments are initially at rest and tangent to each other, find (a) the initial force and potential energy, (b) their final relative velocity, and (c) the final velocity of each fragment relative to their center of mass.

1.51 Four protons are placed at the vertices of a square of side 2×10^{-9} m. Another proton is initially on the perpendicular to the square through its center and is a distance 2×10^{-9} m from the center. Calculate (a) the minimum initial velocity the fifth proton needs to reach the square, and (b) the initial and final acceleration. (c) Plot the potential energy of the proton as a function of its distance from the center of the square. (d) Describe the motion of the proton if the initial energy is either smaller or larger than that found in (a).

1.52 Consider the charges in Problem 1.24. (a) Suppose that a positively charged particle of charge q' is placed precisely at the origin and released from rest. What happens? (b) What will happen if the charge in part (a) is displaced slightly in the direction of the Y-axis? (c) What will happen if the charge is displaced slightly in the direction of the X-axis?

1.53 Show that the rectangular components of the electric field produced by a charge q at the distance r are

$$\mathscr{E}_x = \frac{qx}{4\pi\epsilon_0 r^3},$$

etc.

1.54 Establish a numerical relation giving the velocity (in m s^{-1}) of an electron and a proton in terms of the potential difference (in volts) through which they have moved. Assume that initially they were at rest.

1.55 (a) What is the maximum potential difference through which an electron can be accelerated if its mass is not to exceed its rest mass by more than 1 percent of the rest mass? (b) Express the velocity of such an electron as a fraction of the speed of light c. (c) Make the same calculations for a proton.

1.56 A certain high-energy machine accelerates electrons through a potential difference of 6.5×10^9 V. (a) What is the ratio of the mass m of an electron to its rest mass m_0 when the electron emerges from the accelerator? (b) What is the ratio of the electron's velocity to that of light? (c) What would the velocity be if computed from the principles of classical mechanics?

1.57 An electron in a certain X-ray tube is accelerated from rest through a potential difference of 180,000 V in going from the cathode to the anode. When the electron arrives at the anode, what is (a) the electron's kinetic energy in eV, (b) its mass m, and (c) its velocity?

1.58 Suppose that the potential difference between the spherical terminal of a Van de Graaff generator and the point at which charges are sprayed onto the upward-moving belt is 2×10^6 V. If the belt delivers negative charge to the sphere at the rate of 2×10^{-3} C s^{-1} and removes positive charge at the same rate, what power must be expended to drive the belt against electrical forces?

1.59 The average separation of protons within an atomic nucleus is of the order of 10^{-15} m. Estimate in J and in MeV the order of magnitude of the electric potential energy of two protons in a nucleus.

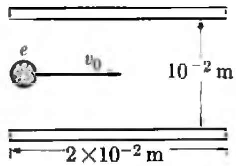


Figure 1-53

1.60 The potential difference between the two parallel plates in Fig. 1-53 is 100 V, their separation is 10^{-2} m, and their length is 2×10^{-2} m. An electron is projected with an initial velocity of 10^7 m s^{-1} in a direction perpendicular to the field. (a) Find the transverse deviation and the transverse velocity of the electron when it emerges from the plate. (b) If a screen is placed at 0.50 m to the right of the end of the plates, at what position on the screen will the electron fall?

1.61 A certain vacuum triode consists basically of the following elements. A plane surface (the

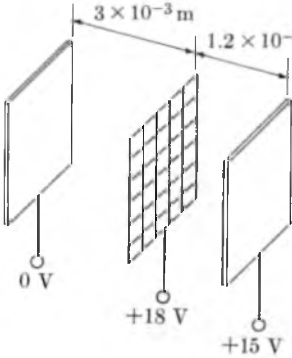


Figure 1-54

cathode) emits electrons with negligible initial velocities. Parallel to the cathode and 3×10^{-3} m away from it is an open grid of fine wire at an electric potential of 18 V with respect to the cathode. The structure of the grid is sufficiently open for electrons to pass through it freely. A second plane surface (the anode) is 1.2×10^{-2} m beyond the grid and is at an electric potential of 15 V with respect to the cathode. (See Fig. 1-54).) We may assume that the electric fields between cathode and grid and between grid and anode are uniform. (a) Draw a diagram of electric potential versus distance, along a line from cathode to anode. (b) With what velocity do the electrons cross the grid? (c) With what velocity do electrons strike the anode? (d) Determine the magnitude and the direction of the electric field between the cathode and the grid and between the grid and the anode. (e) Calculate the magnitude and the direction of the acceleration of the electron in each region.

1.62 An electron is between two horizontal plates separated 2×10^{-2} m and charged with a potential difference of 2000 V. (a) Compare the electric force on the electron with the force caused by gravity. (b) Repeat for a proton. (c) Does this justify having ignored gravitational effects in this chapter?

1.63 Consider a plane carrying a uniform charge density σ. Show that the electric field and potential are

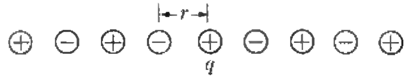


Figure 1-55

$$\mathcal{E} = \frac{\sigma}{2\epsilon_0} \qquad \text{and} \qquad V = -\frac{\sigma z}{2\epsilon_0}.$$

where y is the distance from one plate to the point in question.

1.64 Along a straight line there is an infinite number of alternating positive and negative charges $\pm q$, all adjacent charges being separated the same distance r (Fig. 1-55). Show that the electric potential energy of one charge is $(-q^2/2\pi\epsilon_0 r)\ln 2$.

1.65 A regular plane arrangement of alternate positive and negative charges of the same magnitude is obtained by placing the charges at the center of squares of side a (Fig. 1-56). Find the potential energy of a charge such as A. [*Hint:* Group the charges surrounding A; consider at one time all charges at the same distance from A.]

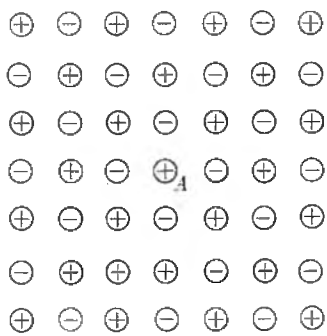


Figure 1-56

1.66 Find the electric potential and field along the points on the axis of a disk having a radius R and a charge σ per unit area. [*Hint:* Divide the disk into rings and add the contributions of all rings; see Problem 1.46.]

1.67 Referring to Problem 1.66, obtain the electric field and potential of a plane distribution of charge having the same charge density as the disk. [*Hint:* Make R very large and keep only the dominant term.]

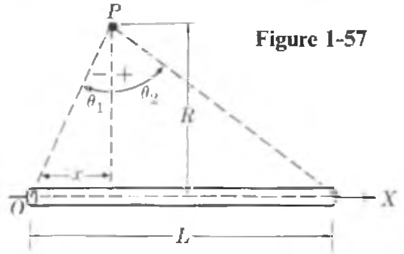


Figure 1-57

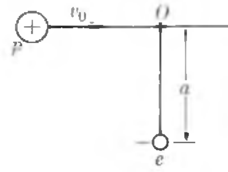


Figure 1-58

1.68 A wire of length L carries a uniform charge density λ per unit length (Fig. 1-57). (a) Show that the electric field at a point a distance from the wire is given by

$$\mathscr{E}_\perp = \frac{\lambda}{4\pi\epsilon_0 R}(\sin\theta_2 - \sin\theta_1)$$

and

$$\mathscr{E}_\parallel = -\frac{\lambda}{4\pi\epsilon_0 R}(\cos\theta_2 - \cos\theta_1)$$

where $\mathscr{E}_\perp$ and $\mathscr{E}_\parallel$ are the components of $\mathscr{E}$ perpendicular and parallel to the wire and θ_1 and θ_2 are the angles that the lines from the point to the ends of the wire make with the perpendicular to the wire. (b) Find the field when the point is equidistant from both ends. The signs of angles θ_1 and θ_2 are as indicated in the figure.

1.69 A wire carrying a uniform charge density λ per unit length is bent in the form of a square of side L. Find the electric field and electric potential at points on the line perpendicular to the square and passing through the center.

1.70 Obtain an expression for the electric field and electric potential of a plane carrying a uniform charge per unit area equal to σ if the plane is composed of a series of filaments of infinite length and width dx.

1.71 A very fast proton with velocity v_0 passes at a distance a from an electron initially at rest (Fig. 1-58). Assume that the motion of the proton is undisturbed because of its larger mass. (a) Plot as a function of x the component of the force perpendicular to v_0 that the proton exerts on the electron. (b) Show that the momentum transferred to the electron is

$$\left(\frac{e^2}{4\pi\epsilon_0}\right)\left(\frac{2}{v_0 a}\right)$$

in a direction perpendicular to v_0. (c) Estimate the deflection of the proton as a function if its velocity. This example provides a crude basis for analyzing the motion of charged particles passing through matter. [*Hint:* If one assumes that the electron practically remains at its initial position during the passage of the proton, the momentum transferred to the electron is given by $\Delta \boldsymbol{p} = \int \boldsymbol{F}\, dt$, and only the component perpendicular to v_0 needs to be computed. Instead of integrating from $-\infty$ to $+\infty$, in view of the symmetry of the force, integrate from 0 to ∞ and multiply by 2.]

1.72 Prove that the internal electric potential energy of a system of charges can be written in either of these alternate forms:

$$\text{(a) } E_p = \sum_{\substack{\text{All}\\\text{pairs}}} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}}, \qquad \text{(b) } E_p = \tfrac{1}{2}\sum_{\substack{\text{All}\\\text{charges}}} q_i V_i$$

where V_i is the potential produced at q_i by *all other* charges. (c) Using the result of (b), show that the electrical energy of a continuous charge distribution of density ρ is $E_p = \frac{1}{2}\int \rho V\, d\tau$. (d) Use this expression to show that the electric potential energy of a spherical conductor having a charge Q uniformly distributed over its volume is $\frac{3}{5}Q^2/4\pi\epsilon_0 R$. (e) Extend the last result to the case of a nucleus of atomic number Z.

1.73 Prove that the differential equations of the lines of force are

$$\frac{dx}{\mathscr{E}_x} = \frac{dy}{\mathscr{E}_y} = \frac{dz}{\mathscr{E}_z}$$

where dx, dy, and dz correspond to two very close points on the line of force. Apply these equations to obtain the equation of the lines of force of an electric dipole. [*Hint:* Note that since in this case the lines of force are plane curves, the component $\mathscr{E}_z$ is not required. Express $\mathscr{E}_x$ and $\mathscr{E}_y$ for an electric dipole in rectangular coordinates.]

CHAPTER TWO

STATIC ELECTRIC FIELD

2.1 Introduction

In the preceding chapter the concept of the electric field was introduced. In this chapter the characteristics of this field will be discussed under the assumption that the field is static or time independent. We shall then examine electric circuits before we begin an investigation of the static magnetic field. In subsequent chapters the time-dependent electromagnetic field will be considered.

2.2 Flux of a Vector Field

The *flux of a vector field* is a concept of great usefulness in many physical problems and will appear many times in this and succeeding chapters. Consider a surface S placed in a region in which there is a vector field V (Fig. 2-1). Divide the surface into very small (or infinitesimal) surfaces of areas dS_1, dS_2, $dS_3, \ldots$, and draw at each of them a unit vector $\boldsymbol{u}_1$, $\boldsymbol{u}_2$, $\boldsymbol{u}_3, \ldots$ perpendicular to the surface at that point. The unit vectors are oriented in the direction given by the thumb of the right hand when the fingers are curled in the sense in which we decide to orient the rim of the surface. Let θ_1, θ_2, $\theta_3, \ldots$ be the angles between the normal vectors $\boldsymbol{u}_1$, $\boldsymbol{u}_2$, $\boldsymbol{u}_3, \ldots$ and the field vectors V_1, V_2, $V_3, \ldots$ at each point on the surface. Then by definition the *flux* Φ of the vector field V through the surface S is

$$\begin{aligned}\Phi &= V_1\, dS_1 \cos\theta_1 + V_2\, dS_2 \cos\theta_2 + V_3\, dS_3 \cos\theta_3 + \cdots \\ &= \boldsymbol{V}_1\cdot\boldsymbol{u}_1\, dS_1 + \boldsymbol{V}_2\cdot\boldsymbol{u}_2\, dS_2 + \boldsymbol{V}_3\cdot\boldsymbol{u}_3\, dS_3 + \cdots \\ &= \sum_i \boldsymbol{V}_i\cdot\boldsymbol{u}_i\, dS_i,\end{aligned}$$

or

$$\Phi = \int_S V\cos\theta\, dS = \int_S \boldsymbol{V}\cdot\boldsymbol{u}_N\, dS \tag{2.1}$$

where the integral extends over all the surface as indicated by the subscript S. For that reason an expression like Eq. (2.1) is called a *surface integral*. Because of the $\cos\theta$ factor in Eq. (2.1), the flux through the surface element dS may be positive or negative, depending on whether θ is smaller or larger than $\pi/2$. If the field V is tangent or parallel to the surface element dS, the angle θ is $\pi/2$ and $\cos\theta = 0$, and the result is zero flux through dS. The total flux Φ may also be positive, negative, or zero. When it is positive, the flux is "outgoing"; and when it is negative, the flux is "incoming." If the surface is closed, such as in a sphere or an ellipsoid, a circle is written on top of the integral sign so that Eq. (2.1) becomes

$$\Phi = \oint_S V\cos\theta\, dS = \oint_S \boldsymbol{V}\cdot\boldsymbol{u}_N\, dS. \tag{2.2}$$

The name flux given to the integral in Eq. (2.1) is due to its application in the study of fluid flow. Suppose that there is a stream of particles, all moving with

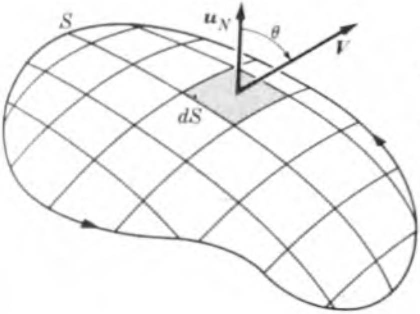


Fig. 2-1. Flux of a vector field through a surface.

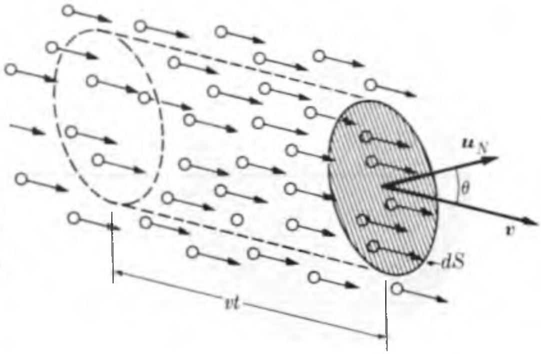


Fig. 2-2. Flux of particles through an area.

velocity $\boldsymbol{v}$ as shown in Fig. 2-2. Those particles passing through a surface dS in time t will be contained in an oblique cylinder whose base is dS, whose generatrix is parallel to $\boldsymbol{v}$, and whose length equals vt. This volume is $vt\, dS \cos\theta$. Given that there are n particles per unit volume, the total number of particles passing through dS in time t is $nvt\, dS \cos\theta$; and the number passing per unit time, or the *flux of particles*, is $nv\, dS \cos\theta = n\boldsymbol{v}\cdot\boldsymbol{u}_N\, dS$. The total flux of particles through a surface S is then

$$\Phi = \int_S n\boldsymbol{v}\cdot\boldsymbol{u}_N\, dS.$$

This is an expression similar to Eq. (2.1), with the vector field V equal to $n\boldsymbol{v}$. It must be realized, however, that the name "flux" as applied to Eq. (2.1) does not in general mean the actual motion of something through a surface.

Example 2.1. Expression of an electric current through a surface as a flux of a current density.

▼ The quantity $n\boldsymbol{v}\cdot\boldsymbol{u}_N\, dS$ expresses the number of particles passing through the surface dS per unit time. If each particle carries a charge q, the charge passing through the surface dS per unit time (that is, the current $dI = dQ/dt$) is

$$qn\boldsymbol{v}\cdot\boldsymbol{u}_N\, dS = \boldsymbol{j}\cdot\boldsymbol{u}_N\, dS$$

where $\boldsymbol{j} = nq\boldsymbol{v}$ is called the *current density* having the units of $\mathrm{A\,m^{-3}}$. Therefore the total charge passing through a surface S per unit time (i.e., the electric current I through the surface) is

$$I = \int_S \boldsymbol{j}\cdot\boldsymbol{u}_N\, dS.$$

In other words the electric current through a surface is equal to the flux of the electric current density through that surface. If the current density is uniform and the surface is plane, the equation reduces to

$$I = \boldsymbol{j}\cdot\boldsymbol{u}_N S = jS\cos\theta. \quad ▲$$

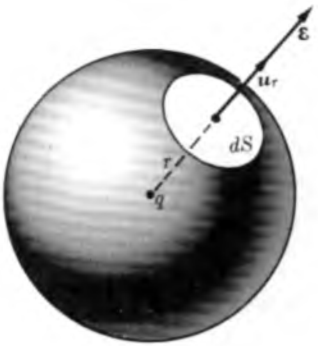


Fig. 2-3. Electric flux of a point charge through a sphere.

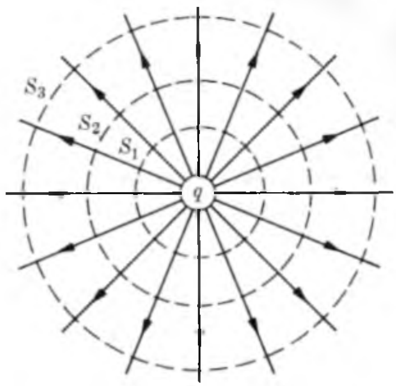


Fig. 2-4. The electric flux through concentric spheres surrounding the same charge is the same.

2.3 Law for the Electric Field

Consider a point charge q (Fig. 2-3). The flux of the electric field $\mathscr{E}$ of the charge through a spherical surface concentric with the charge may be computed using Eq. (2.2). Given that r is the radius of the sphere, the electric field produced by the charge at each point of the spherical surface is

$$\mathscr{E} = \frac{q}{4\pi\epsilon_0 r^2}\boldsymbol{u}_r.$$

The unit vector normal to a sphere coincides with the unit vector $\boldsymbol{u}_r$ along the radial direction. Therefore the angle θ between the electric field $\mathscr{E}$ and the normal unit vector $\boldsymbol{u}_r$ is zero, and $\cos\theta = 1$. Because the electric field has the same magnitude at all points of the spherical surface and the area of the sphere is $4\pi r^2$, it is easy to see that Eq. (2.2) gives the electric flux $\Phi_{\mathscr{E}}$ as

$$\Phi_{\mathscr{E}} = \oint_S \mathscr{E}\, dS = \mathscr{E}\oint_S dS = \mathscr{E}S = \frac{q}{4\pi\epsilon_0 r^2}(4\pi r^2) = \frac{q}{\epsilon_0}.$$

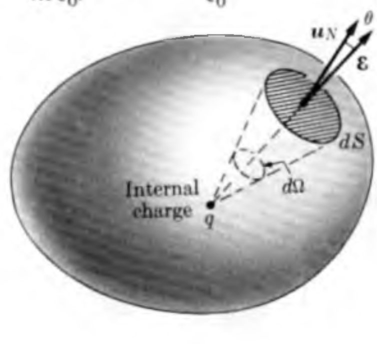


Fig. 2-5. The electric flux through a closed surface surrounding a charge is independent of the shape of the surface.

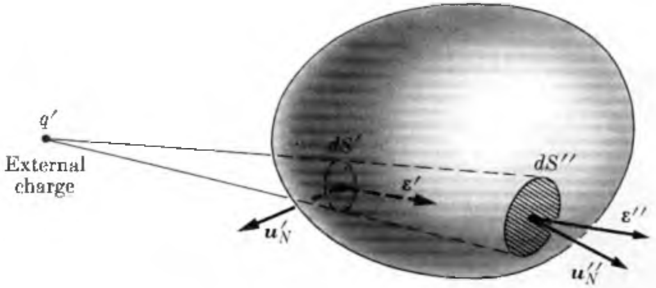


Fig. 2-6. The electric flux through a closed surface due to an external charge is zero.

The electric flux through the sphere, then, is proportional to the charge and independent of the radius of the surface. Therefore, for each concentric spherical surface S_1, S_2, S_3, . . . (Fig. 2-4) around the charge q, the electric flux through each is the same and equal to q/ϵ_0. The result is due to the $1/r^2$ dependence of the field.

Next consider a charge q *inside* an arbitrary *closed* surface S (Fig. 2-5). The total flux through S of the electric field produced by q is given by

$$\Phi_{\mathscr{E}}=\oint_S \mathscr{E}\cos\theta\, dS=\oint_S \frac{q}{4\pi\epsilon_0 r^2}\cos\theta\, dS=\frac{q}{4\pi\epsilon_0}\oint_S \frac{dS\cos\theta}{r^2}.$$

However, $dS\cos\theta/r^2$ is the solid angle $d\Omega$ subtended by the surface element dS as viewed from the charge q. Since the total solid angle around any point is 4π steradians,* then

$$\Phi_{\mathscr{E}}=\frac{q}{4\pi\epsilon_0}\int d\Omega=\frac{q}{4\pi\epsilon_0}(4\pi)=\frac{q}{\epsilon_0}.$$

This result is the same as the previous result for a spherical surface concentric with the charge, and thus is valid for any closed surface, irrespective of the position of the charge within the surface.

If a charge such as q' is *outside* a closed surface (Fig. 2-6), the electric flux is zero because the incoming flux is equal to the outgoing flux; hence a net flux of zero results. For example the electric flux of q' through dS' is equal in magnitude, but opposite in sign, to the electric flux through dS''; and therefore they add to zero. This result is true because even though dS'' is larger than dS', $\mathscr{E}''$ is less than $\mathscr{E}'$. As may be seen from the figure, the subtended solid angle is exactly the same at both surfaces.

*See the appendix for a discussion of plane and solid angles.

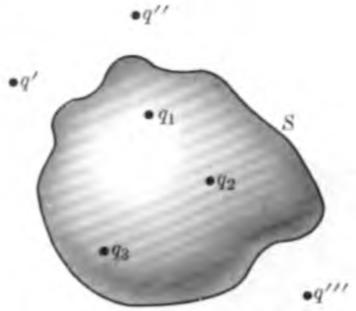


Fig. 2-7. The electric flux through any closed surface is proportional to the net charge contained within the surface.

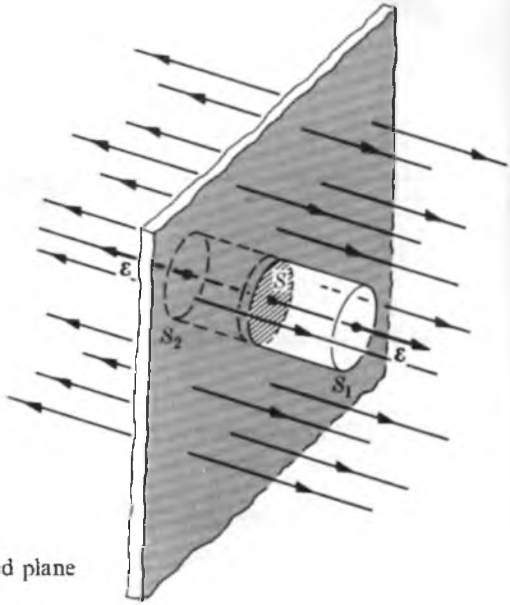


Fig. 2-8. Electric field of a uniformly charged plane surface.

If there are several charges $q_1, q_2, q_3, \ldots$ inside a closed surface S (Fig. 2-7), the total electric flux will be the sum of the fluxes produced by each charge. *Gauss's law* may then be stated:

The electric flux through a closed surface surrounding charges $q_1, q_2, q_3, \ldots$ is

$$\Phi_{\mathscr{E}} = \oint_S \mathscr{E} \cdot \boldsymbol{u}_N \, dS = \frac{q}{\epsilon_0} \tag{2.3}$$

where $q = q_1 + q_2 + q_3 + \cdots$ is the total net charge inside the closed surface.

The law is named to honor the German mathematician Karl F. Gauss (1777–1855), who originally developed the mathematical theory of electromagnetism. If no charges are present *inside* the closed surface, or if the net charge is zero, the total electric flux through the surface is zero. The charges, such as $q', q'', \ldots$, *outside* the closed surface do not contribute to the total flux.

Gauss's law is particularly useful to compute the electric field produced by charge distributions having certain geometrical symmetries as shown in the following examples.

Example 2.2. Using Gauss's law, discuss the electric field of a charge uniformly distributed over a plane.

▼ Consider Fig. 2-8: a charge σ per unit area has been placed on the plane. The symmetry of the problem indicates that if the plane is infinite in extent, the lines of force are perpendicular to the plane; and if the charge is positive they are oriented as indicated in the figure. For the chosen closed surface, the cylinder shown in the figure, the electric flux $\Phi_{\mathscr{E}}$ may be separated into three terms: the flux through S_1 is given by $+\mathscr{E}S$ where S is the area of the base of the cylinder; the flux through S_2 is also given by $+\mathscr{E}S$ since by symmetry the electric field must be the same in magnitude and opposite in direction at points at the same distance on both sides of the plane; and the flux through the lateral surface of the cylinder is zero because the electric field is parallel to the surface. Therefore the total electric flux is $\Phi_{\mathscr{E}}=2\mathscr{E}S$. The charge inside the closed surface is that in the shaded area and is equal to $q=\sigma S$. Therefore applying Gauss's law, Eq. (2.3), gives $2\mathscr{E}S=\sigma S/\epsilon_0$, or

$$\mathscr{E}=\frac{\sigma}{2\epsilon_0}.$$

This result indicates that the electric field is independent of the distance to the plane and is therefore uniform. Using the relation $\mathscr{E}=-dV/dx$ and assuming that the potential of the plane is zero, the electric potential is then

$$V=-\frac{\sigma}{2\epsilon_0}x. \quad ▲$$

Example 2.3. Using Gauss's law, discuss the electric field of two parallel planes with equal but opposite uniformly distributed charge densities.

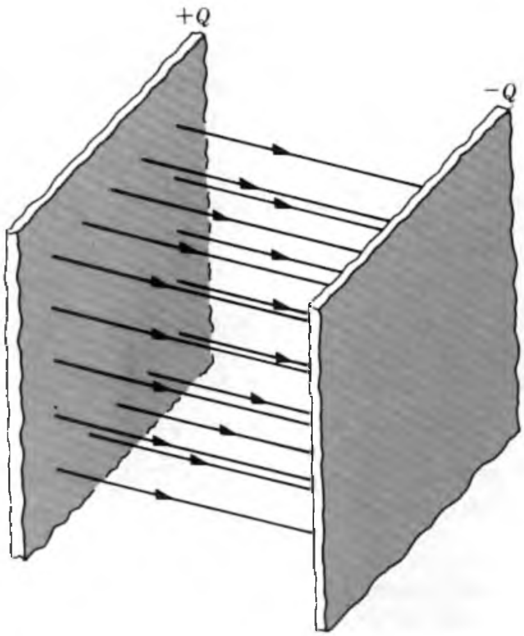


Fig. 2-9. Electric field between a pair of plane parallel surfaces carrying equal but opposite charges.

▼ Figure 2-9 shows two parallel planes with equal but opposite charges. Observe that in the region outside the two oppositely charged planes there are electric fields equal in magnitude but opposite in direction and the resultant field is zero. However in the region between the planes the fields are in the same direction, and the resultant field is twice as large as the field of a single plane, or $\mathscr{E}=\sigma/\epsilon_0$. Thus the two parallel and oppositely charged planes produce a uniform field contained in the region between them. ▲

Example 2.4. Using Gauss's law, discuss the electric field of a spherical distribution of charge.

▼ This problem has already been discussed in a different manner in Volume I for the case of the gravitational field of a spherical body. Consider a sphere of radius a and charge Q (Fig. 2-10). The symmetry of the problem suggests that the field at each point must be radial and depends only on the distance r from the point to the center of the sphere. Therefore drawing a spherical surface of radius r concentric with the charged sphere shows that the electric flux through it is

$$\Phi_{\mathscr{E}}=\oint_S \mathscr{E}\, dS=\mathscr{E}\oint_S dS=\mathscr{E}(4\pi r^2).$$

Consider first $r>a$; the charge inside the surface S is the total charge Q of the sphere. Thus from Gauss's law, Eq. (2.3), $\mathscr{E}(4\pi r^2)=Q/\epsilon_0$, or

$$\mathscr{E}=\frac{Q}{4\pi\epsilon_0 r^2}.$$

This result is the same as that for the field of a point charge. Thus *the electric field at points outside a charged sphere is the same as if all the charge were concentrated at its center.*

Consider next $r<a$; there are two possibilities. If all the charge is at the surface of the charged sphere, the total charge inside the spherical surface S' is zero, and Gauss's law gives $\mathscr{E}(4\pi r^2)=0$ or $\mathscr{E}=0$. Thus *the electric field at points inside a sphere that is charged only on its surface is zero.* If the sphere is uniformly charged throughout its volume and Q' is the charge inside the surface S',

$$Q'=\frac{Q}{4\pi a^3/3}(4\pi r^3/3)=\frac{Qr^3}{a^3}.$$

Therefore Gauss's law now gives $\mathscr{E}(4\pi r^2)=Q'/\epsilon_0=Qr^3/\epsilon_0 a^3$, or

$$\mathscr{E}=\frac{Qr}{4\pi\epsilon_0 a^3};$$

this expression shows that *the electric field at a point inside a uniformly charged sphere is directly proportional to the distance from the point to the center of the sphere.* If γm is replaced by $Q/4\pi\epsilon_0$, these results agree with those for the gravitational case. ▲

Example 2.5. Using Gauss's law, discuss the electric field of a cylindrical charge distribution of infinite length.

▼ Consider a length L of the cylinder C, whose radius is a (Fig. 2-11). If λ is the charge per unit length, the total charge in that portion of the cylinder is $q=\lambda L$. The symmetry of the problem indicates that the electric field at a point depends only on the distance from the point to the axis

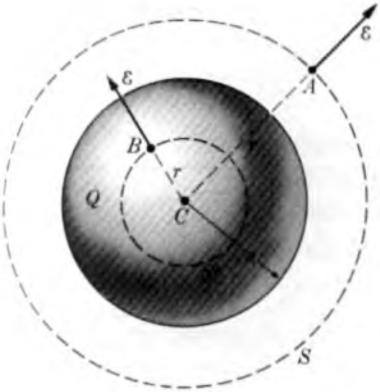


Figure 2-10

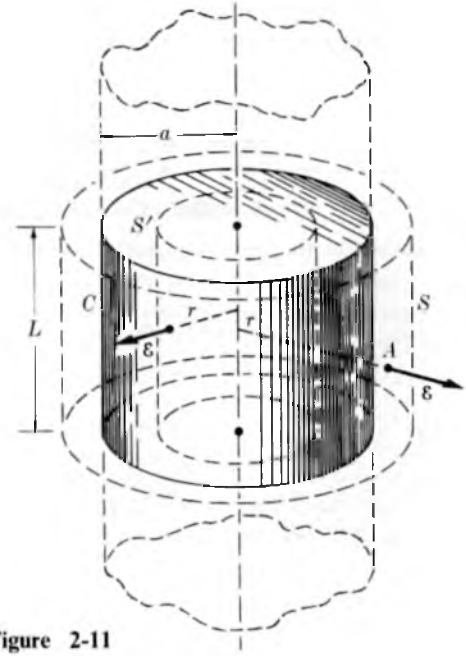


Figure 2-11

of the cylinder and is directed radially. Take as the closed surface of integration a cylindrical surface of radius r, coaxial with the charge distribution. Then the electric flux through that surface has three terms. Two terms represent the flux through each base; but they are zero because the electric field is tangent to each base. Thus the flux through the lateral surface is all that remains and gives $\mathscr{E}(2\pi rL)$. That is,

$$\Phi_{\mathscr{E}} = 2\pi rL\mathscr{E}.$$

If $r > a$, the total charge within the cylindrical surface S is $q = \lambda L$, and from Gauss's law, Eq. (2.3), we get $2\pi rL\mathscr{E} = \lambda L/\epsilon_0$, or

$$\mathscr{E} = \frac{\lambda}{2\pi\epsilon_0 r}.$$

This result agrees with that of Example 1.5 for the electric field of a charged filament. Therefore *the electric field at points external to a cylindrical charge distribution of infinite length is the same as if all the charge were concentrated along the axis.*

For $r < a$ there are again two possibilities as in the previous example. If all the charge is on the surface of the cylinder, there is no charge inside the surface S', and Gauss's law gives $2\pi rL\mathscr{E} = 0$ or $\mathscr{E} = 0$. Thus *the electric field at points inside a cylinder charged only on its surface is zero.* If the charge is distributed uniformly over the volume of the cylinder C, the charge within the surface S' is $q' = \lambda Lr^2/a^2$; and Gauss's law gives $2\pi rL\mathscr{E} = -q'/\epsilon_0$, or

$$\mathscr{E} = \frac{\lambda r}{2\pi\epsilon_0 a^2}.$$

Thus the electric field at a point within a uniformly charged cylinder of infinite length is proportional to the distance of the point from the axis. ▲

2.4 Gauss's Law in Differential Form

As shown in the previous section, Gauss's law can be applied to a closed surface of any shape. Consider now the closed surface surrounding an infinitesimal volume whose edges are parallel to the XYZ-axis as indicated in Fig. 2-12. The sides of the volume element are dx, dy, and dz. The area of the surface $ABCD$ is $dxdz$, and the electric flux through it is

$$\mathscr{E}\, dS \cos\theta = (\mathscr{E} \cos\theta)\, dy\, dz = \mathscr{E}_x\, dy\, dz$$

since $\mathscr{E}_x = \mathscr{E} \cos\theta$. The flux through the face $A'B'C'D'$ has a similar expression but is negative because the field is pointing into the volume; that is, $-\mathscr{E}'_x\, dy\, dz$. The total flux through these two faces is the sum

$$\mathscr{E}_x\, dy\, dz + (-\mathscr{E}'_x\, dy\, dz) = (\mathscr{E}_x - \mathscr{E}'_x)\, dy\, dz.$$

However since the distance $A'A = dx$ between the two surfaces is very small, the quantity $\mathscr{E}_x - \mathscr{E}'_x$ is also very small, and so the difference may be written as a differential

$$\mathscr{E}_x - \mathscr{E}'_x = d\mathscr{E}_x = \frac{\partial \mathscr{E}_x}{\partial x}\, dx$$

where $\partial\mathscr{E}_x/\partial x$ is the rate of change of the x-component of $\mathscr{E}$ in the X-direction. Thus the total flux in the X-direction is given by

$$\left(\frac{\partial \mathscr{E}_x}{\partial x}\, dx\right) dy\, dz = \frac{\partial \mathscr{E}_x}{\partial x}\, dV.$$

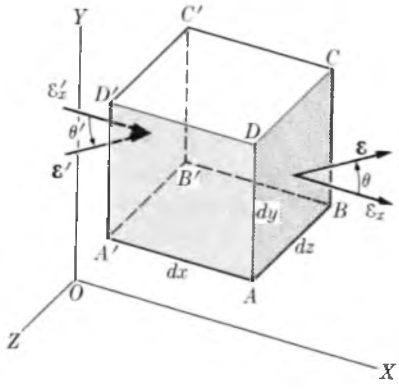


Fig. 2-12. Volume element to evaluate Gauss's law in differential form.

The quantity $dV = dx\,dy\,dz$ is the volume of the box. Since similar results are obtained for the flux through the remaining four faces of the volume element, the *total flux* through the volume element is

$$\Phi_{\mathscr{E}} = \frac{\partial \mathscr{E}_x}{\partial x}\,dV + \frac{\partial \mathscr{E}_y}{\partial y}\,dV + \frac{\partial \mathscr{E}_z}{\partial z}\,dV = \left(\frac{\partial \mathscr{E}_x}{\partial x} + \frac{\partial \mathscr{E}_y}{\partial y} + \frac{\partial \mathscr{E}_z}{\partial z}\right) dV.$$

If dq is the electric charge within the volume element, Gauss's law gives

$$\frac{\partial \mathscr{E}_x}{\partial x} + \frac{\partial \mathscr{E}_y}{\partial y} + \frac{\partial \mathscr{E}_z}{\partial z}\quad dV = \frac{dq}{\epsilon_0}.$$

Setting $dq = \rho\,dV$ in the above expression where ρ is the density of electric charge (or charge per unit volume), and canceling the common factor dV, we obtain

$$\frac{\partial \mathscr{E}_x}{\partial x} + \frac{\partial \mathscr{E}_y}{\partial y} + \frac{\partial \mathscr{E}_z}{\partial z} = \frac{\rho}{\epsilon_0}. \tag{2.4}$$

This equation is Gauss's law expressed in differential form. The expression on the left-hand side of Eq. (2.4) is called the *divergence of* $\mathscr{E}$, abbreviated div $\mathscr{E}$, so that Gauss's law can be written in the compact form

$$\text{div}\ \mathscr{E} = \frac{\rho}{\epsilon_0}. \tag{2.5}$$

The physical meaning of Gauss's law in its differential form is that it relates the electric field $\mathscr{E}$ at a point in space to the charge distribution, expressed by ρ, at the *same* point in space; that is, the law expresses a *local* relation between these two physical quantities. Thus it is correct to say that electric charges are the *sources* of the electric field, and that their distribution and magnitude determine the electric field at each point of space.

Example 2.6. Gauss's law expressed in terms of the electric potential.

▼ If we remember that the components of the electric field $\mathscr{E}$ are expressed in terms of the electric potential V by $\mathscr{E}_x = -\partial V/\partial x$ and similar expressions for $\mathscr{E}_y$ and $\mathscr{E}_z$ (see Eq. 1.15), the rate of change of $\mathscr{E}_x$ with respect to x may be written as

$$\frac{\partial \mathscr{E}_x}{\partial x} = \frac{\partial}{\partial x}\left(-\frac{\partial V}{\partial x}\right) = -\frac{\partial^2 V}{\partial x^2}$$

with similar results for $\mathscr{E}_y$ and $\mathscr{E}_z$. With the substitution in Eq. (2.4), an alternate expression for Gauss's law is

$$\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} + \frac{\partial^2 V}{\partial z^2} = -\frac{\rho}{\epsilon_0}, \tag{2.6}$$

an expression that is called *Poisson's equation*. Eq. (2.6) may be used to obtain the electric potential when the charge distribution is known and conversely, so long as the charge distribution is time independent. In free space where there are no charges, $\rho=0$, and Eq. (2.5) becomes div $\mathscr{E}=0$, and Eq. (2.6) gives

$$\frac{\partial^2 V}{\partial x^2}+\frac{\partial^2 V}{\partial y^2}+\frac{\partial^2 V}{\partial z^2}=0. \tag{2.7}$$

This equation is called *Laplace's equation.* It is one of the most important equations in mathematical physics, and appears in many problems, such as fluid motion and elasticity, outside the theory of the electromagnetic field.

The expression appearing on the left in Eqs. (2.6) and (2.7) is called the *laplacian of V.* ▲

Example 2.7. Verification that the potential of a point charge satisfies Laplace's equation, Eq. (2.7), at all points except at the origin where the charge is located.

▼ The potential of a point charge is $V=q/4\pi\epsilon_0 r$ according to Eq. (1.20). Now $r^2=x^2+y^2+z^2$ so that taking the partial derivative of r relative to x yields

$$2r\left(\frac{\partial r}{\partial x}\right)=2x \qquad \text{or} \qquad \frac{\partial r}{\partial x}=\frac{x}{r}.$$

Therefore with the value of $\partial x/\partial r$ known, the partial derivative of $(1/r)$ with respect to x is

$$\frac{\partial}{\partial x}\left(\frac{1}{r}\right)=-\frac{1}{r^2}\frac{\partial r}{\partial x}=-\frac{x}{r^3}.$$

Similarly,

$$\frac{\partial^2}{\partial x^2}\left(\frac{1}{r}\right)=\frac{\partial}{\partial x}\left(-\frac{x}{r^3}\right)=-\frac{1}{r^3}+\frac{3x}{r^4}\frac{\partial r}{\partial x}=-\frac{1}{r^3}+\frac{3x^2}{r^5}.$$

Then

$$\frac{\partial^2}{\partial x^2}\left(\frac{1}{r}\right)+\frac{\partial^2}{\partial y^2}\left(\frac{1}{r}\right)+\frac{\partial^2}{\partial z^2}\left(\frac{1}{r}\right)=-\frac{3}{r^3}+\frac{3(x^2+y^2+z^2)}{r^5}=0.$$

Multiplying this result by $q/4\pi\epsilon_0$ yields Eq. (2.7). This mathematical method is not valid for $r=0$ because the function $1/r$ goes to infinity at that point, and taking derivatives there is not allowed; therefore the origin must be excluded from the calculation. ▲

Example 2.8. Using Laplace's equation, obtain the electric potential and the electric field in the empty region between two infinite parallel planes charged to potentials V_1 and V_2.

▼ The symmetry of the problem suggests that the field must depend only on the x-coordinate (Fig. 2-13). Therefore since there are no charges in the space between the planes, we may apply Laplace's equation, Eq. (2.7), giving $d^2V/dx^2=0$. Note that the partial derivative notation is not used because there is only one independent variable, x. Integrating Laplace's equation once gives $dV/dx=$ const; but the electric field is $\mathscr{E}=-dV/dx$. Therefore the electric field between the planes is constant. Again integrating and using the expression $\mathscr{E}=-dV/dx$ (keeping in mind that $\mathscr{E}$ is constant), we get

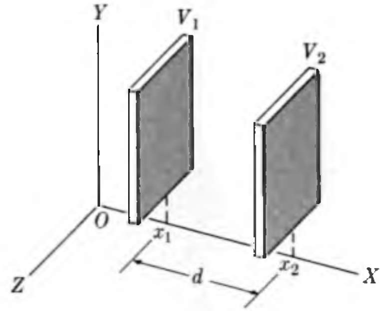


Figure 2-13

$$\int_{V_1}^{V} dV = -\int_{x_1}^{x} \mathscr{E}\, dx = -\mathscr{E}\int_{x_1}^{x} dx,$$

an equation that gives $V - V_1 = -\mathscr{E}(x - x_1)$ or $V = V_1 - \mathscr{E}(x - x_1)$.

This result shows that the electric potential varies linearly with the distance x. Setting $x = x_2$, we have $V = V_2$. Therefore

$$\mathscr{E} = -\frac{V_2 - V_1}{x_2 - x_1} = -\frac{V_2 - V_1}{d}.$$

These results are in agreement with our previous discussion (Section 1.6) leading to Eq. (1.19) since $d = x_2 - x_1$ is the separation of the planes. Comparing the results of this example with those of Example 2.3 shows that two uniformly charged planes produced the uniform electric field seen here. ▲

Example 2.9. Assuming that there is a uniform charge distribution *between* the planes, solve the same problem as in Example 2.8. The planes are still at electric potentials V_1 and V_2, respectively. (This may be, for example, the situation between the plates of a vacuum tube.)

▼ For this case Poisson's equation, Eq. (2.6), must be used. Because of the symmetry of the problem, the potential depends only on the coordinate x; $d^2V/dx^2 = -\rho/\epsilon_0$, with $\rho = \text{const}$. Integration gives

$$\int_{x_1}^{x} \frac{d^2V}{dx^2}\, dx = -\frac{1}{\epsilon_0}\int_{x_1}^{x} \rho\, dx = -\frac{\rho}{\epsilon_0}\int_{x_1}^{x} dx,$$

which results in

$$\frac{dV}{dx} - \left(\frac{dV}{dx}\right)_{x=x_1} = -\frac{\rho}{\epsilon_0}(x - x_1)$$

or

$$\frac{dV}{dx} = -\mathscr{E}_1 - \frac{\rho}{\epsilon_0}(x - x_1) \tag{2.8}$$

where $\mathscr{E}_1 = -(dV/dx)_{x=x_1}$ is the electric field at $x = x_1$. Since $\mathscr{E} = -dV/dx$, the field between the planes is

$$\mathscr{E} = \mathscr{E}_1 + \frac{\rho}{\epsilon_0}(x - x_1),$$

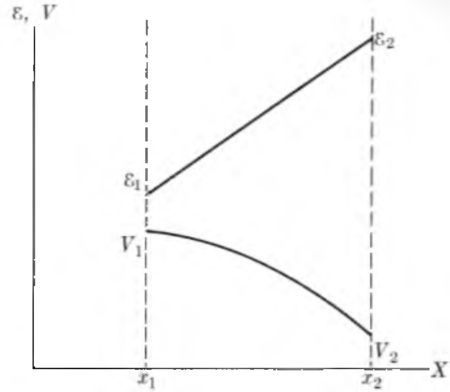


Figure 2-14

showing that the electric field varies linearly with x as illustrated in Fig. 2-14. After integration of Eq. (2.8) to obtain the electric potential as a function of x,

$$\int_{V_1}^{V} dV = -\int_{x_1}^{x} \mathscr{E}_1\, dx - \frac{\rho}{\epsilon_0}\int_{x_1}^{x} (x-x_1)\, dx,$$

we get

$$V = V_1 - \mathscr{E}_1(x-x_1) - \frac{\rho}{2\epsilon_0}(x-x_1)^2. \tag{2.9}$$

The electric potential varies quadratically with x as shown also in Fig. 2-14. ▲

2.5 The Polarization of Matter

Dielectrics

In this section the effect of an electric field on a piece of matter will be considered. Recall that atoms do not have permanent electric dipole moments because of their spherical symmetry; however when they are placed in an electric field, they become polarized, acquiring *induced* electric dipole moments in the direction of the field. This polarization results from the perturbation of the motion of the electrons produced by the applied electric field.

On the other hand, many molecules do have *permanent* electric dipole moments. A molecule with a permanent electric dipole moment tends to be oriented parallel to the applied electric field because of the torque it experiences [given by Eq. (1.36)]. As a consequence of either of these two effects, a piece of matter placed in an electric field becomes electrically *polarized*. That is, the molecules (or atoms) become electric dipoles oriented in the direction of the local electric field (Fig. 2-15) because of either the distortion of the electronic motion or the orientation of their permanent dipoles.

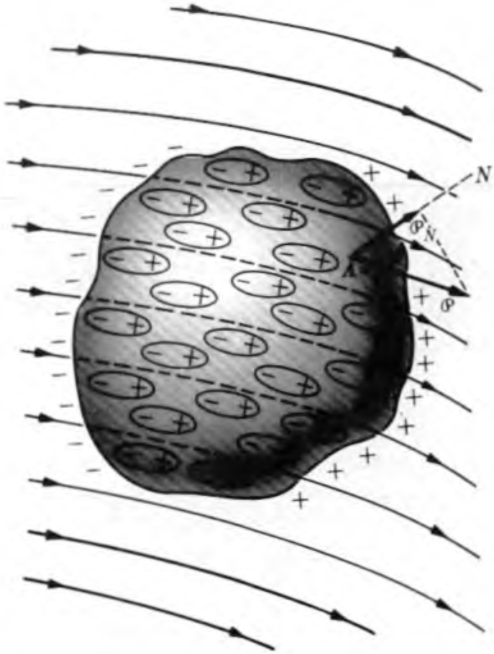


Fig. 2-15. Polarization of matter by an electric field.

A medium that can be polarized by an electric field is called a *dielectric*. The polarization gives rise to a net positive charge on one side of the piece of matter and a net negative charge on the opposite side. The piece of matter then becomes a large electric dipole that *tends* to move in the direction in which the field increases as discussed in Section 1.10. This tendency explains the phenomenon (described in Section 1.1) in which an electrified glass rod or a comb attracts small pieces of paper or a cork ball.

The *polarization* $\mathscr{P}$ of a material is defined as the electric dipole moment of the medium per unit volume. Therefore if $\boldsymbol{p}$ is the dipole moment induced in each atom or molecule and n is the number of atoms or molecules per unit volume, the polarization is $\mathscr{P}=n\boldsymbol{p}$. For most dielectrics $\mathscr{P}$ is proportional to the applied electric field $\mathscr{E}$. Since $\mathscr{P}$ is measured in $(\mathrm{C\ m})\mathrm{m}^{-3}=\mathrm{C\ m}^{-2}$, or charge per unit area, and since $\epsilon_0\mathscr{E}$ is also measured in $\mathrm{C\ m}^{-2}$ [cf. Eq. (1.8)], it is customary to write

$$\mathscr{P}=\chi_e\epsilon_0\mathscr{E}. \tag{2.10}$$

The quantity χ_e is called the *electric susceptibility* of the material. The electric susceptibility is a pure number and is a positive quantity for most substances. The electric susceptibility describes the response of the material to an external electric field.

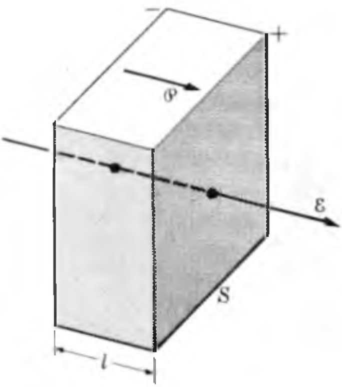


Fig. 2-16. A slab of polarized metal.

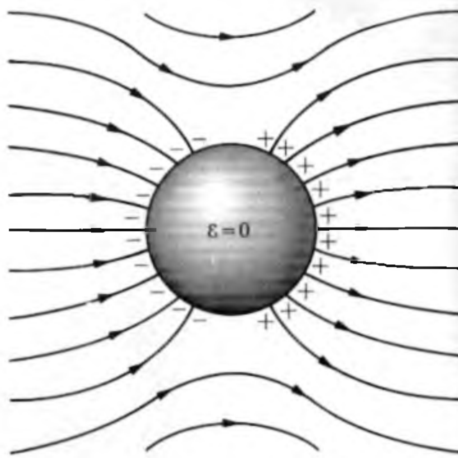


Fig. 2-17. The electric field within a conductor is zero.

Consider now a slab of material of thickness l and surface area S placed perpendicular to a uniform field $\mathscr{E}$ (Fig. 2-16). The polarization $\mathscr{P}$, being parallel to $\mathscr{E}$, is also perpendicular to S. The volume of the slab is lS, and therefore its total electric dipole moment by definition is $\mathscr{P}(lS)=(\mathscr{P}S)l$; but l is just the separation between the positive and negative charges that appear on the two surfaces. Since again by definition the electric dipole moment is equal to charge times distance, the total electric charge that appears on each of the surfaces is $\mathscr{P}S$, and therefore the charge per unit area on the faces of the polarized slab is $\mathscr{P}$. Although obtained for a particular geometrical arrangement, this result has general validity; and

> *the charge per unit area on the surface of polarized matter, $\sigma_{\mathscr{P}}$, is equal to the component of the polarization $\mathscr{P}$ in the direction of the normal to the surface of the body.*

Therefore in Fig. 2-15, the charge per unit area on the surface at A is $\mathscr{P}_N=\mathscr{P}\cos\theta=\sigma_{\mathscr{P}}$.

Conductors

Some materials, such as most metals, contain charged particles that can move more or less freely through the medium. These materials are called *conductors*. In the presence of an electric field they are also polarized but in a way that is essentially different from the polarization of dielectrics. Unless properly removed, the mobile charges in a conductor accumulate on the surface until the field they produce completely cancels the external applied field within the conductor and thereby produces equilibrium (Fig. 2-17). That is, *inside a conductor that is in electrical equilibrium, the electric field is zero*. For the same reason *the electric field at the surface must be normal to the surface* since if there is a parallel component, the charges will move along the surface of the conductor. Furthermore because the field inside the conductor is zero, *all points of a conductor that is in equilibrium must be at the same electrical potential*. Because the electric field inside the conductor is zero, it also follows that div $\mathscr{E}=0$; and therefore Gauss's law in differential form, Eq. (2.5), gives $\rho=0$; and thus the charge density within the volume of the conductor is zero. That statement means that *the entire electric charge of a conductor in equilibrium resides on the surface of the conductor*. "Surface" as used here refers not to an infinitesimally thin surface in the geometric sense, but rather to a surface region several atomic layers thick; i.e., a real finite surface.

Example 2.10. Relation between the electric field at the surface of a conductor and the surface electric charge.

▼ Consider a conductor of arbitrary shape as in Fig. 2-18. To find the electric field at a point immediately outside the surface of the conductor, we construct a flat cylindrical surface similar to a pillbox, with one base immediately outside the surface of the conductor and the other base at a depth such that all the surface charge is within the cylinder and the electric field is already zero at the inner surface. The electric flux through that surface is composed of three terms: the flux through the inner base is zero because the field is zero; the flux through the side is zero because the field is tangent to this surface: the flux through its outer base is all that remains. Given that the area of the base is S, the electric flux is $\Phi_{\mathscr{E}}=\mathscr{E}S$. On the other hand if σ is the surface charge density of the conductor, the charge within the cylinder is $q=\sigma S$. Therefore from Gauss's law $\mathscr{E}S=\sigma S/\epsilon_0$, or

$$\mathscr{E}=\sigma/\epsilon_0. \tag{2.11}$$

This equation gives the electric field at a point immediately outside the surface of a charged conductor while the field inside is zero. Therefore as the surface of a charged conductor is crossed, the electric field varies in some continuous manner such as illustrated in Fig. 2-19. ▲

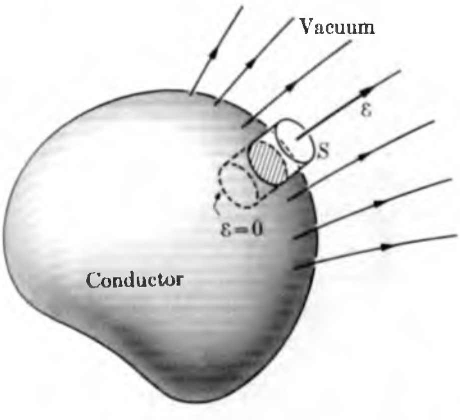


Fig. 2-18. The electric field at the surface of a conductor is normal to the surface.

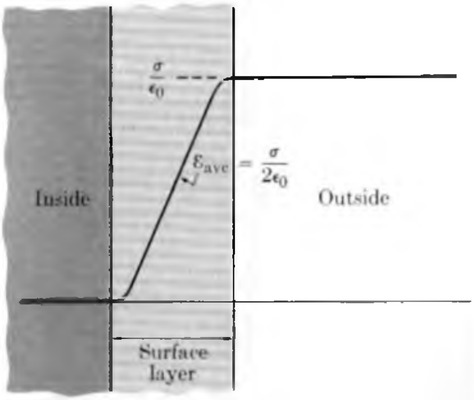


Fig. 2-19. Variation of the electric field when crossing the surface of a conductor.

Example 2.11. Determination of the force per unit area on the charges on the surface of a conductor.

▼ Each charge on the surface of a conductor is subject to a repulsive force because of the other charges present on the surface. The force per unit area, or *electric stress*, can be computed by multiplying the average electric field by the charge per unit area. From Fig. 2-19, the average field is $\mathscr{E}_{\text{ave}} = \sigma/2\epsilon_0$. Therefore the electric stress is

$$F = \sigma\mathscr{E}_{\text{ave}} = \sigma^2/2\epsilon_0.$$

The electric stress is always positive since it depends on σ^2 and therefore corresponds to a force pulling the charges away from the conductor. ▲

2.6 Electric Displacement

In the preceding section it was shown that a polarized dielectric has certain charges on its surface (and also throughout its volume unless the polarization is uniform). These polarization charges, however, are "frozen" in that they are bound to specific atoms or molecules and are not free to move through the dielectric. In other materials, such as a metal or an ionized gas, there may be electric charges capable of moving through the material, and therefore called *free* charges. In many instances a clear distinction between free charges and polarization charges must be made. The discussion in this section is a case in point.

Again consider a slab of a dielectric material placed between two conducting parallel plates (Fig. 2-20), carrying equal and opposite free charges. The surface charge density on the left-hand plate is $+\sigma_{\text{free}}$ and on the right-hand plate is $-\sigma_{\text{free}}$. These charges produce an electric field that polarizes the slab so that polarization charges appear on each surface of the slab. These polarization charges have a sign

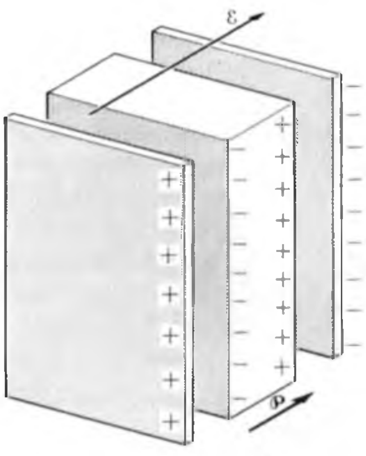


Fig. 2-20. Dielectric placed between oppositely charged plates. The charges on the plates are free charges and the charges on the dielectric surfaces are bound polarization charges.

opposite to that of the adjacent plate. Therefore the polarization charges on the faces of the dielectric slab partially balance the free charges on the conducting plates. Given that $\mathscr{P}$ is the magnitude of the polarization in the slab, the surface charge density is $-\mathscr{P}$ on the left face of the slab and $+\mathscr{P}$ on the right face. The effective, or net, surface charge density on the left is $\sigma=\sigma_{\text{free}}-\mathscr{P}$, with an equal and opposite result on the right. These net surface charges give rise to a uniform electric field that according to Eq. (2.11), is given by $\mathscr{E}=\sigma/\epsilon_0$. Thus using the effective value of σ, we have

$$\mathscr{E}=\frac{1}{\epsilon_0}(\sigma_{\text{free}}-\mathscr{P}) \qquad \text{or} \qquad \sigma_{\text{free}}=\epsilon_0\mathscr{E}+\mathscr{P},$$

an expression that gives the free charges on the surface of a conductor surrounded by a dielectric in terms of the electric field in the dielectric and the polarization of the dielectric. The result above suggests the introduction of a new vector field, which is called the *electric displacement*, defined by

$$\mathscr{D}=\epsilon_0\mathscr{E}+\mathscr{P}. \tag{2.12}$$

Obviously $\mathscr{D}$ is expressed in C m^{-2} since this is the unit of the two terms that appear on the right-hand side of Eq. (2.12). In the special case considered here, $\sigma_{\text{free}}=\mathscr{D}$; that is, the free charges per unit area on the surface of the conductor are equal to the electric displacement in the dielectric. This result has general validity and may be extended to conductors of any shape. Thus *the component of $\mathscr{D}$ along the normal to the surface of a conductor embedded in a dielectric gives the surface charge density on the conductor*. That is,

$$\sigma_{\text{free}}=\mathscr{D}\cdot\boldsymbol{u}_N$$

while the normal component of $\epsilon_0\mathscr{E}$ gives the effective or net charge, which takes into account the compensation due to the charges on the surface of the dielectric [recall Eq. (2.11), $\mathscr{E}=\sigma/\epsilon_0$]. That is, $\sigma=\epsilon_0\mathscr{E}\cdot\boldsymbol{u}_N$. The total free charge on a conductor is then

$$q_{\text{free}}=\oint_S \sigma_{\text{free}}\,dS=\oint_S \mathscr{D}\cdot\boldsymbol{u}_N\,dS=\Phi_{\mathscr{D}}. \tag{2.13}$$

A more detailed analysis, omitted here, indicates that *the flux of $\mathscr{D}$ over any closed surface is equal to the total "free" charge inside the surface, excluding all charges caused by the polarization of the medium*. Therefore Eq. (2.13) has general validity for any closed surface.

For cases in which Eq. (2.10) holds so that the polarization is directly proportional to the electric field,

$$\mathscr{D}=\epsilon_0\mathscr{E}+\epsilon_0\chi_e\mathscr{E}=(1+\chi_e)\epsilon_0\mathscr{E}=\epsilon\mathscr{E} \tag{2.14}$$

where the coefficient

$$\epsilon=\frac{\mathscr{D}}{\mathscr{E}}=(1+\chi_e)\epsilon_0 \tag{2.15}$$

is called the *permittivity* of the medium and is expressed in the same units as ϵ_0; that is, $m^{-3}\ kg^{-1}\ s^2\ C^2$. The *relative permittivity* is defined as

$$\epsilon_r = \epsilon/\epsilon_0 = 1 + \chi_e, \tag{2.16}$$

and is a pure number, independent of any system of units. The relative permittivity is also called the *dielectric constant*. For most substances it is larger than one.

When the relation $\mathcal{D} = \epsilon\mathcal{E}$ holds for a medium, Eq. (2.13) may be written as $q_{\text{free}} = \oint \epsilon\mathcal{E}\cdot\boldsymbol{u}_N\, dS$ and, if the medium is homogeneous so that ϵ is constant,

$$\Phi_{\mathcal{E}} = \oint \mathcal{E}\cdot\boldsymbol{u}_N\, dS = \frac{q_{\text{free}}}{\epsilon}.$$

Comparing Eq. (2.17) with Eq. (2.3) shows that the effect of the dielectric on the electric field $\mathcal{E}$ is to replace ϵ_0 by ϵ if only the free charges are taken into account. Therefore the electric field and the electric potential produced by a point charge embedded in a dielectric are

$$\mathcal{E} = \frac{q}{4\pi\epsilon r^2}\boldsymbol{u}_r \qquad \text{and} \qquad V = \frac{q}{4\pi\epsilon r}. \tag{2.18}$$

The magnitude of the force of interaction between two point charges embedded in a dielectric is then

$$F = \frac{q_1 q_2}{4\pi\epsilon r^2}. \tag{2.19}$$

Since ϵ is in general larger than ϵ_0, the presence of the dielectric effectively reduces the interaction because of the screening caused by the polarization of the molecules of the dielectric.

2.7 Calculation of Electric Susceptibility

The concept of electric susceptibility χ_e was introduced to describe the response of a substance to the action of an external field. The electric susceptibility of a substance must be related to the properties of the atoms and the molecules of the substance. This section will be a brief investigation of how this property, of macroscopic character, is related to the atomic properties of the substance.

As seen before, an atom placed in an electric field becomes polarized because of a relative displacement of the positive and the negative charges. If $\boldsymbol{p}$ is the electric dipole moment induced in the atom by an external field $\mathcal{E}$, and if $\boldsymbol{p}$ is assumed proportional to $\mathcal{E}$, a result confirmed by experience, the proportionality may be written

$$\boldsymbol{p} = \alpha\epsilon_0\mathcal{E} \tag{2.20}$$

where α is a constant characteristic of each atom, is called *polarizability*, and is expressed in m^3. The constant ϵ_0 is written into the equation explicitly for convenience. If there are n atoms or molecules per unit volume, the polarization of the medium is $\mathscr{P}=np=n\alpha\epsilon_0\mathscr{E}$. Comparison with Eq. (2.10) for the electric susceptibility of the material* gives $\chi_e=n\alpha$.

Thus the calculation of the electric susceptibility reduces to the calculation of the polarizability of the atoms (or molecules) of the substance. This calculation amounts to determining the effect of an external field on the motion of atomic electrons; but that determination in turn requires that some detailed information about the electronic motion in an atom must be available. The calculation of the perturbative effect of the external field must be carried out using the techniques of quantum mechanics and is thus beyond the scope of this book. Therefore only the main results, separating the effect for nonpolar substances from that for polar substances, will be presented.

Distortion effect

When the atoms or the molecules of a substance do not have a permanent electric dipole moment, polarization arises entirely from the distortion effect produced by the electric field on the electronic orbits. This effect may be described as a displacement of the center of the electronic charge distribution relative to the nucleus. The result is an induced electric dipole that in atoms and most molecules is parallel to the applied electric field.

Each atom (or molecule) has a characteristic set of frequencies $\omega_1, \omega_2, \omega_3, \ldots$ corresponding to the frequencies of the electromagnetic radiation that the substance can emit or absorb. These frequencies constitute the *electromagnetic spectrum* of the substance. The atomic polarizability when the electric field is constant is called the *static polarizability* and is given by the expression

$$\alpha=\frac{e^2}{\epsilon_0 m_e}\sum_i \frac{f_i}{\omega_i^2} \tag{2.21}$$

where ω_i refers to any of the resonant frequencies of the electromagnetic spectrum of the substance. and the summation extends over all these frequencies. The quantities designated by f_i are called the *oscillator strengths* of the substance. They are all positive and smaller than one, and represent the relative proportion in which each of the resonant frequencies of the spectrum contributes to the polarizability of the atom. These oscillator strengths satisfy the relation $\Sigma_i f_i=1$. The other quantities in Eq. (2.21) have their standard meanings.

The presence of a frequency ω_i associated with an effect produced by a static field may puzzle a student, but can be justified by using a very simple phenomenological model as will be indicated in Example 2.12.

*Strictly speaking, when Eq. (2.20) is written for an atom or molecule that is embedded in a material medium and is not isolated, the electric field appearing on the right-hand side of the equation must be the resultant electric field in the medium minus the electric field produced by the atom itself. When this correction is included, the relation between χ_e and α becomes $\chi_e=n\alpha/(1-n\alpha/3)$. However for most materials (mainly gases) the relation $\chi_e=n\alpha$ is a good approximation.

From the relation $\chi_e = n\alpha$, the static electric susceptibility is

$$\chi_e = \frac{ne^2}{\epsilon_0 m_e} \sum_i \frac{f_i}{\omega_i^2} = 3.19 \times 10^3 n \sum_i \frac{f_i}{\omega_i^2} . \tag{2.22}$$

This expression relates a macroscopic property, χ_e, to the atomic properties n, ω_i, and f_i, of the substance. These results will now be compared with experiment. If the radiation of the atom falls in the visible region, the frequencies ω_i are of the order of 5×10^{15} Hz so that the summation that appears in Eq. (2.22) is of the order of 4×10^{-32}. Also n is of the order of 10^{28} atoms per cubic meter for most solids and liquids and about 10^{25} atoms/m^3 for gases at STP. Therefore Eq. (2.22) shows that the static electric susceptibility χ_e of nonpolar materials that radiate in the visible region is of the order of 10^0 (or one) for solids and 10^{-3} for gases. Since these estimates are very crude, a precise reproduction of experimental results should not be expected. However, comparison with experimental values of the electric susceptibility for a few materials as given in Table 2-1 shows agreement insofar as the order of magnitude is concerned.

The foregoing discussion is valid only for static fields. If a field is time dependent, the *dynamic polarizability* should be different from the static value because the distortion of the electronic motion under a time-dependent electric field will obviously be different from that for a static electric field. Assume that the electric field oscillates with a definite frequency ω. This oscillating field will superpose an oscillatory perturbation on the natural motion of the electrons; this perturbation is analogous to the forced oscillations discussed in Section 12.13 of Volume I.* When damping is not considered, the result of the calculation, using the techniques of quantum mechanics, gives the dynamic susceptibility as

$$\chi_e = \frac{ne^2}{\epsilon_0 m_e} \sum_i \frac{f_i}{\omega_i^2 - \omega^2} \tag{2.23}$$

where all quantities have the meanings previously stated. A simple phenomenological justification of this result is given in Example 2.12. Note that the dynamic result (2.23) reduces to the static case, Eq. (2.22), if $\omega = 0$.

The dielectric constant or relative permittivity of the medium, from Eq. (2.23), is in the dynamic case

$$\epsilon_r = 1 + \chi_e = 1 + \frac{ne^2}{\epsilon_0 m_e} \sum_i \frac{f_i}{\omega_i^2 - \omega^2} . \tag{2.24}$$

If the permitivity ϵ_r is plotted against ω, it is seen that ϵ_r is infinite for ω equal to each characteristic frequency ω_i, in contradiction to observation. This unphysical result is due to the exclusion of a damping term when the dynamic susceptibility was calculated. The damping that occurs is not due to the electron moving in a viscous fluid;

*See Volume I, Chapter 12, for a discussion of free and forced oscillators.

Table 2-1. Electric Susceptibilities at Room Temperature

Substance	χ_e	Substance	χ_e
Solids		*Gases**	
Mica	5	Hydrogen	5.0×10^{-4}
Porcelain	6	Helium	0.6×10^{-4}
Glass	8	Nitrogen	5.5×10^{-4}
Bakelite	4.7	Oxygen	5.0×10^{-4}
		Argon	5.2×10^{-4}
Liquids		Carbon dioxide	9.2×10^{-4}
Oil	1.1	Water vapor	7.0×10^{-3}
Turpentine	1.2	Air	5.4×10^{-4}
Benzene	1.84	Air (10 MPa)	5.5×10^{-2}
Alcohol (ethyl)	24		
Water	78		

* At STP (100 k Pa and 298 K).

rather the damping corresponds to the energy lost by the electron as radiation as a result of the forced oscillations. (This point will be discussed later).

The observed variation of ϵ_r in terms of ω is illustrated in Fig. 2-21. The pattern repeats itself for the characteristic frequencies ω_1, ω_2, ω_3, . . . of each substance. This variation has a profound influence on the optical and electrical behavior of the substance.

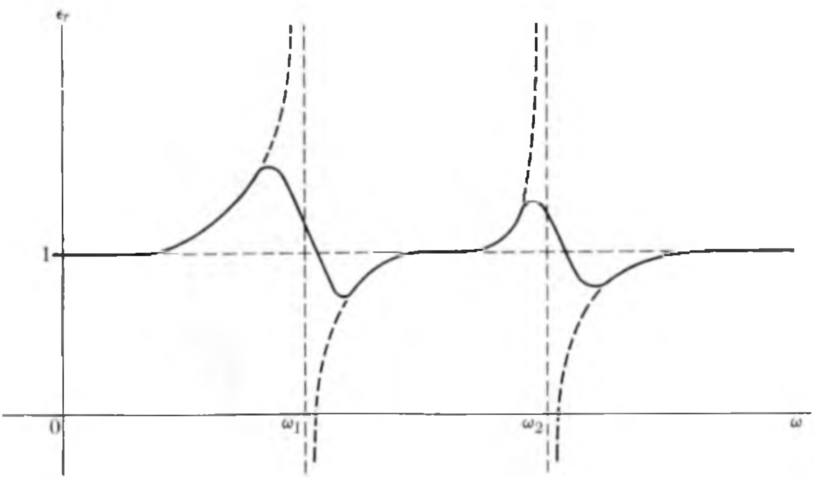


Fig. 2-21. Variation of relative permittivity as a function of the frequency of the electric field. The dashed line is the value predicted by Eq. (2.24); the solid line shows the typical observed behavior of a medium.

(a) No field

(b) Electric field without molecular interactions

(c) Electric field with molecular interactions

Fig. 2-22. Orientation of electric dipoles in an electric field.

Molecules with permanent dipole moment

The susceptibilities obtained in Eqs. (2.22) and (2.23) are "induced" because they result from a distortion of the electronic motion by an external field. However when molecules have a permanent electric dipole moment, another effect enters into play. Consider a polar gas whose molecules have a permanent dipole moment p_0. In the absence of any external electric field, these dipole moments are oriented at random, and no macroscopic or collective dipole moment is observed (Fig. 2-22). However when a static electric field is applied, the electric dipoles tend to orient in the direction of the field. The alignment would be perfect in the absence of molecular interactions (Fig. 2-22b); but molecular collisions tend to disarrange the electric dipoles. The disarrangement is not complete because the applied electric field favors orientation along the field over orientation against it (Fig. 2-22c). As a result, the average value of the component of the electric dipole moment of a molecule parallel to the electric field is given by

$$p_{\text{ave}}=\frac{p_0^2}{3kT}\mathscr{E} \tag{2.25}$$

where k is the Boltzmann constant and T is the absolute temperature of the gas. Note that p_{ave} decreases when the temperature increases. This temperature dependence occurs because molecular agitation increases with an increase in temperature; the more rapidly the molecules move about, the more effective they become in offsetting the aligning effect of the applied electric field. A decreased average dipole moment along the field direction results.

From a comparison of Eq. (2.25) with Eq. (2.20) the average, or effective, polarizability of a molecule is given by $\alpha=p_0^2/3\epsilon_0 kT$; and if there are n molecules per unit volume, the effective susceptibility $\chi_e=n\alpha$ is*

$$\chi_e=\frac{np_0^2}{3\epsilon_0 kT}, \tag{2.26}$$

*The relations given in Eqs. (2.25) and (2.26) are good approximations only when $p_0\mathscr{E}/kT \ll 1$.

a result known as *Langevin's formula*. The molecular electric dipole moments are of the order of magnitude of the electronic charge (1.6×10^{-19} C) multiplied by the molecular dimensions (10^{-10} m), or about 10^{-30} C m (remember Table 1-2). If the values of the other constants are introduced into Eq. (2.26), at room temperature ($T = 298$ K) the electric susceptibility of a substance composed of polar molecules is again of the order of 10^0 (or one) for solids and 10^{-3} for gases, a result in agreement with the values for most polar gases.

Note that the electric susceptibility caused by the orientation of molecules with permanent dipole moments is inversely proportional to the absolute temperature while the induced electric susceptibility due to the distortion of electronic motion in atoms or molecules, Eq. (2.22), is essentially temperature independent except that n varies with temperature because of thermal expansion. This difference in temperature dependence offers a means of separating the two effects experimentally. A temperature dependence of the form

$$\chi_e = A + \frac{B}{T}$$

should be expected in measuring χ_e at different temperatures. When the electric field is time dependent, a more complex result is obtained.

A special class of substances called *ferroelectrics* exhibit a permanent polarization in the absence of an external electric field; this characteristic suggests a natural tendency for the permanent dipoles of their molecules to align. The alignment probably results from the mutual interactions of the molecules producing strong local fields that favor alignment. Among the better-known ferroelectric substances are $BaTiO_3$, $KNbO_3$, $LiTaO_3$, and Rochelle salt: $NaK(C_4H_4O_6) \cdot 4H_2O$.

Example 2.12. Polarization of an atom because of an external electric field.

▼ From an oversimplified and phenomenological model an attempt will be made to determine the effect an external electric field produces on the electronic motion in atoms.

Suppose that when the center of the electronic motion is displaced a distance x relative to the nucleus, an average force $-kx$ acts on the electron. This force tends to restore the electron to the normal configuration. Equilibrium requires that this force balance the force $-e\mathscr{E}$ caused by the applied electric field. Therefore $-kx - e\mathscr{E} = 0$ or $x = -e\mathscr{E}/k$. The negative sign indicates that the electron's orbit is displaced in the direction opposite to that of the electric field. The electric dipole moment induced in the atom by the perturbation of the electronic motion is $p = -ex = (e^2/k)\mathscr{E}$, and thus is in the same direction as the electric field. This relation may be expressed in a slightly different manner by associating a frequency ω_0 with the constant k, corresponding to harmonic oscillations; that is, $k = m_e\omega_0^2$. Then in vector form

$$\boldsymbol{p} = \frac{e^2}{m_e\omega_0^2}\boldsymbol{\mathscr{E}}.$$

When this result is compared with the definition given in Eq. (2.20), the atomic polarizability for this simple model is

$$\alpha = \frac{e^2}{\epsilon_0 m_e \omega_0^2}.$$

From the relation $\chi_e = n\alpha$, the static electric susceptibility is

$$\chi_e = \frac{ne^2}{\epsilon_0 m_e \omega_0^2} = (3.19 \times 10^3)\frac{n}{\omega_0^2}. \tag{2.27}$$

If the model is to have physical meaning, the frequency ω_0 must be identified with some atomic property. When the field $\mathscr{E}$ is removed, the restoring force $-kx$ superposes on the natural motion of the electron an oscillation of frequency ω_0. Later on, in Chapter 11 it will be shown that an oscillating charge radiates energy. Thus ω_0 may be identified with the frequency of the radiation emitted by the atom. Hence if the spectrum of the substance contains only one frequency ω_0, the model basically coincides with Eq. (2.21).

Now consider the time-dependent case in which the applied electric field varies with time according to $\mathscr{E} = \mathscr{E}_0 \cos \omega t$. It is then reasonable to assume that an oscillatory perturbation is superposed on the natural motion of the electron and results in an equation of motion given by

$$m_e \frac{d^2x}{dt^2} = -kx - e\mathscr{E}_0 \cos \omega t \tag{2.28}$$

where the last term is the force produced by the oscillating field. If $k = m_e\omega_0^2$, the equation above may be written in the form

$$\frac{d^2x}{dt^2} + \omega_0^2 x = -\frac{e\mathscr{E}_0}{m_e}\cos \omega t. \tag{2.29}$$

This equation describes forced oscillations of a harmonic oscillator.* (The main difference from the analysis of a forced oscillation in Volume I is that here there is no damping term.) Assume a solution of the form $x = A \cos \omega t$: when substituted into Eq. (2.29), this solution gives $A = -e\mathscr{E}_0/(\omega_0^2 - \omega^2)$. Therefore

$$x = -\frac{e}{m_e(\omega_0^2 - \omega^2)}\mathscr{E}_0 \cos \omega t = -\frac{e}{m_e(\omega_0^2 - \omega^2)}\mathscr{E}$$

since $\mathscr{E} = \mathscr{E}_0 \cos \omega t$. The induced electric dipole is

$$p = -ex = \frac{e^2}{m_e(\omega_0^2 - \omega^2)}\mathscr{E},$$

from which the dynamic polarizability of the atom is

$$\alpha = \frac{e^2}{\epsilon_0 m_e(\omega_0^2 - \omega^2)}. \tag{2.30}$$

To obtain the dynamic susceptibility, again use the relation $\chi_e = n\alpha$ and find that

$$\chi_{e\,(\mathrm{dynamic})} = \frac{ne^2}{\epsilon_0 m_e(\omega_0^2 - \omega^2)}, \tag{2.31}$$

which is essentially identical to Eq. (2.23) if there is only one resonant frequency ω_0 in the electro-

*See Volume I, Section 12-13.

magnetic spectrum of the substance. Once more note that the crude phenomenological model cannot give precise results. One of the reasons is the assumption of a single natural frequency ω_0 as in the static case. Another reason is that the electron's motion follows the laws of quantum mechanics rather than newtonian mechanics, a fact that has been ignored here. ▲

2.8 Electric Capacitance; Capacitors

It has been proved (Section 1.6) that the electric potential at the surface of a spherical conductor of radius R and charge Q is $V=Q/4\pi\epsilon_0 R$. If the sphere is surrounded by a dielectric, ϵ_0 is replaced by ϵ:

$$V=\frac{Q}{4\pi\epsilon R}.$$

The relation Q/V for the sphere is then $4\pi\epsilon R$, a constant quantity, independent of the charge Q. This constancy is understandable because if the potential is proportional to the charge producing it, the ratio of the two must be a constant. This last statement is valid for all charged conductors of any geometrical shape. Accordingly, the electric *capacitance* of an isolated conductor is defined as the ratio of its charge to its potential:

$$C=\frac{Q}{V}. \tag{2.32}$$

This equation indicates how much charge is stored on a conductor for a given electric potential.

The capacitance of an isolated spherical conductor may be written as

$$C_{\text{sphere}}=4\pi\epsilon R.$$

If the sphere is surrounded by vacuum instead of a dielectric, its capacitance is $C_{\text{sphere}}=4\pi\epsilon_0 R$. Therefore surrounding a sphere and in general any conductor by a dielectric increases its electric capacitance by the factor ϵ/ϵ_0 because of the screening effect of the opposite charges that have been induced on the surface of the dielectric adjacent to the conductor. These charges reduce the effective charge of the conductor and decrease the potential of the conductor by the same factor.

The capacitance of a conductor is expressed in C V^{-1}, a unit called the *farad* (F) in honor of the Englishman Michael Faraday (1791–1867). The farad is defined as the capacitance of an isolated conductor whose electric potential is one volt after the conductor receives a charge of one coulomb. In terms of the fundamental units, $\text{F}=\text{C V}^{-1}=\text{m}^{-2}\,\text{kg}^{-1}\,\text{s}^2\,\text{C}^2$.

The concept of electric capacitance can be extended to a system of conductors. Consider the case of two conductors having charges Q and $-Q$ (Fig. 2-23). If V_1

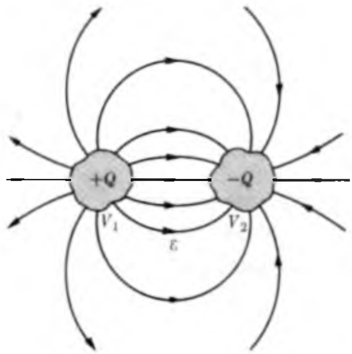


Fig. 2-23. System of two conductors with equal but opposite charges.

and V_2 are their respective potentials so that $\Delta V = V_1 - V_2$ is their potential *difference*, the capacitance of the system is defined as

$$C = \frac{Q}{V_1 - V_2} = \frac{Q}{\Delta V}. \tag{2.33}$$

This arrangement constitutes what is defined as a *capacitor*. Capacitors have wide application in electric circuits. A typical capacitor is formed by two parallel plane conductors separated a distance d, with the space between them filled by a dielectric (Fig. 2-24). The electric field in the space between the conductors is uniform and is

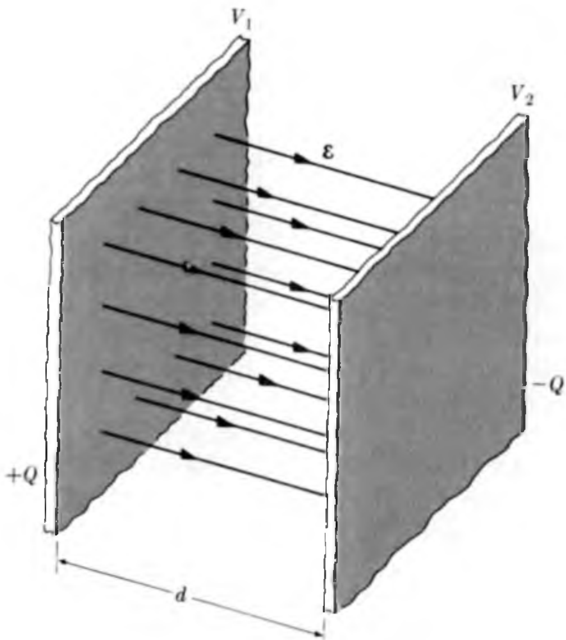


Fig. 2-24. Parallel plate capacitor.

given by $\mathscr{E}=(V_1-V_2)/d$ according to Eq. (1.19). When σ is the surface charge density on the plates according to Example 2.2, the intensity of the electric field in the space between the plates is $\mathscr{E}=\sigma/\epsilon$ where ϵ_0 has been replaced by ϵ because of the presence of a dielectric. Therefore

$$V_1-V_2=\mathscr{E}d=\frac{\sigma d}{\epsilon}.$$

On the other hand if S is the area of the metal plates, it then follows that $Q=\sigma S$. Therefore with the substitutions made in Eq. (2.33), the capacitance of the system is

$$C=\frac{Q}{\Delta V}=\frac{\sigma S}{(\sigma d/\epsilon)}=\frac{\epsilon S}{d}. \tag{2.34}$$

This equation suggests a practical means for measuring the permittivity or the dielectric constant of a material. First measure the capacitance of a capacitor with *no* material between the plates:

$$C_0=\frac{\epsilon_0 S}{d}.$$

Next fill the space between the plates with the material being investigated, and measure the new capacitance, given by Eq. (2.34). Then

$$\frac{C}{C_0}=\frac{\epsilon}{\epsilon_0}=\epsilon_r.$$

Therefore the ratio of the two capacitances gives the relative permittivity or dielectric constant of the material placed between the plates.

Example 2.13. Combinations of capacitors.

▼ Capacitors can be combined in two kinds of arrangements: *series* and *parallel*. In the series combination (see Fig. 2-25a) the negative plate of one capacitor is connected to the positive of the

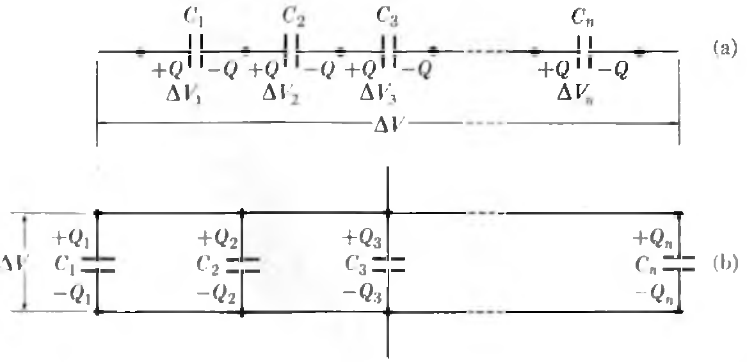


Fig. 2-25. Series and parallel arrangements of capacitors.

next, and so on. As a result, all capacitors carry the same charge, positive or negative, on their plates. Call $\Delta V_1, \Delta V_2, \ldots, \Delta V_n$ the potential differences across each capacitor. If $C_1, C_2, \ldots, C_n$ are their respective capacitances, then $\Delta V_1 = Q/C_1, \Delta V_2 = Q/C_2, \ldots, \Delta V_n = Q/C_n$. Thus the overall potential difference is

$$\Delta V = \Delta V_1 + \Delta V_2 + \cdots + \Delta V_n = \left(\frac{1}{C_1} + \frac{1}{C_2} + \cdots + \frac{1}{C_n}\right) Q.$$

The system can be equated to a single capacitor whose capacitance C satisfies the relation $\Delta V = Q/C$. Therefore

$$\frac{1}{C} = \frac{1}{C_1} + \frac{1}{C_2} + \cdots + \frac{1}{C_n}, \tag{2.35}$$

which gives the resultant capacitance for a series arrangement of capacitors.

In the parallel combination (Fig. 2-25b) all positive plates are connected to a common point, and the negative plates are also connected to another common point so that the potential difference ΔV is the same for all the capacitors. Thus if their charges are $Q_1, Q_2, \ldots, Q_n$, we must have $Q_1 = C_1 \Delta V, Q_2 = C_2 \Delta V, \ldots, Q_n = C_n \Delta V$. The total charge on the system is

$$Q = Q_1 + Q_2 + \cdots + Q_n = (C_1 + C_2 + \cdots + C_n) \Delta V.$$

The system can be equated to a single capacitor whose capacitance C satisfies the relation $Q = C \Delta V$. Therefore

$$C = C_1 + C_2 + \cdots + C_n \tag{2.36}$$

gives the resultant capacitance for a parallel arrangement of capacitors. It is possible to make combinations of capacitors in series and parallel. The resultant capacitance of the combination may usually be worked through, subnetwork by subnetwork, by using the rules developed in this example. ▲

2.9 Energy of the Electric Field

Charging a conductor requires expending energy because to bring more charge to a conductor, work must be done to overcome the repulsion of the charge already present. This work results in an increase in the energy of the conductor. For example consider a conductor of capacitance C having a charge q. The electric potential of the capacitor is $V = q/C$. If a charge dq is added to the conductor by bringing it from infinity to the plates, the work done is $dW = V\,dq$ according to Eq. (1.24). This work is equal to the increase in energy dE_e of the conductor. Therefore with the value of V, the change in energy is

$$dE_e = \frac{q\,dq}{C}.$$

When the charge is increased from zero to the value Q, the total increase in energy of the conductor equals the work done during the process and

$$E_e = \frac{1}{C}\int_0^Q q\,dq = \frac{Q^2}{2C}. \tag{2.37}$$

For the case of a spherical conductor, $C = 4\pi\epsilon R$, and the energy is

$$E_e = \frac{1}{2}\left(\frac{Q^2}{4\pi\epsilon R}\right). \tag{2.38}$$

The energy of a charged sphere can be related to the electric field of the charged sphere in a very interesting way. The magnitude of the electric field of a spherical charged conductor at a distance r, larger than its radius, is

$$\mathscr{E} = \frac{Q}{4\pi\epsilon r^2}.$$

The integral of $\mathscr{E}^2$ over all the volume exterior to the sphere will now be calculated. To obtain the volume element for the integration, divide the outer space into thin spherical shells of radius r and thickness dr (Fig. 2-26). The surface area of each shell is $4\pi r^2$, and its volume is $dv = \text{area} \times \text{thickness} = 4\pi r^2\,dr$. Therefore

$$\int_R^\infty \mathscr{E}^2\,dv = \int_R^\infty \left(\frac{Q}{4\pi\epsilon r^2}\right)^2 (4\pi r^2\,dr) = \frac{Q^2}{4\pi\epsilon^2}\int_R^\infty \frac{dr}{r^2} = \frac{Q^2}{4\pi\epsilon^2 R}.$$

Comparison of this result with Eq. (2.38) shows that the energy of a charged spherical conductor may be written as

$$E_e = \tfrac{1}{2}\epsilon \int_R^\infty \mathscr{E}^2\,dv.$$

A more general mathematical calculation indicates that this result has general validity, and the energy required to assemble a system of charges can thus be expressed as

$$E_e = \tfrac{1}{2}\epsilon \int_{\text{All space}} \mathscr{E}^2\,dv. \tag{2.39}$$

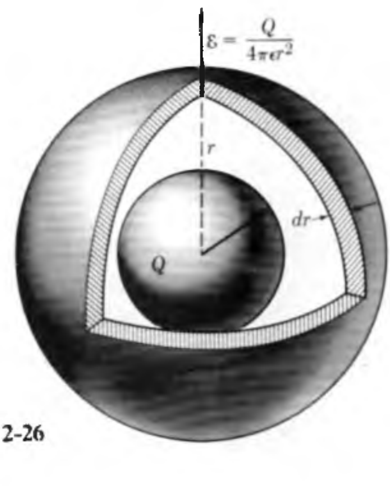


Figure 2-26

This expression may be given an important physical interpretation. We may say that the energy spent in assembling the charges has been *stored* in the surrounding space so that each infinitesimal volume dv has a corresponding energy $\frac{1}{2}\epsilon\mathscr{E}^2\,dv$. Hence the energy per unit volume, or energy density E_e "stored" in the electric field, is

$$\mathrm{E}_e=\tfrac{1}{2}\epsilon\mathscr{E}^2. \tag{2.40}$$

This interpretation of the energy of a system of charged particles distributed throughout all the space where the electric field is present is very useful in the discussion of many processes.

Example 2.14. The energy required to assemble a spherical charge distributed *uniformly* throughout the volume of the sphere (Fig. 2-27).

▼ Call R the radius of the sphere and Q the charge that is distributed uniformly throughout its volume (Fig. 2-27). Divide the volume of the sphere into a series of thin spherical shells of increasing radius from zero up to the radius R of the sphere. Imagine that the spherical distribution of charge has been built in an onionlike fashion by adding successive spherical shells until the final radius is attained. To compute the total energy of the spherical charge distribution, we sum the energy spent in adding each of the shells.

The charge density throughout the sphere is

$$\rho=\frac{Q}{4\pi R^3/3}.$$

When the radius of the sphere is r, the charge q contained in it is

$$q=\rho(\tfrac{4}{3}\pi r^3)=\frac{Qr^3}{R^3} \tag{2.41}$$

and the electric potential at this surface is

$$V=\frac{q}{4\pi\epsilon_0 r}=\frac{Qr^2}{4\pi\epsilon_0 R^3}.$$

To increase the radius by the amount dr by adding a new shell requires adding a charge dq. The quantity dq is obtained by differentiating Eq. (2.41) to yield

$$dq=\frac{3Qr^2}{R^3}\,dr.$$

The energy required to add this charge to the sphere is

$$dE_e=V\,dq=\frac{3Q^2r^4}{4\pi\epsilon_0 R^6}\,dr.$$

The total energy required to build up the charge to its final value is then

$$E_e=\int_0^Q V\,dq=\int_0^R \frac{3Q^2r^4}{4\pi\epsilon_0 R^6}\,dr=\frac{3Q^2}{4\pi\epsilon_0 R^6}\int_0^R r^4\,dr.$$

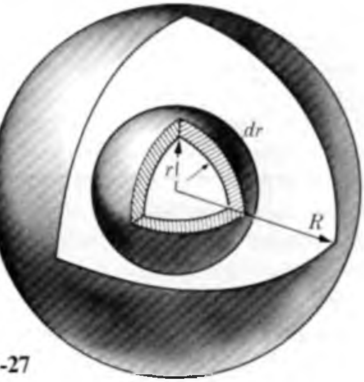


Figure 2-27

The integration yields

$$E_e = \frac{3}{5}\left(\frac{Q^2}{4\pi\epsilon_0 R}\right), \tag{2.42}$$

a result that differs from Eq. (2.38). The reason is that when Eq. (2.38) was derived, the charge was added to a sphere of constant radius, while for Eq. (2.42) a sphere charged uniformly throughout its volume was produced by adding successive layers until the final size was attained. We leave it to the student to verify that in this case relation (2.39) still holds, but the energy associated with the electric field *inside* the sphere must be included in the computation.

An interesting application of Eq. (2.42) is to estimate the electric (or coulomb) energy of a nucleus. If the nuclear charge is $Q = Ze$,

$$E_e = \frac{3}{5}\frac{Z^2 e^2}{4\pi\epsilon_0 R}. \tag{2.43}$$

However in the case of a nucleus composed of protons and neutrons, there is not a uniform distribution of the charge throughout the volume of the sphere. The charge is only on the protons, and a more careful analysis yields a slightly different result, in which Z^2 is replaced by $Z(Z-1)$. ▲

Example 2.15. The "radius" of the electron.

▼ There is very little known about the geometrical shape of an electron. All that can be said for certain is that an electron is a negatively charged particle of charge $-e$. We are interested in estimating the size of the region where that charge is concentrated. To simplify the calculation, assume that the electron is a sphere of radius R. We may compute its electrical energy by using the methods above after making some assumptions about how the charge is distributed over the volume of the electron. For example if the electron resembles a solid sphere of radius R and has a uniform charge density with charge $-e$, the energy will be

$$E_e = \frac{3}{5}\frac{e^2}{4\pi\epsilon_0 R}.$$

This energy may be compared to the rest mass energy $m_e c^2$ of the electron. If the two energies are equated, then

$$m_ec^2=\frac{3}{5}\frac{e^2}{4\pi\epsilon_0 R} \qquad \text{or} \qquad R=\frac{3}{5}\left(\frac{1}{4\pi\epsilon_0}\right)\frac{e^2}{m_ec^2} \tag{2.44}$$

This expression gives the radius of the electron according to the model chosen. If instead of being a uniformly charged sphere, the electron is assumed to be charged only on its surface, Eq. (2.37) must be used for the energy. The expression obtained for the radius is similar to Eq. (2.44), but with the factor $\frac{3}{5}$ replaced by the factor $\frac{1}{2}$. Since the electron probably does not correspond to either of these models, it is customary to adopt as the *definition* of the radius of the electron the quantity

$$r_e=\left(\frac{1}{4\pi\epsilon_0}\right)\frac{e^2}{m_ec^2}=2.8178\times 10^{-15}\text{ m}. \tag{2.45}$$

We repeat—this radius cannot be considered in a strictly geometrical sense, but mainly as an estimate of the size of the region where the electron may be "concentrated." ▲

Problems

2.1 The cubical closed surface of side a shown in Fig. 2-28 is placed in a region in which there is an electric field parallel to the X-axis. Find the electric flux across the surface and the total charge inside the surface if the electric field (a) is uniform, and (b) varies according to $\mathscr{E}=Cx$.

2.2 Find the electric flux, the total charge, and the charge density inside the cube of side a (Fig. 2-28) if the cube is in a region in which the electric field is (a) $\boldsymbol{\mathscr{E}}=\boldsymbol{u}_x cx^2$, (b) $\boldsymbol{\mathscr{E}}=c(\boldsymbol{u}_x y+\boldsymbol{u}_y x)$.

2.3 In an ionized medium (such as a gas or an electrolyte), there are both positive and negative ions. Show that if each ion carries a charge $\pm ve$, the current density is $\boldsymbol{j}=ve(n_+\boldsymbol{v}_+-n_-\boldsymbol{v}_-)$ where the n_+ and n_- are the number of ions of each class per unit volume.

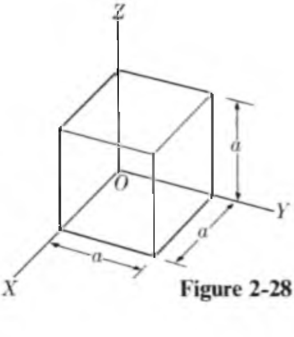


Figure 2-28

2.4 A conducting sphere of radius R_1 has a central cavity of radius R_2. At the center of the cavity there is a charge q. (a) Find the charge on the inner and the outer surfaces of the conductor. (b) Compute the electric field and the electric potential outside the sphere, inside the sphere, and in the cavity. (c) Plot the electric field and the electric potential as functions of the distance from the center. [*Hint:* Remember that the field inside a conductor is zero.]

2.5 Two conducting spheres of radii R_1 and R_2 are placed a large distance from each other but are connected by a wire so they are always at the same electric potential. There is an excess charge Q placed on the system. (a) Show that charge is distributed on the electrically joined spheres such that $\sigma_1/\sigma_2=R_2/R_1$ where σ is the surface density of electric charge. (b) Show therefore that the surface value of the electric field at each sphere is such that $\mathscr{E}_{1,\text{surface}}/\mathscr{E}_{2,\text{surface}}=R_2/R_1$. In solving this problem, ignore the effect of the wire.

2.6 Two conducting spheres of radii 1×10^{-3} m and 1.5×10^{-3} m have charges of $+1 \times 10^{-7}$ C and $+2 \times 10^{-7}$ C, respectively. The spheres are placed in contact and separated. Calculate the charge on each sphere after they are separated.

2.7 As a result of Eqs. (2.10) and (2.11), it can be proved that at the surface of separation of two dielectrics, the tangential component of the electric field and the normal component of the electric displacement are continuous; i.e., they have the same value on both sides of the surface. (The second statement holds only if the surface is uncharged.) Show then that the angles the lines of force make with the normal to the surface satisfy the relation

$$\frac{\tan \theta_1}{\tan \theta_2} = \frac{\epsilon_1}{\epsilon_2}.$$

2.8 The permittivity of diamond is 1.46×10^{-10} m^{-3}kg^{-1}s^{2}C^2. (a) What is the dielectric constant of diamond? (b) What is the susceptibility of diamond?

2.9 An air capacitor consisting of two closely spaced parallel plates has a capacitance of 1000 pF. The charge on each plate is 1 μC. (a) What is the potential difference between the plates? (b) If the charge is kept constant, what will be the potential difference between the plates if the separation is doubled?

2.10 A capacitor can be made by sandwiching a sheet of paper 4×10^{-5} m thick between sheets of tinfoil. The paper has a relative dielectric constant of 2.8 and will conduct electricity if it is in an electric field of strength 5×10^7 V m^{-1} (or greater). That is, the *dielectric strength* of the paper is 50 MV m^{-1}. (a) Determine the plate area needed for a 0.3 μF paper-and-foil capacitor. (b) What is the maximum charge that may be applied if the electric field in the paper is not to exceed one-half the dielectric strength?

2.11 A parallel-plate capacitor is to be constructed using rubber as a dielectric. This rubber has a dielectric constant of 3 and a dielectric strength of 20 MV m^{-1}. The capacitor is to have a capacitance of 0.15 μF and must be able to withstand a maximum potential difference of 6000 V. What is the minimum area the plates of the capacitor may have?

2.12 (a) Show that the capacity of a capacitor composed of spherical shells with radii a and b is $4\pi\epsilon_r ab/(a-b)$ where ϵ_r is the relative permittivity of the medium between the spheres. (b) Show that the capacity of a capacitor composed of two cylindrical shells with radii a and b is $4\pi\epsilon_r/2 \ln (b/a)$.

2.13 A certain capacitor is made of 25 thin metal sheets, each having an area of 6×10^{-2} m^2, and separated from each other by paraffin paper 6×10^{-4} m thick (relative permittivity equals 2.6). Find the capacitance of the system.

2.14 Three capacitors of 1.5 μF, 2 μF, and 3 μF are connected in (1) series and (2) parallel; a potential difference of 20 V is applied. Determine in each case (a) the capacity of the system, (b) the charge and potential difference on each capacitor, and (c) the energy of the system.

2.15 (a) Determine the capacity of the arrangement of capacitors illustrated in Fig. 2-29. If the applied voltage is 120 V, find (b) the charge and (c) the potential difference on each capacitor.

2.16 In the capacitor arrangement of Fig. 2-30, the capacitors are $C_1 = 3$ μF, $C_2 = 2$ μF, and

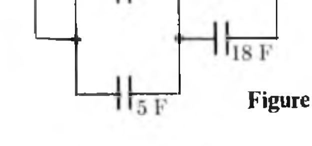


Figure 2-29

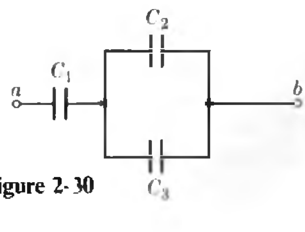


Figure 2-30

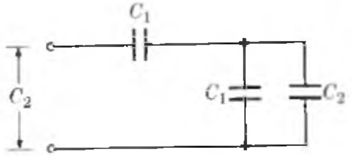


Figure 2-31

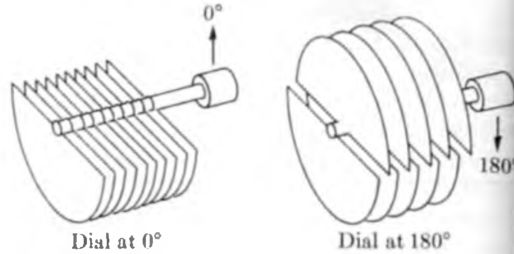


Figure 2-32

$C_3 = 4\ \mu F$. The voltage applied between points a and b is 300 V. Find the charge and the potential difference on each capacitor.

2.17 Given the capacitor arrangement shown in Fig. 2-31, show that the relation between C_1 and C_2 must be $C_2 = 0.618C_1$ in order that the capacity of the system be equal to C_2.

2.18 (a) Show that the electric energy of an isolated charged conductor is $\frac{1}{2}CV^2$. (b) Also show that the same result holds for a parallel-plate capacitor and, in general, for any capacitor.

2.19 The capacitance of a variable capacitor can be changed from 50 pF to 950 pF by turning a dial from 0° to 180° as shown in Fig. 2-32. With the dial set at 180°, the capacitor is connected to a 400-V battery. After it has been charged, the capacitor is disconnected from the battery and the dial is turned to 0°. (a) What is the charge on the capacitor? (b) What is the potential difference across the capacitor when the dial reads 0°? (c) What is the energy of the capacitor in this position? (d) Neglecting friction, determine the amount of work required to turn the dial.

2.20 A 20-μF capacitor is charged to a potential difference of 1000 V. The terminals of the charged capacitor are then connected to those of an uncharged 5-μF capacitor. Compute (a) the original charge of the system, (b) the final potential difference across each capacitor, (c) the final energy of the system, and (d) the decrease in energy when the capacitors are connected.

2.21 A metal sphere of radius 1 m has net electric charge of 10^{-9} C. This sphere is connected by a conducting wire to an initially uncharged sphere (far away from the larger sphere) of radius 0.30 m so that both have the same electric potential. (a) What will be the equilibrium charge on each sphere after the connection is made? (b) What is the energy of the charged sphere before connections are made? (c) What is the energy of the system after the spheres are joined? (d) If there is any loss, explain where the energy has gone.

CHALLENGING PROBLEMS

2.22 A charge Q is uniformly distributed along a semicircular arc of radius R as shown in Fig. 2-33. Find the magnitude and direction of the electric field at the center C of the arc.

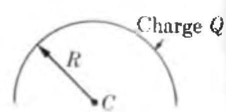


Figure 2-33

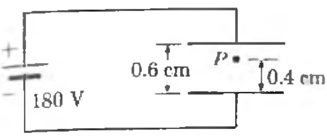


Figure 2-34

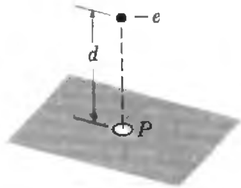


Figure 2-35

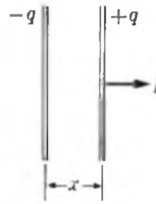


Figure 2-36

2.23 A battery having a potential difference of 180 volts is connected across a capacitor as shown in Fig. 2-34. The electric field in the evacuated region between the capacitor plates is uniform. The distance between the plates is 0.6 centimeter and point P is 0.4 centimeter from the lower plate. (a) Find the magnitude and direction of the electric field at point P. (b) Find the electric potential difference between point P and the lower plate. (c) An electron is released from rest on the lower plate. Find the instantaneous speed of the electron as it passes point P. (AP-B; 1971)

2.24 Consider an infinite plane sheet of positive charge with uniform surface charge density σ as shown in Fig. 2-35. (a) Use Gauss's law to determine the electric field due to this sheet of charge at a point a distance d above its surface. Suppose there is a tiny hole in the sheet at point P and that an electron (charge $-e$, mass m) is released from rest a distance d above the hole. The electron will pass through the hole and undergo periodic motion. (Neglect the distortion of the electric field due to the presence of the hole, and neglect the radiation of the accelerated charge.) (b) In terms of the given quantities, determine the speed of the electron when it passes through the hole. (c) In terms of the given quantities, determine the period of the motion. (AP-C; 1972)

2.25 The plates of an isolated parallel plate capacitor are pulled apart very slowly by a force F as shown in Fig. 2-36. Each plate has charge q and area A. Assume that edge effects are negligible; i.e., the spacing x is much smaller than the plate dimensions. (a) Determine the change in capacitance as x is increased by dx. (b) Determine the change of stored energy in the capacitor as x is increased by dx. (c) How is the force F related to the change in stored energy? Determine F in terms of x, q, and A. (AP-C; 1973)

2.26 Two concentric conducting spherical shells of radii a and b have charges of equal magnitude and opposite sign as shown in Fig. 2-37. (a) Determine the electric field at a distance r from the center. Give separate expressions for $r<a$, $a<r<b$, and $r>b$. (b) Determine the electric potential at $r=b$ and at $r=a$, taking the potential equal to zero at infinity. (AP-C; 1974)

2.27 A parallel-plate capacitor with spacing b and area A is connected to a battery of voltage V as shown in Fig. 2-38a. Initially the space between the plates is empty. Make the following determinations in terms of the given symbols. (a) Determine the electric field

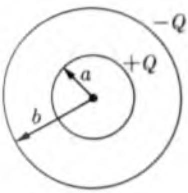


Figure 2-37

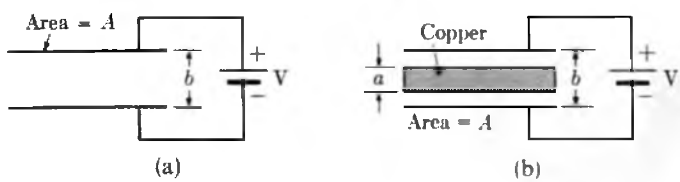


Figure 2-38

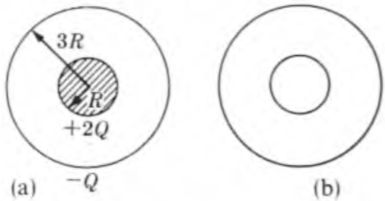


Figure 2-39

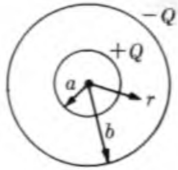


Figure 2-40

between the plates. (b) Determine the charge stored on each capacitor plate. A copper slab of thickness a is now inserted midway between the plates as shown in Fig. 2-38b. (c) Determine the electric field in the spaces above and below the slab. (d) Determine the ratio of capacitances

$$\frac{C_{\text{with copper}}}{C_{\text{original}}}$$

when the slab is inserted. (AP-C; 1974)

2.28 A solid metal sphere of radius R has charge $+2Q$. A hollow spherical shell of radius $3R$ placed concentric with the first sphere has net charge $-Q$ (Fig. 2-39a). (a) On a diagram like Fig. 2-39b, make a sketch of the electric field lines inside and outside the spheres. (b) Use Gauss's law to find an expression for the magnitude of the electric field between the spheres at a distance r from the center of the inner sphere ($R<r<3R$). (c) Calculate the potential difference between the two spheres. (d) What would be the final distribution of the charge if the spheres were joined by a conducting wire? (AP-C; 1976)

2.29 A capacitor is composed of two concentric spherical shells of radii a and b, respectively, that have equal and opposite charges as shown in Fig. 2-40. Just outside the surface of the inner shell, the electric field is directed radially outward and has magnitude $\mathscr{E}_0$. (a) With the use of Gauss's law, express the charge $+Q$ on the inner shell as a function of $\mathscr{E}_0$ and a. (b) Write an expression for the electric field strength E between the shells as a function of $\mathscr{E}_0$, a, and r. (c) What is the potential difference V between the shells as a function of $\mathscr{E}_0$, a, and b? (d) Express the energy U stored in this capacitor as a function of $\mathscr{E}_0$, a, and b. (e) Determine the value of a that should be chosen in order to maximize U, if $\mathscr{E}_0$ and b are fixed. (AP-C; 1978)

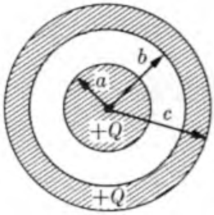


Figure 2-41

2.30 A solid conducting sphere of radius a is surrounded by a hollow conducting shell of inner radius b and outer radius c as shown in Fig. 2-41. The sphere and the shell each have a charge $+Q$. Express your answers to parts (a), (b), and (e) in terms of Q, a, b, c, and the Coulomb's law constant. (a) Using Gauss's law, derive an expression for the electric field magnitude at $a<r<b$, where r is the distance from the center of the solid sphere. (b) Write expressions for the electric field magnitude at $r>c$, $b<r<c$, and $r<a$. Full credit will be given for statements of the correct expressions. It is not necessary to show your work on this part. (AP-C; 1979)

2.31 A thin plastic rod has uniform linear positive-charge density λ. The rod is bent into a semicircle of radius R as shown in Fig. 2-42a. (a) Determine the electric potential V_0 at

Figure 2-42

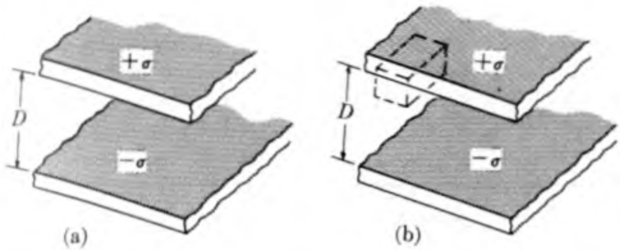


Figure 2-43

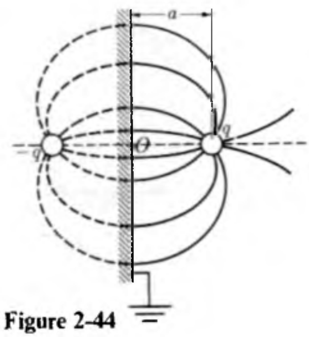


Figure 2-44

point O, the center of the semicircle. (b) Indicate on a copy of the diagram the direction of the electric field at point O. Explain your reasoning. (c) Calculate the magnitude $\mathscr{E}_0$ of the electric field at point O. (d) Write an approximate expression, in terms of q, V_0, and $\mathscr{E}_0$, for the work required to bring a positive point charge q from infinity to point P, located a small distance s from point O as shown in Fig. 2-42b. (AP-C; 1980)

2.32 A parallel-plate capacitor consists of two conducting plates separated by a distance D as shown in Fig. 2-43a. The plates may be considered very large so that the effects of the edges may be ignored. The two plates have an equal but opposite surface charge per unit area, σ. The charge on either plate resides entirely on the inner surface facing the opposite plate. (a) On a diagram, draw the electric-field lines in the region between the plates. (b) By applying Gauss's law to the rectangular box whose upper surface lies entirely within the top conducting plate as shown in the Fig. 2-43b, determine the magnitude of the electric field E in the region between the plates. (c) A dielectric is inserted and fills the region between the plates. Is the electric field greater than, less than, or equal to the electric field when there is no dielectric? Describe the mechanism responsible for this effect. Recognize that the plates are not connected to a battery. (AP-C: 1980)

2.33 The electron in a hydrogen atom may be assumed to be "spread" over all space with a density $\rho = Ce^{-2r/a_0}$ where $a_0 = 0.53 \times 10^{-10}$ m. (a) Find the constant C such that the total charge is $-e$. (b) Determine the total charge within a sphere of radius a_0, which corresponds to the orbit radius of the electron. (c) Obtain the electric field as a function of r. (d) At what distance does the electric field differ from $-e/4\pi\epsilon_0 r^2$ by 1 percent? [*Hint:* For part (a), divide the space into spherical shells, each of volume $4\pi r^2\,dr$.]

2.34 A sphere of radius R_1 has a central cavity of radius R_2. A charge q is uniformly distributed over the volume of the sphere. (a) Find the electric field and the electric potential outside the sphere, inside the sphere, and in the central cavity. (b) Plot the electric field and electric potential as functions of the distance from the center.

2.35 A charge q is placed a distance a from an infinite plane conductor held at zero electric potential. It can be shown that the resultant electric field in front of the plane is the same as if a negative charge $-q$, at a distance $-a$, were to replace the plane (see Fig. 2-44). This second charge is called the *image* of the first. (a) Show that the electric field is normal to the plane. (b) Show that the charge density on the plane is qa/r^3 where r is the radial distance from point O on the plane. (c) Verify that the total charge on the plane is equal to $-q$.

2.36 A conducting sphere of radius a is placed in a uniform electric field $\mathscr{E}_0$ as shown in Fig. 2-45. Since the sphere must be at a constant electric potential, we shall assign to it the value zero. The electric field acts on the free charges on the sphere that are moved to the surface until the electric field inside the sphere is zero. The sphere becomes polarized, distorting the electric field around it although at large distances

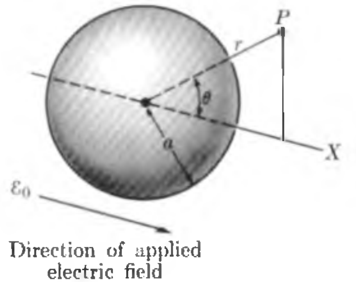


Figure 2-45

the field remains essentially uniform. It can be shown that to satisfy the conditions of this problem, the electric potential solution of *Laplace's equation* is

$$V = -\mathscr{E}_0 r \cos\theta\left(1 - \frac{a^3}{r^3}\right).$$

(a) Verify that the potential of the sphere is zero. (b) Show that at very large distances the potential corresponds to that of a uniform field. (c) Note that the potential V is the sum of a potential for a uniform field and a potential for an electric dipole. Obtain the electric dipole moment of the sphere. (d) Obtain the radial and transverse components of the electric field. (e) Verify that the electric field at the surface of the conductor is perpendicular to it. (f) Plot the lines of force of the resultant electric field. (g) Find the surface charge density. Discuss its variation over the surface of the sphere. (h) Verify that the total charge on the sphere is zero. (i) Show that at the center of the sphere the electric field produced by the surface charge is $-\mathscr{E}_0$. The same situation occurs for any point inside the sphere. Was this outcome to be expected?

2.37 Using the result of problem 2.17, show that the capacity of the system in Fig. 2-46 is $0.618C_1$. [*Hint:* Note that, if the system is cut along the dashed line, the section to the right is still equal to the original system because it is composed of an infinite number of capacitors.]

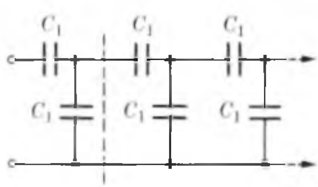


Figure 2-46

2.38 A dielectric slab is partially introduced between the two plates of a parallel-plate capacitor as shown in Fig. 2-47. Calculate as a function of x (a) the capacity of the system, (b) the energy of the system, and (c) the force on the slab. Assume that the potential applied to the capacitor is constant. [*Hint:* Note that the system may be considered as two capacitors in parallel.]

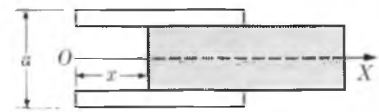


Figure 2-47

2.39 The plates of a parallel-plate capacitor in a vacuum have charges $+Q$ and $-Q$, and the distance between the plates is x. The plates are disconnected from the charging voltage and pulled apart a short distance dx. (a) What is the change dC in the capacity of the capacitor? (b) What is the change dE_g in the energy of the capacitor? (c) Equate the work $F\,dx$ to the increase in energy dE_g, and find the force of attraction F between the plates. (d) Explain why F is not equal to $Q\mathscr{E}$ where $\mathscr{E}$ is the electric field strength between the plates.

2.40 Rework Problem 2.39 for a case in which the electric potential V is kept constant.

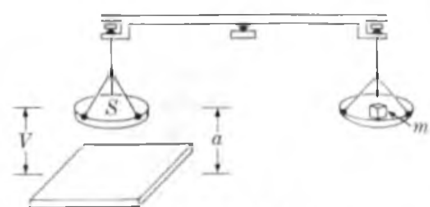


Figure 2-48

2.41 An *electrometer*, diagrammed in Fig. 2-48, is used to determine electric potential differences. It consists of a balance whose left pan is a disk of area S placed at a distance a from a horizontal plane to form a capacitor. When a potential difference is applied between the disk and the plane, a downward force is produced on the disk. To restore the balance to equilibrium, a mass m is placed in the other pan. Show that $V = a\sqrt{2mg/\epsilon_0 S}$. [*Note:* In the actual instrument, the disk is surrounded by a ring kept at the same potential to assure that the field is uniform over all the disk.]

2.42 Four capacitors are arranged as shown in Fig. 2-49. A potential difference ΔV is applied between the terminals A and B and a meter M is connected between C and D to determine their potential difference. Show that the meter reads zero if $C_1/C_2 = C_3/C_4$. This is a bridge arrangement that may be used to determine the capacitance of a capacitor in terms of a standard capacitor and the ratio of two capacitances.

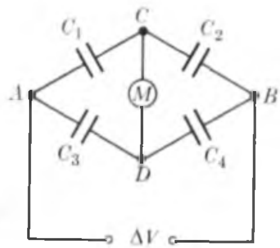


Figure 2-49

CHAPTER THREE

ELECTRIC CIRCUITS

3.1 Introduction

A most important electric phenomena from the point of view of its many practical applications is the motion of electric charge through a substance and the resulting generation of an electric current. The magnitude of an electric current was defined in Section 1.9 as the amount of charge passing through a cross section of the material per unit time and per unit area. The SI unit for electric current is the *ampere*. In this chapter we investigate the case in which the substance is a solid metal and the current has a constant magnitude; at the end of the chapter we briefly consider the passage of an electric current through other substances. The set of conductors and the sources of the electric field required to keep the charges moving through them constitute an *electric circuit*. Three important concepts of electric circuits will be discussed in this chapter: the electric resistance of a conductor, the electromotive force applied to a circuit, and Ohm's Law.

3.2 Electrical Conductivity; Ohm's Law

In Chapter 2 certain aspects of the behavior of a substance under an applied electric field were considered. This behavior has been represented by the electric susceptibility of the material. *Electrical conductivity*, another important property related to an external electric field, will now be discussed in connection with electrical conduction in a metal.

When an electric field is applied to a dielectric, a polarization of the dielectric results. However if the field is applied in a region where free charges exist, the charges are set in motion and an electric current, instead of a polarization of the medium, results. The charges are accelerated by the field and therefore gain energy.

When free charges are present within a body as electrons are in a metal, their motion is hindered by their interaction with the positive ions that form the crystal lattice of the metal. Consider, for example, a metal with the positive ions regularly arranged in three dimensions as in Fig. 3-1. The free electrons move in an electric

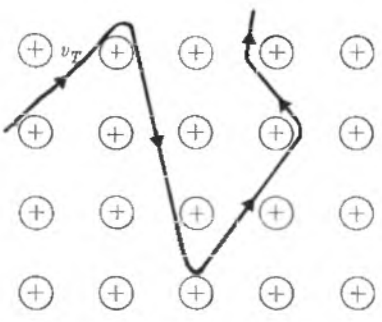


Fig. 3-1. Electron motion through the crystal lattice of a metal. In the figure, v_T is the thermal velocity of the electrons.

field that exhibits the same periodicity as the lattice. During their motion the free electrons are very frequently scattered by the field. A correct description of this type of electronic motion must use the methods of quantum mechanics. Here a simple classical description will have to suffice. When no external electric field is present, the electrons move in all directions, and no net charge transport or electric current results. However if an external electric field is applied, a drift motion is superposed on the natural random motion of the electrons; and an electric current results. It seems natural to assume that the strength of the current must be related to the intensity of the electric field, and that the relation must be a direct consequence of the internal structure of the metal.

For the clue to this relation let us first turn to the experimental results. One of the laws of physics that is perhaps most familiar to the student is *Ohm's law*, which states that *for a metallic conductor at constant temperature the ratio of the potential difference ΔV between two points of the conductor to the electric current I through the conductor is constant*. This constant is called the *electrical resistance R* of the conductor between the two points. Thus Ohm's law may be expressed as

$$\frac{\Delta V}{I}=R \qquad \text{or} \qquad I=\frac{\Delta V}{R}. \tag{3.1}$$

This law, formulated by the German physicist Georg Ohm (1787–1854), is obeyed with surprising accuracy by many conductors over a wide range of values of ΔV, I, and temperatures of the conductor. However, many substances, especially the semiconductors, do not obey Ohm's law. A graph of the relation between ΔV and I given by Ohm's law yields the straight line shown in Fig. 3-2.

From Eq. (3.1) it is seen that R is expressed in SI units as volts/ampere or $m^2\ kg\ s^{-1}\ C^{-2}$, a unit called an *ohm* (Ω). Thus one ohm is the resistance of a conductor through which there is a current of one ampere when a potential difference of one volt is maintained across the ends of the conductor.

Consider now a cylindrical conductor of length l and cross section S (Fig. 3-3). The current may be expressed as $I=jS$ where j is the current density. The electric field along the conductor is $\mathscr{E}=\Delta V/l$. (Remember Eq. 1.18.) Therefore Eq. (3.1)

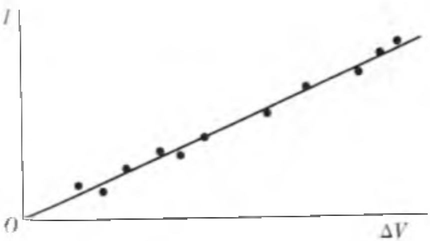


Fig. 3-2. Graphical statement of Ohm's law.

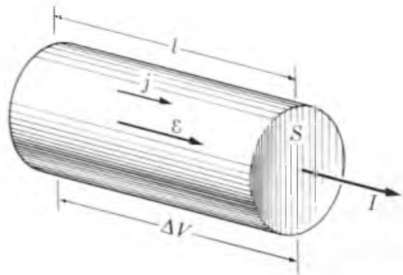


Figure 3-3

may be written in the form $\mathscr{E}l = RjS$ or

$$j=\left(\frac{l}{RS}\right)\mathscr{E}=\sigma\mathscr{E} \tag{3.2}$$

where $\sigma = l/RS$ is a new constant called the *electrical conductivity* of the material, and is expressed in $\Omega^{-1}\ m^{-1}$ or $m^{-3}\ kg^{-1}\ s\ C^2$. The relation between σ and R is more frequently written in the form

$$R=\frac{l}{\sigma S}. \tag{3.3}$$

Table 3-1 gives the electrical conductivity of several materials. The reciprocal of the conductivity is called the *resistivity*, ρ; that is

$$\rho=\frac{1}{\sigma}.$$

Resistivity is expressed in Ω m.

Equation (3.2) expresses a relation between the magnitudes of the vectors $\boldsymbol{j}$ and $\boldsymbol{\mathscr{E}}$. If they have the same direction, a situation found in most substances, Eq. (3.2) may be replaced by the vector equation

$$\boldsymbol{j}=\sigma\boldsymbol{\mathscr{E}}, \tag{3.4}$$

which is merely another way of writing Ohm's law.

3.3 Origin of Electric Resistance

If we use Eq. (3.4) and recall the definition of current density from Example 2.1, that $\boldsymbol{j} = -en\boldsymbol{v}_e$ where n is the number of electrons per unit volume and $\boldsymbol{v}_e$ is the electrons' drift velocity caused by the applied electric field $\boldsymbol{\mathscr{E}}$, the drift velocity may be written

$$\boldsymbol{v}_e=-\frac{\sigma}{en}\boldsymbol{\mathscr{E}}. \tag{3.5}$$

This equation shows that the conduction electrons in the metal attain a *constant* drift velocity as a result of the external applied electric field. This conclusion is quite different from that reached in the discussion of the motion of an ion along the evacuated tube of an accelerator (Section 1.7). There it was found that the acceleration is $\boldsymbol{a} = -(e/m)\boldsymbol{\mathscr{E}}$, resulting in a velocity $\boldsymbol{v} = -(e/m)\boldsymbol{\mathscr{E}}t$, which increases continuously with time.

However this is not the first time a situation like this has been encountered. For example, a freely falling body in vacuum has a velocity $\boldsymbol{v} = \boldsymbol{g}t$ that increases continuously with time; but if the body falls through a viscous fluid, the motion becomes uniform with a constant limiting velocity (as discussed in Section 7.9 of Volume I).

Table 3-1. Electrical Conductivities at Room Temperature

Substance	σ, Ω^{-1} m^{-1}	Substance	σ, Ω^{-1} m^{-1}
Metals		*Semiconductors*	
Copper	5.81×10^7	Carbon	2.8×10^4
Silver	6.14×10^7	Germanium	2.2×10^{-2}
Aluminum	3.54×10^7	Silicon	1.6×10^{-5}
Iron	1.53×10^7		
Tungsten	1.82×10^7	*Insulators*	
		Glass	10^{-10} to 10^{-14}
Alloys		Lucite	$< 10^{-13}$
Manganin	2.27×10^6	Mica	10^{-11} to 10^{-15}
Constantan	2.04×10^6	Quartz	1.33×10^{-18}
Nichrome	1.0×10^6	Teflon	$< 10^{-13}$
		Paraffin	3.37×10^{-17}

It may be said by analogy that the effect of the crystal lattice may be represented by a "viscous" force, acting on the conduction electrons when their natural motion is disturbed by the applied electric field. The exact nature of this viscous force depends on the dynamics of the electronic motion through the crystal lattice; this concept will be elaborated in Example 3.1; here a general qualitative analysis is given.

For a perfect crystal lattice with all positive ions at rest in fixed, regularly spaced positions, it can be proved with the methods of quantum mechanics that a conduction electron moves freely through the lattice under the action of an external field. However, no metal is composed of a perfect crystal lattice. In some instances the imperfection is due to impurities that replace some of the metal ions (Fig. 3-4); in other cases

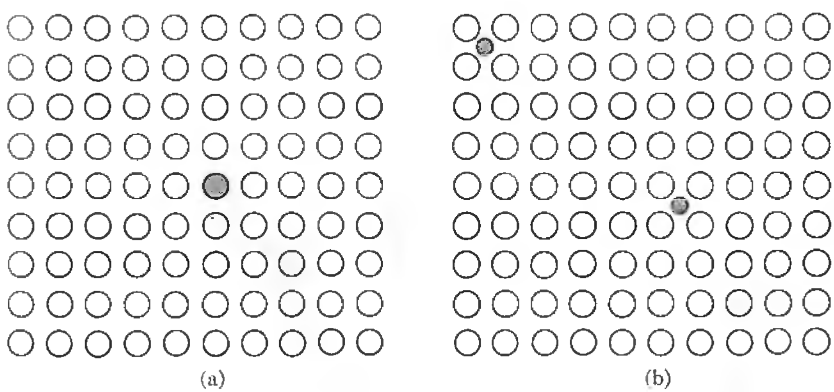
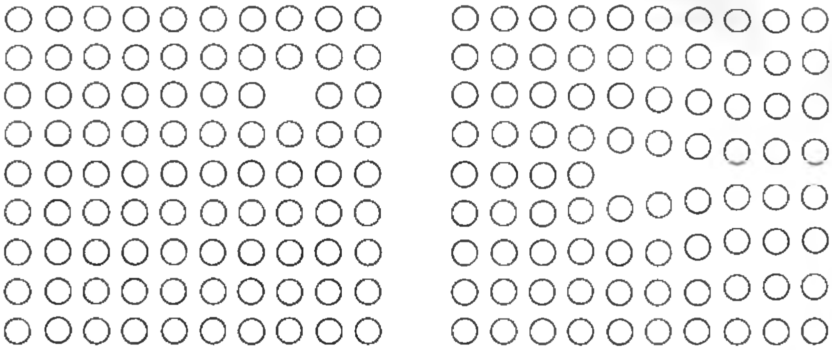


Fig. 3-4. Crystal imperfections due to impurities. (a) Substitutional impurity. (b) Interstitial impurity.



(a) (b)

Fig. 3-5. Crystal lattice imperfections due to missing atoms. (a) Vacancy site. (b) Edge dislocation.

some ions may be missing (Fig. 3-5). In addition the ions are always vibrating as a result of their thermal energy. Since the ions do not vibrate in phase, the distances between the ions fluctuate; this fluctuation is equivalent to imperfections in the crystal lattice. The result of these conditions is that electron motion is hindered; the electron suffers numerous scatterings and sometimes even moves backwards. Therefore rather than picking up energy continuously from the electric field, the electron transfers some energy to the lattice. After a short period of time a steady state is reached in which the average velocity of the electron becomes constant and has a value equal to the drift velocity given by Eq. (3.5).

It seems reasonable to assume that the effect on the resistivity because of the imperfections resulting from impurities or lattice irregularities is temperature independent. Likewise, it is reasonable to assume that thermal oscillations of the ions must increase the resistivity with temperature since the amplitude of the oscillations also increases with the temperature. These assumptions have been verified experimentally, and in most substances the resistance increases as the temperature increases.

Example 3.1. Motion of the conduction electrons in a metal.

▼ The effect of the interaction of the crystal lattice and the conduction electrons in a metal may be represented phenomenologically by a "viscous" force. If this force is of the same form as that considered in the case of motion in a fluid (Section 7.9 in Volume I), that is, $-kv$, the equation of motion of an electron in a metal is written as

$$m_e \frac{dv}{dt} = -e\mathscr{E} - kv. \tag{3.6}$$

Thus the limiting drift velocity, obtained by making $dv/dt = 0$, is $v_e = -e\mathscr{E}/k$. If we compare this result with Eq. (3.5), the electrical conductivity is $\sigma = ne^2/k$.

This result may be expressed in a different way by introducing a quantity called *relaxation time*. Suppose that the electric field $\mathcal{E}$ is suddenly cut off after the limiting drift velocity has been attained. The equation of motion for the electron is then

$$m_e \frac{d\mathbf{v}}{dt} = -k\mathbf{v},$$

whose solution is $\mathbf{v} = \mathbf{v}_e e^{-(k/m)t}$. The student may check this result either by direct substitution or by direct integration. Then the time required for the drift velocity and therefore the current to drop by the factor e is $\tau = m/k$. This is the relaxation time of the electron's motion, similar to that (introduced in Example 7.9 of Volume I) for the motion of a body through a viscous fluid. Thus for the conductivity the relation becomes

$$\sigma = \frac{ne^2\tau}{m_e}. \tag{3.7}$$

If σ is known, τ can be computed, and conversely since n, e, and m_e are known quantities. If each atom contributes one valence electron to the current, the value of n is about 10^{28} electrons m^{-3} in most metals. From the values of e and m_e with σ of the order of $10^7\ \Omega^{-1}\ m^{-1}$, the relaxation time τ is of the order of 10^{-14} s.

It must be understood at this point that the only thing that has been done is to devise a phenomenological model by which the result required by Ohm's law is obtained; but this has led to introducing a new quantity τ. To "explain" Ohm's law and electrical conduction in metals, we must relate τ to the dynamics of the motion of electrons. However as indicated before, since this motion takes place according to the laws of quantum mechanics, further discussion of Eq. (3.7) must be postponed.

However, the correctness of our model may be estimated by checking the orders of magnitude of the quantities involved. It is reasonable to assume that the relaxation time is of the same order of magnitude as the time between two successive collisions of an electron with the ions of the crystal lattice. If l is the average separation of the ions and v is the average velocity of the electrons, the collision time can be estimated by the ratio l/v. For most solids l is of the order of 5×10^{-9} m. To obtain v, assume that the same relation devised for gas molecules may be used; that is, $v = \sqrt{3kT/m_e}$. Thus at room temperature v is of the order of 10^5 m s^{-1}. Then τ is about 5×10^{-14} s. This result agrees with the estimates made previously using Eq. (3.7) and the experimental values of σ. ▲

3.4 The Joule Effect

Maintaining a current in a conductor requires the expenditure of energy. Energy must also be expended to accelerate an ion in an accelerator or an electron tube, but there is a difference. In the accelerator all the supplied energy is spent in speeding up the ions. In a conductor, because of the interaction of the electrons and the positive ions of the crystal lattice, the energy supplied to the electrons is transferred to the lattice, and thus increases the vibrational energy of the lattice. The resulting increase

Fig. 3-6. Symbolic representation of a resistor.

in the temperature of the material is the well-known heating effect of a current and is called the *Joule effect*.

The rate at which energy is transferred to the crystal lattice may be easily estimated. The work done per unit time on an electron is $\boldsymbol{F}\cdot\boldsymbol{v}_e = -e\boldsymbol{\mathscr{E}}\cdot\boldsymbol{v}_e$, and the work done per unit time and unit volume (or *power* per unit volume) is $P_{\text{vol}} = n(-e\boldsymbol{\mathscr{E}}\cdot\boldsymbol{v}_e)$. Equations (3.2) and (3.5) serve to eliminate $\boldsymbol{v}_e$;

$$P_{\text{vol}} = \sigma\mathscr{E}^2 = j\mathscr{E}. \tag{3.8}$$

Again consider Fig. 3-3, in which the cylindrical conductor has a volume of Sl. The power required to maintain the current in the conductor is

$$P = (Sl)P_{\text{vol}} = (Sl)(j\mathscr{E}) = (jS)(\mathscr{E}l).$$

However $jS = I$ and $\mathscr{E}l = \Delta V$. Therefore the power required to maintain the current in the conductor is

$$P = I\Delta V. \tag{3.9}$$

This equation is identical to Eq. (1.29), which was obtained in a more general way, and is independent of the nature of the conduction process. For conductors that follow Ohm's law, $\Delta V = RI$, and Eq. (3.9) may be written in the alternate form

$$P = RI^2. \tag{3.10}$$

Many materials, however, do not follow Ohm's law; and for them Eq. (3.10) is not correct although Eq. (3.9) remains valid. A conductor with resistance, also called a *resistor*, is represented diagrammatically in Fig. 3-6.

Example 3.2. Combination of resistors.

▼ Resistors can be combined in two kinds of arrangements, similar to those discussed in Example 2.12 for capacitors: *series* and *parallel*. In the series combination (Fig. 3-7a), the resistors are connected in such a way that the same current I is present in each of them. According to Ohm's law, the potential drop across each resistor is $\Delta V_1 = R_1 I$, $\Delta V_2 = R_2 I, \ldots, \Delta V_n = R_n I$. Thus the overall potential difference is

$$\Delta V = \Delta V_1 + \Delta V_2 + \cdots + \Delta V_n = (R_1 + R_2 + \cdots + R_n)I.$$

The system can be reduced effectively to a single resistor R satisfying $\Delta V = RI$. Therefore

$$R = R_1 + R_2 + \cdots + R_n \tag{3.11}$$

gives the resultant resistance for a series arrangement of resistors.

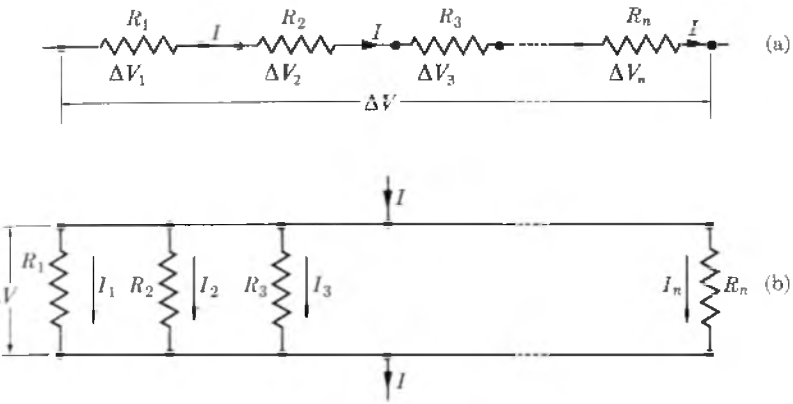


Fig. 3-7. (a) Series and (b) parallel arrangements of resistors.

In the parallel combination (Fig. 3-7b), the resistors are connected in such a way that the potential difference ΔV across each resistor is the same for all of them. According to Ohm's law, the current through each resistor is $I_1 = \Delta V/R_1$, $I_2 = \Delta V/R_2, \ldots, I_n = \Delta V/R_n$. The total current I supplied to the system is

$$I = I_1 + I_2 + \cdots + I_n = \left(\frac{1}{R_1} + \frac{1}{R_2} + \cdots + \frac{1}{R_n}\right)\Delta V.$$

The system can be reduced effectively to a single resistor R satisfying $I = \Delta V/R$. Therefore

$$\frac{1}{R} = \frac{1}{R_1} + \frac{1}{R_2} + \cdots + \frac{1}{R_n} \tag{3.12}$$

gives the resultant resistance for a parallel arrangement of resistors. ▲

3.5 Conductors, Insulators, and Semiconductors

Not all substances are conductors of electricity; and among the conducting substances, not all follow Ohm's law. Only a few substances, mainly metals, are good conductors of electricity and obey Ohm's law. This situation stems from the fact that metals are composed of atoms that have an electronic structure consisting of filled shells plus one, two, or perhaps three valence electrons in a level beyond the outermost, completely filled shell. In the solid state the valence electrons of metals are easily excited, become detached from the atoms, and constitute a sort of free-electron gas pervading the space between the lattice ions. Even a small electric field applied to the metal sets these electrons in motion and produces an electric current.

In many substances, however, the valence electrons are "frozen" in fixed positions between the positive lattice ions because the valence electrons are responsible for the bonds among the ions in the solid much as electrons bind atoms in a molecule together. The lack of free electrons in these substances makes it very difficult to produce a current within them; these substances are designated *dielectrics* or, in the limit of absolutely no free carriers, *insulators*. Under the action of an external field a dielectric becomes polarized as explained in Chapter 2, but no electric current is produced. However when the electric field is very strong, some valence electrons may be pulled from their positions; and a current is produced that often results in permanent damage to the dielectric. This process is called *dielectric breakdown*.

There is another class of substances, called *semiconductors*, in which it is relatively easy to excite some of the bound electrons because the excitation energy is about one-fifth to one-tenth that of the excitation energy in an insulator. (Among the most widely used semiconductors are silicon, germanium, and tellurium.) An increase in temperature may result in freeing some electrons from their bonds as a result of interactions between neighboring ions. In such a case an electric current is produced when an electric field is applied. Semiconductors do not obey Ohm's law, and their resistivity decreases as the temperature increases because more electrons become available with increasing temperature. For example in silicon a temperature increase from 250 K to 450 K increases the number of excited electrons by a factor of 10^6, and results in an increase in the conductivity and a consequent decrease of the electric resistance.

It is interesting to note that when an electron in a semiconductor is removed from the bond the electron occupied, it leaves behind a "hole" that behaves like a positive electron (Fig. 3-8). A valence electron from a neighboring bond can then jump into the hole; the jumping electron in turn leaves a hole in the bond it previously occupied. Since this process can be repeated many times, especially in the presence of an electric field, the process amounts to the movement of the positive hole through the substance. Thus a current in a semiconductor is due both to free electrons and to holes. This phenomenon is called *intrinsic conductivity*.

Another way for enhancing the conductivity of a semiconductor is by adding certain impurities. Two different situations arise in this case: and the impurities are called either *donors* or *acceptors*, depending upon the effect on the semiconductor. A donor impurity is composed of atoms having *more* valence electrons than the

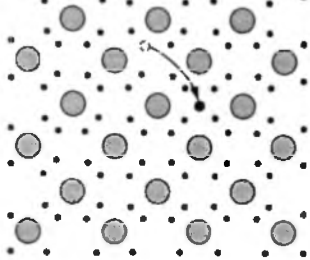


Fig. 3-8. The motion of an electron (●) leaves a hole (○) behind in a semiconductor.

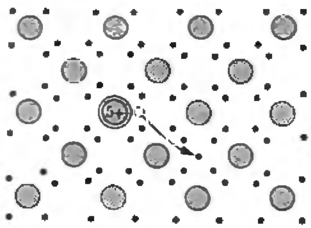


Fig. 3-9. Semiconductor with a donor impurity.

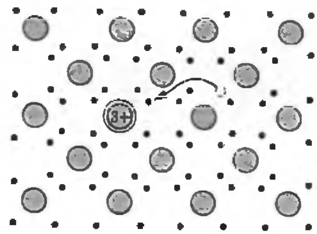


Fig. 3-10. Semiconductor with an acceptor impurity.

host lattice has; these extra electrons require very little energy to be set free and become conduction electrons. For example if in a silicon or germanium crystal a few atoms of Si or Ge (with four valence electrons) are replaced by phosphorus or arsenic atoms (with five valence electrons), the extra electrons that cannot be accommodated in the valence bonds associated with the crystal structure can be easily set free to move through the lattice (Fig. 3-9). Conversely an acceptor impurity is composed of atoms having *fewer* valence electrons than the atoms of the substance have; this situation produces bonds with one electron missing (i.e., with an electron hole). This is the case if a few atoms of boron or aluminum (with three valence electrons) are added to silicon or germanium. Under an external field an electron from a nearby bond can move to fill the hole and thereby leave a hole behind. Thus the hole can be said to move through the lattice (Fig. 3-10) just as a positive electron would.

Semiconductors with donor impurities are called n-type, and semiconductors with acceptor impurities are called p-type. Semiconductors have wide industrial application as rectifiers, modulators, detectors, photocells, transistors, etc. In fact the electronics industry today manufactures virtually all its devices with semiconductor elements.

Example 3.3. Discussion of the p–n junction.

▼ One important application of semiconductors to modern electric circuitry is the p–n junction. Suppose that there are two samples of the same semiconductor—say germanium—one of p-type and the other of n-type (Fig. 3-11a). If the two samples are placed in contact (Fig. 3-11b), there is a diffusion, or flow, of holes from the left to the right and of electrons from the right to the left. This double flow produces a double layer of positive and negative charges on both sides of the junction, and sets up a potential difference across the junction (as shown on the right in Fig. 3-11 b); when equilibrium is reached, the potential difference opposes further flow of holes and electrons across the junction. The following discussion will concentrate on the holes; the situation for the electrons is just the reverse.

Because of the recombination of holes and electrons, the number of holes in the n-type semiconductor tends to decrease. This decrease allows a small, continuous hole current I_1 from the p-side to the n-side. At the same time, because of thermal excitation, hole–electron pairs are

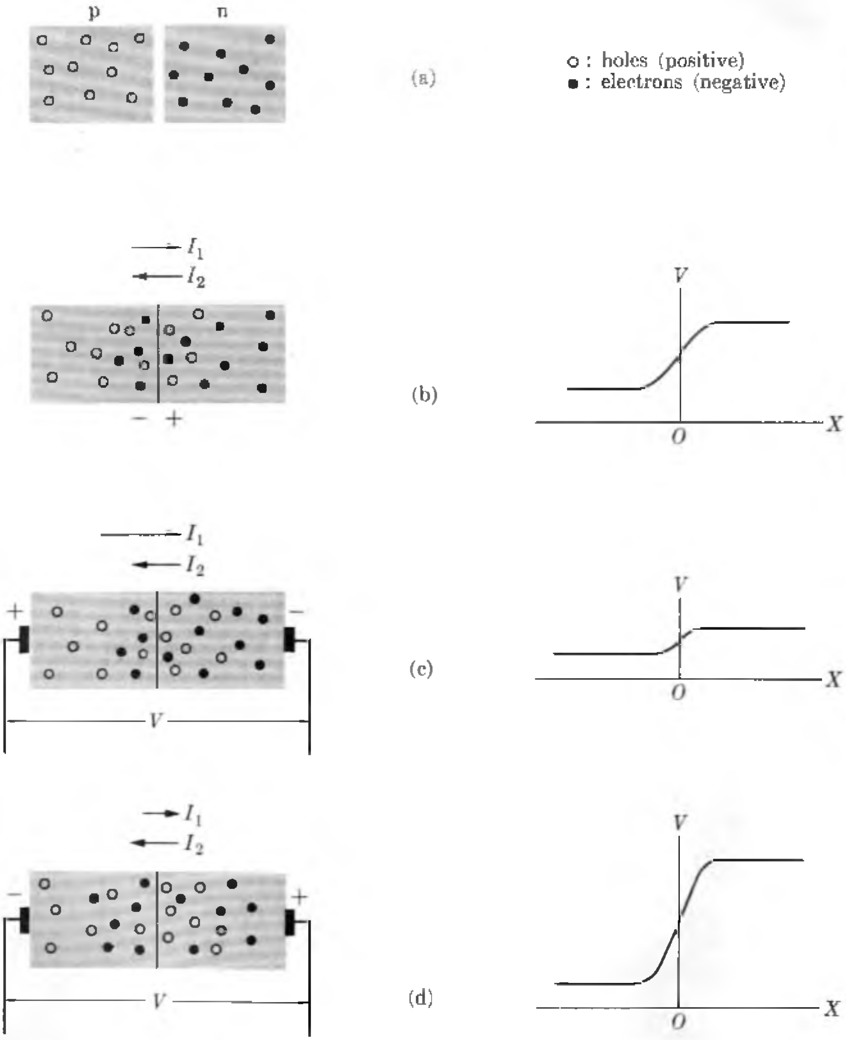


Fig. 3-11. The p-n junction.

produced in the n-type semiconductor, and these excess holes can flow very readily across the junction into the p-side with a current I_2. At equilibrium both hole currents are identical; that is, $I_1 = I_2$ (similar logic can be applied to the electrons). If a potential difference ΔV is applied as shown in Fig. 3-11(c) with the p-side joined to the positive terminal and the n-side to the negative terminal of the source of the potential difference ΔV, the height of the potential difference across the junction decreases. This decrease allows a larger current I_1 to the right without actually

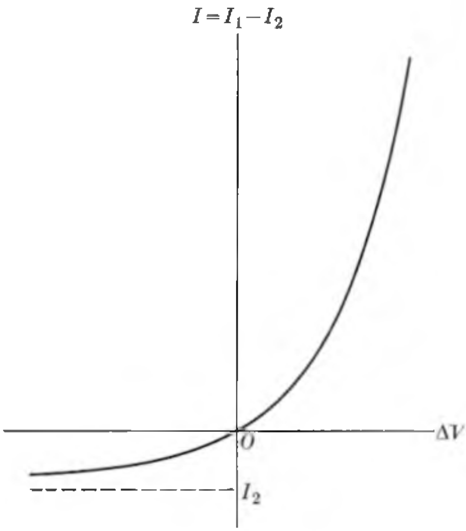


Fig. 3-12. Current as a function of voltage across a p–n junction. The voltage V is considered positive when applied in the direction p→n and negative when applied in the opposite direction.

changing the thermally generated current I_2 to the left. Thus a net hole current $I_1 - I_2$ results across the junction to the right, and this current increases very rapidly with ΔV because of the large supply of holes from the p-side. On the other hand if the potential difference ΔV is reversed as in Fig. 3-11(d), the potential difference across the junction increases. This increase reduces the value of I_1 again without substantially affecting I_2 since the supply of holes from the n-side is temperature limited. Thus a net current to the left will exist across the junction. This current will approach the constant value I_2 with increasing ΔV.

Figure 3-12 shows the graph of the net current across the junction as a function of ΔV, with ΔV considered positive when applied as in Fig. 3-11(c) and negative otherwise. The net current is expressed fairly accurately by the expression

$$I = I_1 - I_2 = I_2(e^{\Delta V/kT} - 1).$$

It is thus seen that a p–n junction acts as a *rectifier* or a detector device favoring the passage of a current in the direction p→n. This function is the same as that performed by diode and triode electron tubes, but the p–n junction performs the function with considerably less expenditure of energy. ▲

3.6 Electromotive Force

Suppose that a particle moves from A to B along a path L under the action of a force $\boldsymbol{F}$. The work done by the force is $W = \int_L \boldsymbol{F} \cdot d\boldsymbol{l}$ where the subscript L means that the integration is performed along the path and $d\boldsymbol{l}$ is a line element of the path. When the force is conservative (i.e., the force is related to the potential energy by

$F = -\text{grad } E_p$), the work is independent of the path and results in $\int F \cdot dl = E_{p,A} - E_{p,B}$. An important consequence of this fact is that when the path is closed, the work of a conservative force is zero since point B is the same as point A and thus $E_{p,A} = E_{p,B}$.

These results can be extended to any vector field, such as the electric or magnetic fields. Designate the vector field by V. The *line integral* of the vector field V from point A to point B along a path L is defined as

$$\text{Line integral of } V = \int_L V \cdot dl. \tag{3.13}$$

In general the line integral depends on the path. If the path along which the line integral is calculated is a *closed* path, the line integral is called the *circulation* of the vector field and is indicated by a circle on top of the integral sign:

$$\text{Circulation of } V = \oint V \cdot dl. \tag{3.14}$$

An important case is that in which the field V can be expressed as the gradient of a function. This situation is the same as that found in the case of conservative forces; and therefore

> *when a vector field can be expressed as the gradient of a function, the line integral of the field between two points is independent of the path joining the points; and the circulation around an arbitrary closed path is zero.*

The student will discover that the concepts of line integral and circulation of a vector field are very useful in formulating the laws of electromagnetism. These two new definitions will now be applied to the electric field.

Since the electric field is equal to the force per unit charge, the line integral of the electric field, $\int_L \mathscr{E} \cdot dl$, is equal to the work done when moving one unit of charge along the path L. If the path is closed (Fig. 3-13), the line integral becomes the circulation of the electric field and is called the *electromotive force* (emf) applied to the closed path. With the emf designated by V

$$\text{emf} = V = \oint_L \mathscr{E} \cdot dl. \tag{3.15}$$

Therefore *the electromotive force applied to a closed path is equal to the work done by the electric field when moving one unit of charge around the path.* (The word "force" is misleading since we are referring to "energy," but the term has been accepted by common usage.) Naturally emf is expressed in volts.

Consider the special case of a *static* electric field. Recall that the static electric field is related to the electric potential by $\mathscr{E} = -\text{grad } V$; then the line integral of the electric field is

$$\int_L \mathscr{E} \cdot dl = \int_A^B -\text{grad } V \cdot dl = \int_B^A dV = V_A - V_B = \Delta V_{AB} \tag{3.16}$$

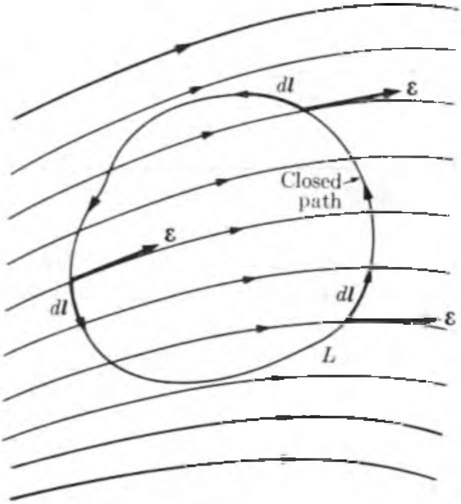


Figure 3-13

where A and B are the two points joined by the path L. Thus the line integral of a static electric field between two points is equal to the potential difference between the points. If the path is closed, points A and B coincide, and Eq. (3.16) gives

$$V=\oint \mathscr{E} < dl=0. \tag{3.17}$$

This may be expressed in words as

> *the emf, or circulation, of a static electric field around an arbitrary closed path is zero.*

This statement means that the work done by a static electric field in moving a charge around a closed path is zero, and by definition a static electric field is a conservative field.

If the electric field is applied to a conductor, Eq. (3.16) may be combined with Ohm's law; and Eq. (3.1) becomes

$$\Delta V=\int_L \mathscr{E}\cdot dl=RI \tag{3.18}$$

where L is a path along the conductor and R is the electric resistance between the points of the conductor joined by the path L.

As previously mentioned, maintaining a current between two points in a conductor implies that energy must be supplied to the system by the source of the potential difference. The question now arises as to whether or not a current can be maintained in a *closed* conductor or *electric circuit*. Ohm's law, which essentially describes

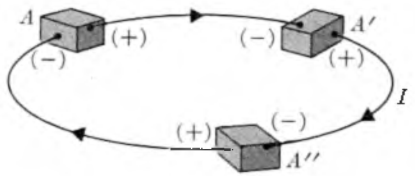


Fig. 3-14. An electric current is maintained in a closed circuit by electric generators.

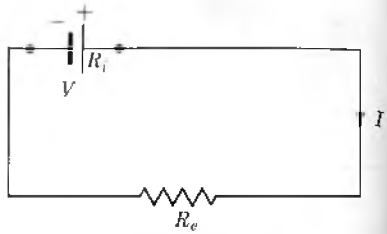


Fig. 3-15. Symbolic representation of a circuit with an electromotive force.

energy conservation in the conductor, when applied to a closed conductor is

$$V = \oint_L \mathscr{E} \cdot dl = RI. \tag{3.19}$$

The left-hand side of this equation is the emf applied to the circuit and R is the total resistance of the closed circuit.

If the conductor is placed in a *static* electric field, which is a time-independent and conservative field, then according to Eq. (3.17), we have that the emf is zero ($V=0$) and Eq. (3.19) gives $I=0$. In other words

a static electric field cannot maintain a current in a closed circuit.

The reason is that since a static electric field is conservative, the total net energy supplied to a charge describing a closed path is zero. However, a charge moving inside a conductor is transferring the energy received from the electric field to the crystal lattice, and this process is irreversible; that is, the lattice does not give the energy back to the electrons. Therefore, unless a net amount of energy is supplied to the electrons, they cannot move steadily around a closed circuit.

Accordingly, to maintain a current in a closed circuit it is necessary to feed energy into the circuit at certain places A, A', A'', . . . (Fig. 3-14). The suppliers of energy are called *electric generators* and may be considered as the sources of the emf. Therefore the electric field $\mathscr{E}$ appearing in Eq. (3.19) is not a static field, and at points A, A', A'' . . . , corresponds to local fields produced by the generators.

There are many ways of generating an electromotive force. A common method is by a chemical reaction, such as in a dry cell or a storage battery, in which the internal energy released in the chemical reaction is transferred to the electrons. Another important method is by the phenomenon of electromagnetic induction, to be discussed in a future chapter.

A source of emf is represented diagrammatically in Fig. 3-15, in which the sense of the current that is produced in the circuit *external to the source of emf* is from the long bar, or positive pole, to the short bar or negative pole.

When Ohm's law is applied to a simple circuit such as that of Fig. 3-15, it must be recognized that the total resistance R is the sum of the internal resistance R_i of the source of emf and the external resistance R_e of the conductor connected to the generator (or battery). Thus $R = R_i + R_e$, and Ohm's law becomes

$$V = (R_e + R_i)I \tag{3.20}$$

where V is the emf applied to the circuit. This equation may also be written in the form $V - R_i I = R_e I$. Each side of the equation gives the potential difference between the poles of the generator (or battery). Note that this potential difference between the terminals of the battery is smaller than the emf.

Example 3.4. Calculation of the currents in an electric network: Kirchhoff's laws.

▼ An *electric network* is a combination of conductors and emf's, such as the one illustrated in Fig. 3-16. We shall now consider only the case in which the emf's are constant, and steady conditions have been reached in the network so that the currents are also constant. Usually the problem consists in finding the currents in terms of the emf's and the resistances. The rules to solve this kind of problem, rules known as *Kirchhoff's laws*, merely express the conservation of electric charge and of energy. Kirchhoff's laws may be stated as follows:

(1) *The sum of all currents at a junction in a network is zero.*
(2) *The sum of all potential drops along any closed path in a network is zero.*

In writing the first law those currents directed away from the junction are considered positive and those directed toward the junction are considered negative. The first law expresses the conservation of charge: since charges are not accumulated at a junction, the number of charges that arrive at a junction in a certain time interval must leave it in the same time interval so that charges do not build up at the junctions.

In applying the second law, the following rules must be taken into account. (1) A potential drop across a resistance is considered positive or negative depending on whether the closed path is traced in the same sense as the current or in the opposite sense. (2) When passing through an emf, the potential drop is taken as negative or positive depending on whether we pass through in the direction the emf acts (increase in potential) or in the opposite direction (drop in potential). The second law expresses the conservation of energy since the net change in the energy of a charge after the charge completes a closed path must be zero. This requirement has already been met in Eq. (3.20), which reads $RI - V = 0$ for a single circuit where $R = R_i + R_e$.

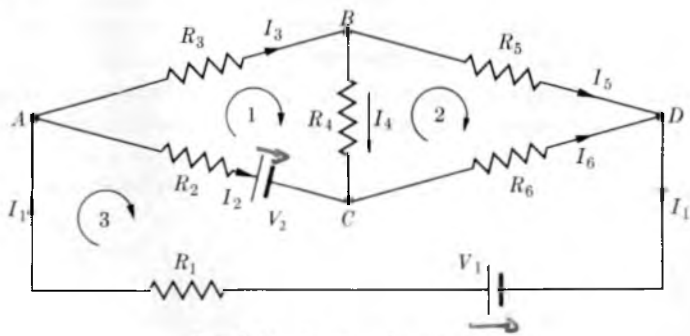


Fig. 3-16. An electric network.

The practical use of Kirchhoff's laws can be illustrated by applying them to the network of Fig. 3-16. The first law applied to junctions A, B, and C gives

$$\begin{aligned} &\text{Junction } A: && -I_1+I_2+I_3=0; \\ &\text{Junction } B: && -I_3+I_4+I_5=0; \\ &\text{Junction } C: && -I_2-I_4+I_6=0. \end{aligned}$$

The second law applied to the paths marked 1, 2, and 3 gives

$$\begin{aligned} &\text{Path 1:} && -R_2I_2+R_3I_3+R_4I_4-V_2=0; \\ &\text{Path 2:} && R_5I_5-R_6I_6-R_4I_4=0; \\ &\text{Path 3:} && R_1I_1+R_2I_2+R_6I_6-V_1+V_2=0. \end{aligned}$$

These six equations are enough to determine the six currents in the network.

A practical rule to follow in finding the currents in a network having n junctions is to apply the first law to only $n-1$ junctions because once the law is satisfied for $n-1$ junctions, it is automatically satisfied for the remaining junction. (The student should verify this statement for the network in Fig. 3-16.) The second law must be applied to as many closed paths as required in order for each conductor to be part of a path at least once. ▲

3.7 Nonohmic Conductors

Conducting materials that obey Ohm's law, $\Delta V=RI$ or $\boldsymbol{j}=\sigma\boldsymbol{\mathscr{E}}$, are designated as ohmic or linear conductors because of the direct proportionality between the voltage and the current. Most solid conductors (mainly metals) and many liquids follow Ohm's law to a very good approximation. On the other hand, gaseous conduction departs markedly from Ohm's law, primarily because conduction in gases is due not to the presence of free electrons but to the production of ionized atoms or molecules. These ions are produced by passing high-frequency electromagnetic radiation, such as x-rays or γ-rays, through the gas or by increasing the temperature of the gas to several thousand kelvin. When an electric potential difference is maintained across the gas, there will then be an electric current. Also metals depart from Ohm's law at very low temperatures because when the temperature is very low, the energy of the conduction electrons and of the lattice ions is extremely small. The motion of the electrons is less hindered by lattice vibrations, and conductivity increases appreciably. In some substances the conductivity effectively becomes infinite near absolute zero: such substances are called *superconductors*.

There are some conducting solids that do not obey Ohm's law even at room temperature. These solids are called nonohmic conductors. They are predominately ceramic compounds, semiconductors, p–n junctions, and boundary layers between metals and their oxides. Figure 3-17 shows the relation between the voltage and the current for some metallic filaments that are used in incandescent lamps. It should be noted that the current through the iron alloy filament in Fig. 3-17 is practically in-

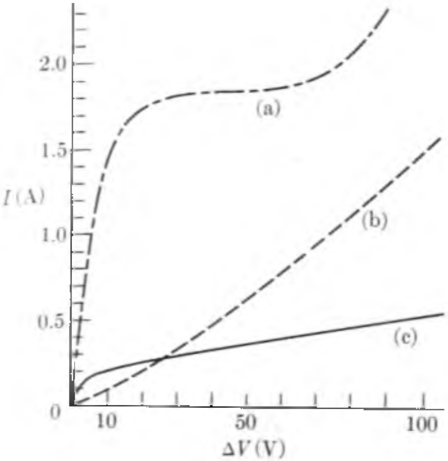


Fig. 3-17. Typical lamp-filament characteristics. (a) Iron alloy filament in hydrogen atmosphere; (b) carbon filament; (c) tungsten filament.

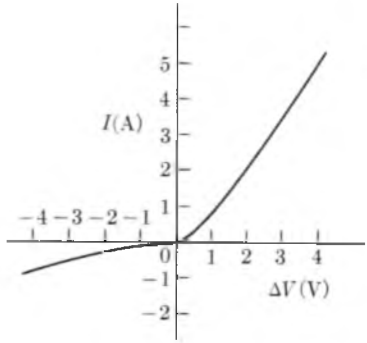


Fig. 3-18. Voltage-current relation in Cu-CuO system.

sensitive to voltage variations in the range of 30 to 60 V; this property has a number of interesting practical applications.

Furthermore some conducting systems are asymmetric in that the current depends not only on the magnitude of the applied voltage, but also on the direction in which the voltage is applied. One example of an asymmetric system is the p–n junction, for which the relation between ΔV and I has been illustrated in Fig. 3-13. A second example is a piece of metallic copper covered with a layer of copper oxide. The relation between ΔV and I is shown in Fig. 3-18. These asymmetric conducting elements find many applications, especially as rectifiers of alternating currents.

Problems

3.1 It has been estimated that there are about 10^{29} free electrons per cubic meter in copper. Using the value of copper's conductivity given in Table 3-1, estimate the relaxation time for an electron in copper.

3.2 How is the resistance of a wire changed if (a) the length is doubled, (b) the cross-sectional area is doubled, and (c) the radius is doubled?

3.3 (a) Determine the total resistance in the circuit shown in Fig. 3-19. Also determine (b) the current in and (c) the potential difference across each resistor.

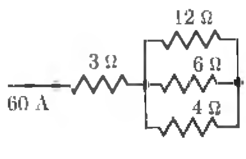


Figure 3-19

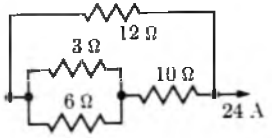


Figure 3-20

3.4 (a) Determine the total resistance in the circuit shown in Fig. 3-20. Also determine (b) the current in and (c) the potential difference across each resistor.

Figure 3-21

3.5 (a) Determine the total resistance in the circuit shown in Fig. 3-21. Also determine (b) the current in and (c) the potential difference across each resistor.

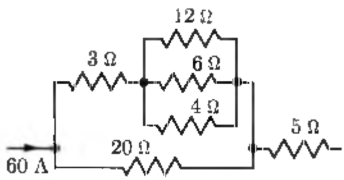


Figure 3-22

3.6 (a) Determine the total resistance in the circuit shown in Fig. 3-22. Also determine (b) the current in and (c) the potential difference across each resistor.

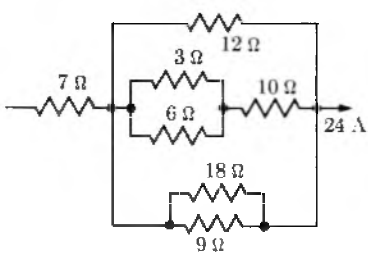


Figure 3-23

3.7 (a) Determine the total resistance of the circuit shown in Fig. 3-23. Also determine (b) the current in and (c) the potential difference across each resistor.

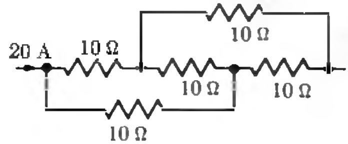


Figure 3-24

3.8 (a) Determine the total resistance of the circuit shown in Fig. 3-24. Also determine (b) the current in and (c) the potential difference across each resistor.

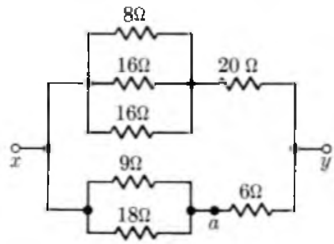


Figure 3-25

3.9 (a) Calculate the equivalent resistance between x and y of the circuit in Fig. 3-25. (b) What is the potential difference between x and a if the current in the 8-ohm resistor is 0.5 amp?

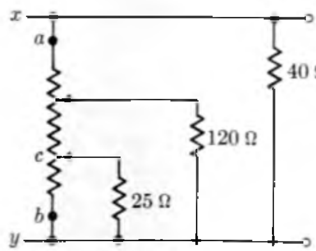


Figure 3-26

3.10 (a) The long resistor between a and b in Fig. 3-26 has a resistance of 300 ohms and is tapped at the one-third points. What is the equivalent resistance between x and y? (b) The potential difference between x and y is 320 volts. What is the potential difference between b and c?

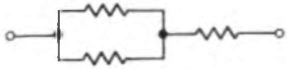


Figure 3-27

3.11 Each of the three resistors in Fig. 3-27 has a resistance of 2 ohms and can dissipate a maximum of 18 watts without becoming excessively heated. What is the maximum power the circuit can dissipate?

3.12 Three equal resistors are connected in series. When a certain potential difference is applied across the combination, the total power consumed is 10 watts. What power would be consumed if the three resistors were connected in parallel across the same potential difference?

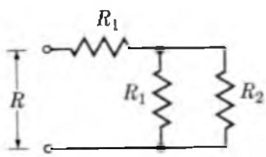


Figure 3-28

3.13 Given the resistor arrangement shown in Fig. 3-28, prove that the relation between R_1 and R_2 must be $R_2 = 1.618R_1$ in order that the resistance of the system be equal to R_2. [*Hint*: Recall Problem 2.17.]

3.14 The maximum permissible current in the coil of an electrical instrument is 2.5 A. Its resistance is 20 Ω. What must be done to the instrument so that it may be inserted in an electric circuit carrying a current of 15 A and not destroy the coil? An *ammeter* is a galvanometer modified so that it may be placed *in* a circuit and used to measure electric current.

3.15 How may the instrument given in Problem 3.14 (resistance of 20 Ω and maximum permissible current of 2.5 A) be modified so that it may be placed between two points having a potential difference of 110 V and not have the coil destroyed? A *voltmeter* is a galvanometer modified so that it may be placed *across* two points of an electric circuit and used to measure electric potential difference.

Table 3-2

I, amp	ΔV, volts
0.5	4.75
1.0	5.81
2.0	7.05
4.0	8.56

3.16 The measurements in Table 3-2 are for the current in and the potential differences between the ends of a wire of certain material. (a) Make a graph of ΔV versus I. Does the material follow Ohm's law? (b) From your graph estimate the resistance of the material when the current is 1.5 A. This resistance is defined as the ratio $\Delta V/\Delta I$ when the changes are small, and is obtained by drawing a tangent to the curve at the given point. (c) Compare your result with the average resistance between 1.0 A and 2.0 A.

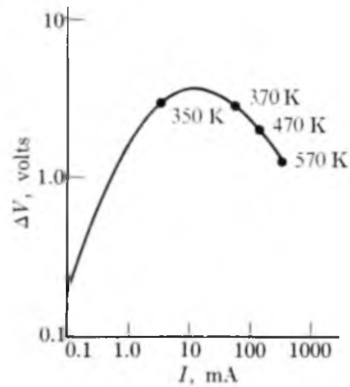


Figure 3-29

3.17 The graph in Fig. 3-29 illustrates potential difference versus current (on a logarithmic plot) for different temperatures of a semiconductor. (a) Estimate the resistance of the semiconductor at the temperatures marked. (b) Plot the log of the resistance against the temperature. (c) Assuming that the change in

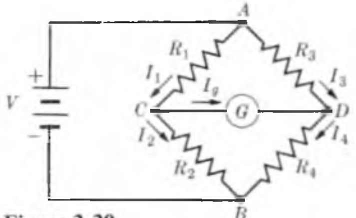


Figure 3-30

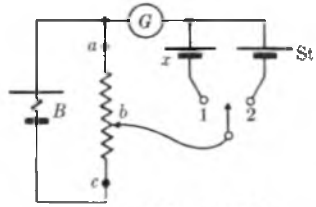


Figure 3-31

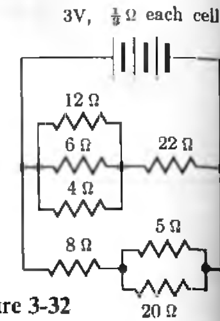


Figure 3-32

resistance is all due to a change in the number of charge carriers per unit volume, estimate the ratio of their number at 570 K to that at 370 K.

3.18 The circuit of Fig. 3-30 is called a *Wheatstone bridge*. It is used for measuring resistance. Show that when the current through the galvanometer G is zero (so that points D and C are at the same potential), then

$$\frac{R_1}{R_2}=\frac{R_3}{R_4}.$$

Thus if we know R_2 and the ratio R_3/R_4, we can obtain the resistance R_1.

3.19 Figure 3-31 shows a *potentiometer* set up to measure the emf V_x of cell x; B is a battery and St is a standard cell of emf V_{St}. When the switch is set at either 1 or 2, the tap b is moved until the galvanometer G reads zero. Show that if l_1 and l_2 are the corresponding distances from b to a, then $V_x = V_{St}(l_1/l_2)$.

3.20 Referring to the potentiometer of Fig. 3-31. The emf of B is approximately 3 V and its internal resistance is unknown; St is a standard cell of emf 1.0183 V. The switch is set at point 2; thus the standard cell is placed in the galvanometer circuit. When the tap b is 0.36 of the distance from a to c, the galvanometer G reads zero. (a) What is the potential difference across the entire length of resistor ac? (b) The switch is set at point 1, and a new zero reading of the galvanometer is obtained when b is 0.47 of the distance from a to c. What is the emf of cell x?

3.21 The potential difference across the terminals of a battery is 8.5 V when there is a current of 3 A in the battery from the negative to the positive terminal. When the current is 2 A in the reverse direction, the potential difference becomes 11 V. (a) What is the internal resistance of the battery? (b) What is its emf?

3.22 In the circuit of Fig. 3-32, determine (a) the current in the battery, (b) the potential difference at its terminals, and (c) the current in each conductor.

3.23 Determine the current in each conductor in the network shown in Fig. 3-33.

3.24 Determine the current in each conductor in the network shown in Fig. 3-34.

3.25 Determine the current in each conductor in the network shown in Fig. 3-35.

3.26 (a) Determine the potential difference between points a and b in Fig. 3-36. (b) Given that a and b are connected, calculate the current in the 12-V cell.

3.27 (a) In Fig. 3-37a what is the potential difference ΔV_{ab} when switch S is open? (b) What is the current through switch S when it is

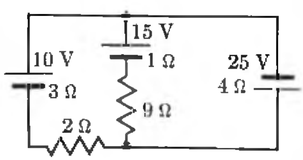


Figure 3-33

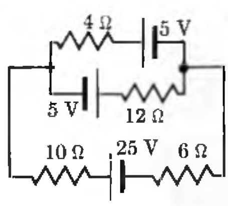


Figure 3-34

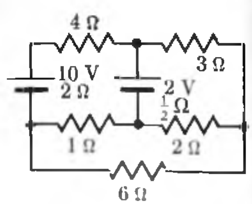


Figure 3-35

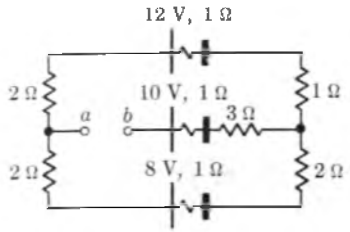


Figure 3-36

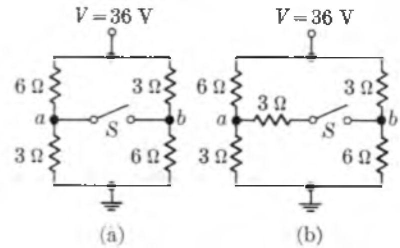


Figure 3-37

closed? (c) In Fig. 3-37b what is the potential difference ΔV_{ab} when switch S is open? (d) What is the current through switch S when it is closed? What is the equivalent resistance of the circuit in Fig. 3-37b, (e) when switch S is open and (f) when it is closed?

CHALLENGING PROBLEMS

3.28 Determine the charge on the plates of the capacitor in the circuit shown in Fig. 3-38. (AP-B: 1971)

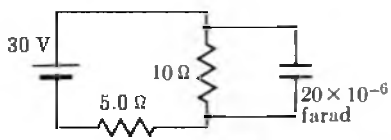


Figure 3-38

3.29 In the circuit shown in Fig. 3-39, all currents and voltages are at their steady-state values. (a) Calculate the current in the 10-ohm resistor. (b) Calculate the charge on either plate of the capacitor. (c) Calculate the power dissipation in the 2.0-ohm resistor. (AP-B: 1972)

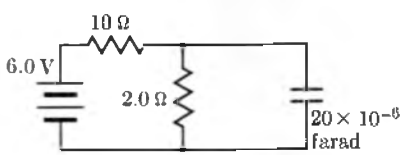


Figure 3-39

3.30 In the circuit shown in Fig. 3-40, the current delivered by the 9-volt battery of internal resistance 1 ohm is 3 amperes. The power dissipated in R_2 is 12 watts. (a) Determine the reading of voltmeter V in the diagram. (b) Determine the resistance of R_2. (c) Determine the resistance of R_1. (AP-B: 1976)

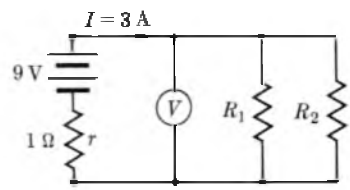


Figure 3-40

3.31 Suppose that you are provided with the apparatus shown in Fig. 3-41. (a) Draw a diagram, using the symbols in the figure, to show how you should connect these components to heat the water as rapidly as possible. The meters should be connected so that from the two meter readings alone you could determine at what rate the water is being heated. (b) Suppose the emf V of the battery is 50 volts and the current through the battery is 5 amperes. Assume the specific heat of the water is 4 joules per gram per Celsius degree, and the heat of vaporization is 2,200 joules

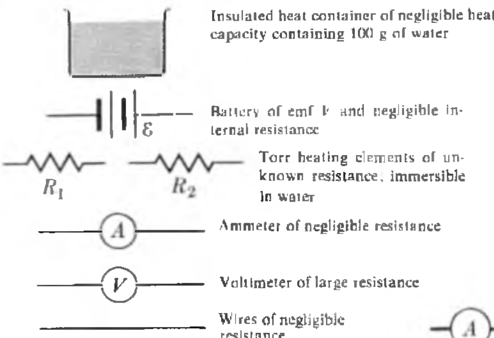


Figure 3-41

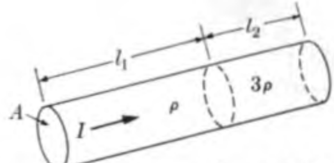


Figure 3-42

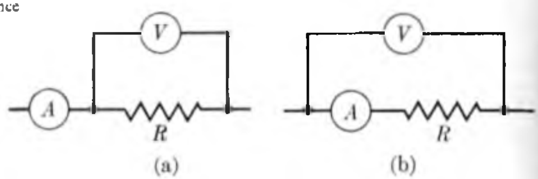


Figure 3-43

per gram. Calculate the number of seconds required for all the water to boil away. (AP-B: 1977)

3.32 A resistor is made in the form of a cylinder of cross-sectional area A. One portion, of length l_1, is made of material whose resistivity is ρ: the other, of length l_2, is made of material whose resistivity is 3ρ. There is a current I uniformly distributed over the area A. Express all answers in terms of fundamental constants and the symbols shown in Fig. 3-42. (a) Determine expressions for the electric field strengths $\mathscr{E}_1$ and $\mathscr{E}_2$ in the two portions of the resistor. (b) Determine the potential difference V between the opposite ends of the resistor. (c) By applying Gauss's law to a surface which encloses the boundary between the two materials, determine the sign and magnitude of the electric charge which is present on this boundary. (AP-C: 1977)

3.33 The needle of a galvanometer suffers a full-scale (50 divisions) deviation when the current is 0.1 mA. The resistance of the galvanometer is 5 Ω. What must be done to change it into (a) an ammeter with each division corresponding to 0.2 A, and (b) a voltmeter with each division corresponding to 0.5 V?

3.34 When resistance is measured using a voltmeter and an ammeter as shown in Fig. 3-43, errors are made if the resistances R_V and R_A of the instruments are ignored. Discuss these errors. Which method has the smaller error when R is (a) large, and (b) small? Note that in general R_V is very large and R_A is very small.

3.35 Using the result of Problem 3.13, show that the resistance of the system shown in Fig. 3-44 is equal to $1.618R_1$. [*Hint:* Note that if the system is cut through the dashed line, the section to the right is still equal to the original system because it is composed of an infinite number of resistors.]

3.36 A hollow cylindrical conductor of length L has radii r_1 and r_2. A potential difference is applied between the two ends of the conductor so that there is a current I parallel to the axis. Show that if σ is the conductivity of the material, the resistance is

$$R_{\text{eff}}=\frac{L}{\pi\sigma(r_2^2-r_1^2)}.$$

3.37 A hollow cylindrical conductor of length L has radii r_1 and r_2. A potential difference is applied between the inner and outer surfaces so that there is a current I in the outward radial direction. Show that if σ is the conductivity of the material, the resistance is

$$R_{\text{eff}}=\frac{\ln(r_2/r_1)}{2\pi\sigma L}.$$

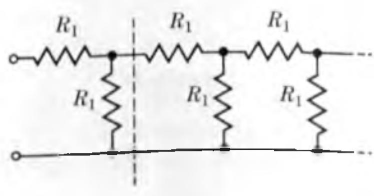


Figure 3-44

CHAPTER FOUR

MAGNETIC INTERACTION

4.1 Introduction

Centuries before Christ. it was observed that certain iron ores. such as the lodestone, have the property of attracting small pieces of iron. This property is also exhibited by iron. cobalt, and manganese, as well as by many compounds of these metals. This attribute is unrelated to gravitation since not only does the property fail to be exhibited naturally by all bodies, but it appears to be concentrated at certain spots in the mineral ore. This attribute is also apparently unrelated to the electrical interaction because neither cork balls nor pieces of paper are attracted at all by these minerals. Therefore a new name, *magnetism*,* was given to this physical property. The regions of a body where the magnetism appears to be concentrated are called *magnetic poles*. A magnetized body is called a *magnet*.

The earth itself is a huge magnet. For example if a magnetized rod is suspended at any point on the earth's surface and allowed to rotate freely about the vertical. the rod orients itself so that the same end always points toward the north geographic pole. This result shows that the earth exerts a force on the magnetized rod.

Experiment suggests also that there are two kinds of magnetic poles, which may be designated by the signs $+$ and $-$, or by the letters N and S, corresponding to the north-seeking and south-seeking poles, respectively. If two magnetized bars are placed as shown in Fig. 4-1, the bars will either repel or attract each other, depending on whether like or unlike poles face each other. Thus from experiment

> *the interaction between like magnetic poles is repulsive and the interaction between unlike magnetic poles is attractive.*

Before physicists clearly understood the nature of magnetism, they attempted to measure the strength of a magnetic pole by defining a *magnetic mass* or charge and then investigated the dependence of the magnetic interaction on the distance between the poles, much as gravitational interaction was first studied. However, a fundamental difficulty appeared when these measurements were attempted: although positive and negative electric charges have been isolated and a definite amount of electric charge is associated with the fundamental particles constituting all atoms, it has not yet been possible to isolate a magnetic pole or identify a fundamental particle having only one kind of magnetism, either N or S. Magnetized bodies always seem to exhibit poles in equal and opposite pairs. On the other hand the notions of magnetic pole and magnetic mass have been found unnecessary for the description of magnetism. Electric and magnetic interactions are intimately related, and in fact are only two different aspects of one property of matter: its electric charge. As will be seen, *magnetism is a manifestation of electric charge in motion*. Thus electric and magnetic

*The name magnetism is derived from the ancient city in Asia Minor called Magnesia, where according to tradition, the phenomenon was first recognized.

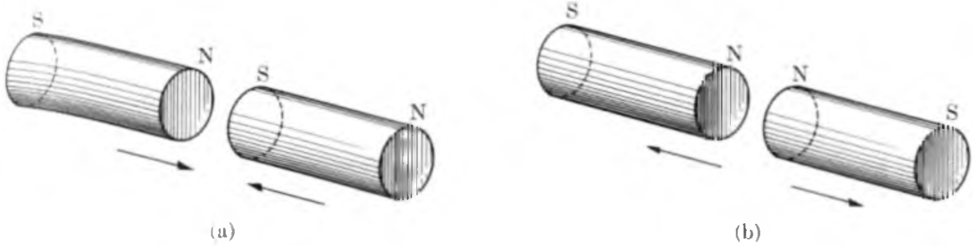


Fig. 4-1. Interaction between two magnetized bars. (a) Unlike poles attract each other. (b) Like poles repel each other.

interactions are often considered together under the more general name of *electromagnetic interaction.*

4.2 Magnetic Force on a Moving Charge

Interactions between magnetized bodies may be described by saying, in analogy with the gravitational and electrical cases, that a magnetized body produces a *magnetic field* in the space around it. When an electric charge *at rest* is placed in a magnetic field, no special force or interaction is observed on the charge. However when an electric charge *moves* in a region in which there is a magnetic field, a new force on the charge is observed in addition to those forces resulting from gravitational and electric interactions.

By measuring the force experienced by different charges moving in different ways at the same point in a magnetic field, a relation between the force, the charge, and its velocity may be deduced. The result is that

> *the force exerted by a magnetic field on a moving charge is proportional to the electric charge and to its velocity, and the direction of the force is perpendicular to the velocity of the charge.*

If the properties of the vector product are recalled, the analysis may be taken one step further by writing the force $\boldsymbol{F}$ on a charge q moving with velocity $\boldsymbol{v}$ in a magnetic field $\mathscr{B}$ as

$$\boldsymbol{F} = q\boldsymbol{v} \times \mathscr{B}. \tag{4.1}$$

which satisfies the experimental requirements mentioned above. Here $\mathscr{B}$ is a vector whose magnitude and direction are found at each point by comparing the observed value of $\boldsymbol{F}$ at the point with the values of q and $\boldsymbol{v}$. It is found experimentally that the vector $\mathscr{B}$ may vary from point to point in a magnetic field, but at any given point $\mathscr{B}$ has the same value for all charges and velocities. Therefore $\mathscr{B}$ describes a property that is characteristic of the magnetic field, and is called the *magnetic field strength.*

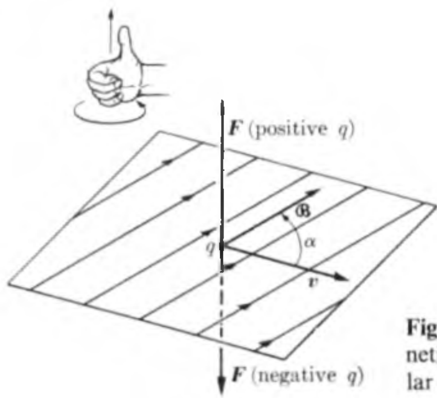


Fig. 4-2. Vector relation between magnetic force, magnetic field, and charge velocity. The force is perpendicular to the plane containing $\mathscr{B}$ and v.

in analogy with the electric field $\mathscr{E}$. (Another name, imposed by usage, is *magnetic induction*.)

When the particle moves in a region in which there are both electric and magnetic fields, the total force on the particle is the sum of the electric force $q\mathscr{E}$ and the magnetic force $q\boldsymbol{v}\times\mathscr{B}$. That is,

$$\boldsymbol{F}=q(\mathscr{E}+\boldsymbol{v}\times\mathscr{B}). \tag{4.2}$$

This expression is called the *Lorentz force*.

Because of the property of the vector product, Eq. (4.1) gives a force perpendicular to the velocity $\boldsymbol{v}$ as indicated before, but also perpendicular to the magnetic field $\mathscr{B}$. Equation (4.1) implies also that when $\boldsymbol{v}$ is parallel to $\mathscr{B}$, the force $\boldsymbol{F}$ is zero. In fact it is observed that at each point in a magnetic field there is a certain direction of motion such that the charge experiences no force. This direction is defined as the direction of the magnetic field at the point. Figure 4-2 illustrates the relation between the three vectors $\boldsymbol{v}$, $\mathscr{B}$, and $\boldsymbol{F}$ for both a positive and a negative charge. The figure shows the method, called the *right hand rule*, for determining the direction of the force.

If α is the angle between $\boldsymbol{v}$ and $\mathscr{B}$, the magnitude of F is

$$F=qv\mathscr{B}\sin\alpha. \tag{4.3}$$

The force is maximum when $\alpha=\pi/2$ or $\boldsymbol{v}$ is perpendicular to $\mathscr{B}$, resulting in

$$F_{\max}=qv\mathscr{B}. \tag{4.4}$$

The force is zero when $\alpha=0$ or when $\boldsymbol{v}$ is parallel to $\mathscr{B}$ as indicated previously.

From Eq. (4.1), the unit of magnetic field may be defined as N/(C m s^{-1}) or kg s^{-1} C^{-1}. This unit is called a *tesla* (T) in honor of the Yugoslavian-born American engineer Nicholas Tesla (1856–1943). That is, T = kg s^{-1} C^{-1}. One tesla corresponds to the magnetic field that produces a force of one newton on a charge of one coulomb moving perpendicular to the field at one meter per second.

Because the magnetic force $F = qv \times \mathscr{B}$ is perpendicular to the velocity, the work associated with the magnetic force is zero; therefore the magnetic force does not produce any change in the kinetic energy of the particle. Although the magnetic force is not conservative in that it is not related to a magnetic potential energy, when a particle moves in combined electric and magnetic fields, its total energy remains constant. (Total energy means its kinetic energy plus the potential energy from its different interactions.)

Example 4.1. A cosmic-ray proton with a velocity equal to 10^7 m s^{-1} enters the magnetic field of the earth in a direction perpendicular to it. Estimate the magnetic force exerted on the proton.

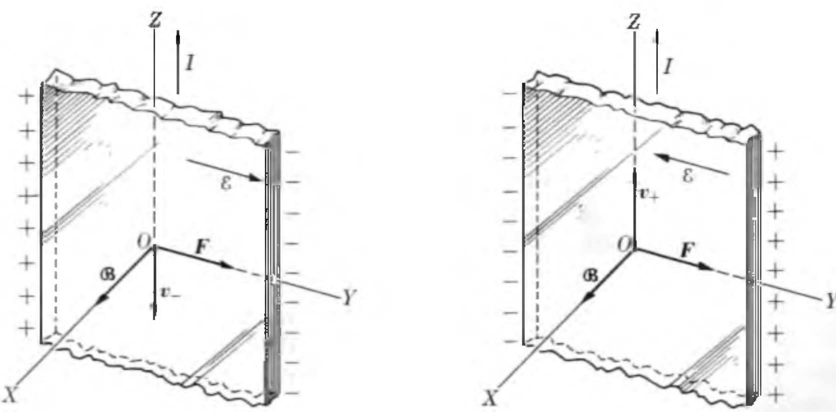
▼ The intensity of the magnetic field near the earth's surface at the equator is about $\mathscr{B} = 1.3 \times 10^{-7}$ T. The electric charge on the proton is $q = +e = 1.6 \times 10^{-19}$ C. Therefore from Eq. (4.4) the force on the proton is

$$F = qv\mathscr{B} = 2.08 \times 10^{-19} \text{ N}.$$

Since $m = m_p = 1.67 \times 10^{-27}$ kg, the acceleration caused by this force is $a = F/m_p = 1.24 \times 10^8$ m s^{-2}. Thus the acceleration of the proton due to the magnetic field is about ten million times the acceleration of gravity. ▲

Example 4.2. The *Hall effect.* In 1879 the American physicist E. C. Hall (1855–1929) discovered that when a metal plate, along which a current I is passing, is placed in a magnetic field perpendicular to the plate, a potential difference appears between opposite points on the edges of the plate.

▼ The Hall effect is a typical application of Eq. (4.1). Suppose first that the carriers of the electric current in the metal plate are electrons, having a negative charge $q = -e$. Consider Fig. 4-3(a), in which the Z-axis has been drawn parallel to the current I: the actual motion of the electron is



(a) Negative carriers ($q = -e$) (b) Positive carriers ($q = +e$)

Fig. 4-3. The Hall effect.

in the $-Z$-direction with the velocity v_-. When the magnetic field $\mathcal{B}$ is applied perpendicular to the plate, along the X-axis, the electrons are subject to the force $F=(-e)v_- \times \mathcal{B}$. The vector product $v_- \times \mathcal{B}$ is along the $-Y$-axis; but when the product is multiplied by $-e$, the result is a vector F along the $+Y$-axis. Therefore the electrons drift to the right-hand side of the plate, which thus becomes negatively charged. The left-hand side, being deficient in the usual number of electrons, becomes positively charged. As a consequence an electric field $\mathcal{E}$ parallel to the $+Y$-axis is produced. When the force $(-e)\mathcal{E}$ on the electrons, which is produced by this electric field and directed to the left, balances the force to the right, which is produced by the magnetic field $\mathcal{B}$, equilibrium results and there is no further separation of charges. The transverse electric field leads to a transverse potential difference between opposite sides of the conductor, the left-hand side being at the higher potential; the value of the potential difference is proportional to the magnetic field. This outcome is the normal, or "negative," Hall effect, exhibited by most metals, such as gold, silver, platinum, copper, etc. However with some metals, such as cobalt, zinc, and iron, and with other materials, such as the semiconductors, an opposite, or "positive," Hall effect is produced as shown in Figure 4-3(b).

To explain the positive Hall effect, suppose that the carriers of the current are positively charged particles with $q=+e$. Then they must move in the same sense as the current so that their velocity v_+ is along the $+Z$-axis as in Fig. 4-3(b). The magnetic force on the moving charges is $F=(+e)v_+ \times \mathcal{B}$, and it is directed toward the $+Y$-axis. However since the charges are positive, the right-hand side of the plate becomes positively charged, the left-hand side becomes negatively charged, and a transverse electric field is produced in the $-Y$-direction. Therefore, the potential difference is the reverse of that in the case of negative carriers, and the result is a positive Hall effect.

When the two types of Hall effect were discovered, physicists were very puzzled because at that time the general belief was that the only carriers of electric current in a solid conductor were the negatively charged electrons. However, it has been found that in certain circumstances it is advantageous to say that the carriers of electric current in a solid are positively charged particles. In these materials there are places in which normally an electron should be present, but from some defect in the solid structure the electron is missing; in other words, there exists what may be described as an *electron hole*. When a nearby electron for some reason moves to fill an existing hole, the electron obviously produces a hole at its original position. Thus electron holes move in a direction exactly opposite to that in which the negatively charged electrons move under an applied electrical field. Electron holes behave entirely similarly to positive particles. Thus the Hall effect provides a very useful method of determining the sign of the charge carriers in a conductor. ▲

4.3 Motion of a Charge in a Magnetic Field

Consider the motion of a charged particle in a uniform magnetic field; i.e., a magnetic field having the same intensity and direction at all points in space. For simplicity the case of a particle moving in a direction perpendicular to the magnetic field will be studied (Fig. 4-4). Since the force is perpendicular to the velocity, the effect of the force is to change the direction of the velocity without changing its magnitude; the result is a uniform circular motion. The acceleration is then centripetal; and from

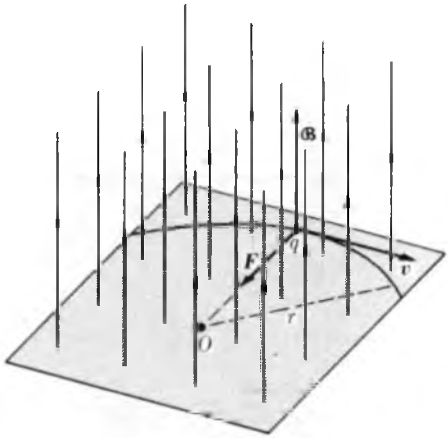


Fig. 4-4. A charge moving perpendicular to a uniform magnetic field follows a circular path.

the equation for circular motion. $F = mv^2/r$ with F given by Eq. (4.4).

$$\frac{mv^2}{r} = qv\mathscr{B}$$

or

$$r = \frac{mv}{q\mathscr{B}}, \tag{4.5}$$

which gives the radius of the circle described by the particle. For example from the data of Example 4.1, the proton would describe a circle of radius 8×10^5 m if the field were uniform. By writing $v = \omega r$ where ω is the angular velocity of the charged particle, Eq. (4.5) may be written as

$$\omega = \frac{q}{m}\,\mathscr{B}. \tag{4.6}$$

Therefore the angular velocity is independent of the velocity v and depends only on the ratio q/m and the field $\mathscr{B}$. The expression (4.6) gives the magnitude of ω but not its direction; however, the acceleration in a uniform circular motion may be written in vector form as $\boldsymbol{a} = \boldsymbol{\omega} \times \boldsymbol{v}$. Therefore the equation of motion $\boldsymbol{F} = m\boldsymbol{a}$ becomes

$$m\boldsymbol{\omega} \times \boldsymbol{v} = q\boldsymbol{v} \times \boldsymbol{\mathscr{B}}$$

or, reversing the vector product on the right-hand side and dividing by m, we get

$$\boldsymbol{\omega} \times \boldsymbol{v} = -(q/m)\,\boldsymbol{\mathscr{B}} \times \boldsymbol{v},$$

indicating that

$$\boldsymbol{\omega} = -(q/m)\,\boldsymbol{\mathscr{B}}, \tag{4.7}$$

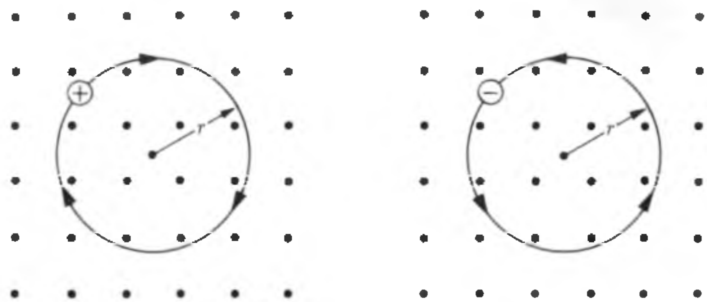


Fig. 4-5. Circular path of positive and negative charges in a uniform magnetic field.

which gives ω both in magnitude and direction.* The minus sign indicates that ω has the opposite direction to $\mathcal{B}$ for a positive charge and the same direction for a negative charge. The angular velocity ω of a charged particle in a uniform magnetic field is called the *cyclotron frequency* for reasons to be explained in Example 4.8 when the cyclotron is discussed.

It is customary to represent a field perpendicular to the paper by a dot ($\cdot$) if the field is directed toward the reader and by a cross ($\times$) if the field is directed into the page. Figure 4-5 represents the paths of a positive (a) and a negative (b) charge moving perpendicularly to a uniform magnetic field directed out of the page. In (a), ω is directed into the page and in (b) toward the reader.

The bending of the path of an ion in a magnetic field therefore provides a means for determining whether the ion's charge is positive or negative if the direction of its motion is known. Figure 4-6 shows the paths of several charged particles made visible in a *cloud chamber*† placed in a magnetic field. The applied magnetic field is many times stronger than the earth's magnetic field so that the radius of the path is of the order of the dimensions of the cloud chamber. Note that the paths are bent in either of two opposite senses, indicating that some particles are positive and others are negative. It may be observed that some of the particles describe a spiral of decreasing radius. This configuration indicates that the particle is being slowed down by collision with the gas molecules. According to Eq. (4.5), the decrease in the velocity of the particle results in a decrease in the radius of the orbit.

*Mathematically speaking, we should have written $\omega = -(q/m)\mathcal{B} + \lambda v$ where λ is an arbitrary constant; but Eq. (4.6) indicates that we must make $\lambda = 0$.

†A cloud chamber is a device containing a gas-and-vapor mixture in which the path of a charged particle is made visible by condensing the vapor on ions of the gas. The ions are produced by the interaction of the charged particle and the gas molecules. The condition for condensation is obtained by cooling the mixture by a rapid (adiabatic) expansion. The mixture may be air and water vapor.

Fig. 4-6. Cloud-chamber photograph of paths of charged particles in a uniform magnetic field. The magnetic field is directed into the page. The particles at the top are positively charged.

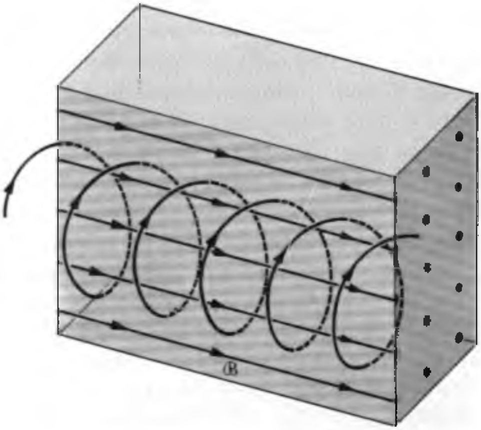


Fig. 4-7. Helical path of a positive ion moving obliquely to a uniform magnetic field.

When a charged particle moves initially in a direction that is not perpendicular to the magnetic field, the velocity may be separated into its parallel and perpendicular components relative to the magnetic field. The parallel component remains unaffected, and the perpendicular component changes continuously in direction but not in magnitude. The motion is then the resultant of a uniform motion parallel to the field and a circular motion around the field, with angular velocity still given by Eq. (4.6). The path is a helix as shown in Fig. 4-7 for a positive ion.

Another fact that results from Eq. (4.5) is that the larger the magnetic field, the smaller the radius of the path of the charged particle. Therefore if the magnetic field is not uniform, the path is not circular. Figure 4-8 shows a magnetic field directed from left to right with its strength increasing in that direction. Thus a charged particle injected at the left-hand side of the field describes a helix whose radius decreases continuously. A more detailed analysis, omitted here, would show that, to conserve energy, the component of the velocity parallel to the field does not remain constant but decreases (and therefore the pitch of the helix also decreases) as the particle moves in the direction of increasing field strength. Eventually the parallel velocity reduces to zero if the path is long enough in the magnetic field, and the particle is forced to move back or antiparallel to the magnetic field. Thus as a magnetic field increases in strength, the field begins to act as a reflector of charged particles, or, as it is popularly called, a *magnetic mirror*. This effect is used for containing ionized gases or plasmas.

Another situation is depicted in Fig. 4-9, in which a magnetic field perpendicular to the page increases in intensity from right to left. The path of a positive ion injected perpendicular to the magnetic field has also been indicated: that path is more curved at the left, where the field is stronger, than at the right, where the field is weaker. The path is not closed, and the particle drifts across the field perpendicular to the direction in which the magnetic field increases.

Example 4.3. Discovery of the positron.

▼ The relation between the velocity of a charged particle and the radius of its orbit in a magnetic field led to the 1932 discovery of the *positron* in cosmic rays. The positron is a fundamental particle having the same mass m_e as the electron but a *positive* charge $+e$. The discovery of the positron was the work of the American physicist Carl D. Anderson (1905–).* Anderson obtained the cloud-chamber photograph in Fig. 4-10. The horizontal band seen in the figure is a lead slab 0.6 cm thick that had been inserted inside the cloud chamber, and through which the particle has passed. That the lower part of the path of the particle is less curved than the upper part indicates that the particle had less velocity (and energy) above the slab than below it. Therefore the particle is moving upward since it must lose energy and slow down in passing through the slab. The curvature of the track of the particle and the sense of the motion relative to the magnetic field indicate that the particle is a positive one. The path looks very much like that of an electron—but a positive electron. Equation (4.2) may be rewritten as $p = mv = q\mathscr{B}r$. Therefore, if r is measured from the photograph and it is assumed that $q = e$, the momentum may be calculated. The order of magnitude of p corresponds to a particle with approximately the same mass as an electron. A more detailed analysis enables the particle's velocity to be evaluated and its mass to be computed; its mass is in fact the electron mass. ▲

Example 4.4. The motion of ions in a magnetic field for the case of charged particles falling on the earth from outer space and constituting part of what are called *cosmic rays*.

▼ Figure 4-11 shows the magnetic field of the earth.† Particles falling along the magnetic axis of the earth do not suffer any deviation and reach the earth even if they have very small energy.

---

*The existence of this particle, however, had been predicted by the British physicist Paul A. M. Dirac (1902–) a few years before its discovery.

†Actually the magnetic field around the earth shows several local anomalies and an overall distortion in the direction away from the sun. The schematic representation of Fig. 4-11 ignores these variations.

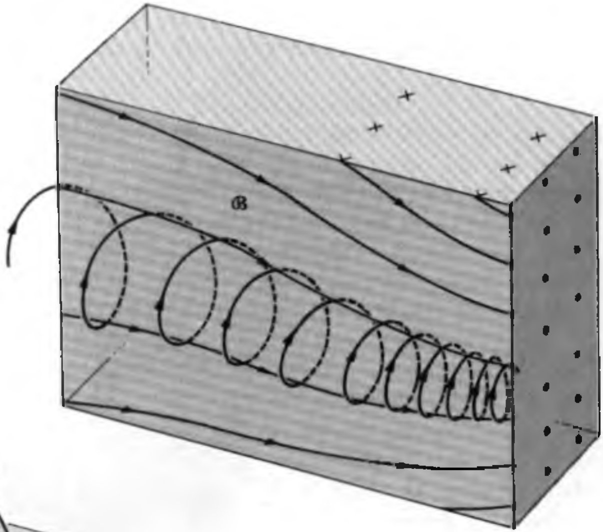


Fig. 4-8. Path of a positive ion in a nonuniform magnetic field.

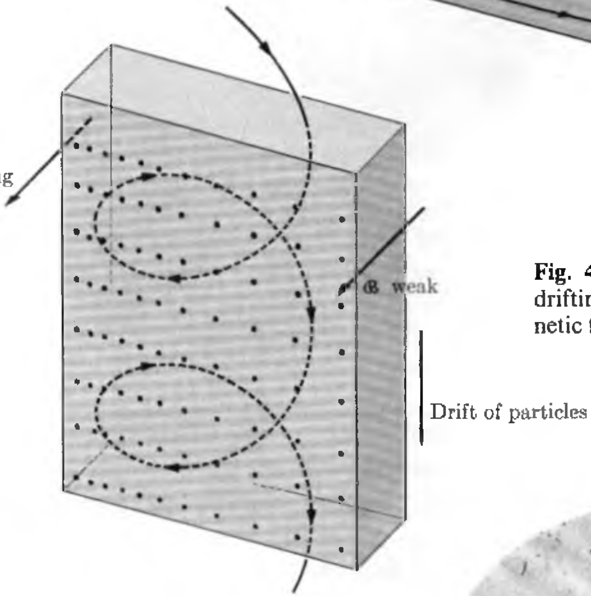


Fig. 4-9. Plane motion of an ion drifting across a nonuniform magnetic field.

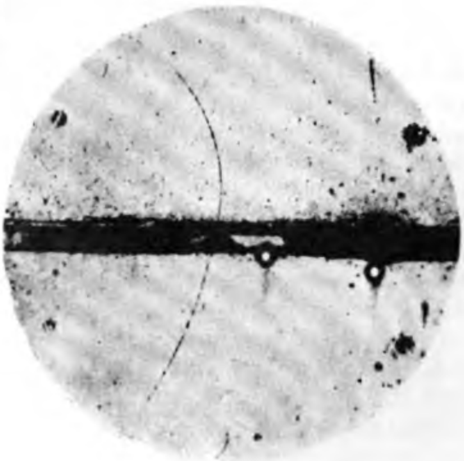


Fig. 4-10. Anderson's cloud-chamber photograph of the path of a positron (positive electron) in a magnetic field directed into the page. This photograph presented the first (1932) experimental evidence of the existence of positrons, previously predicted by Dirac.

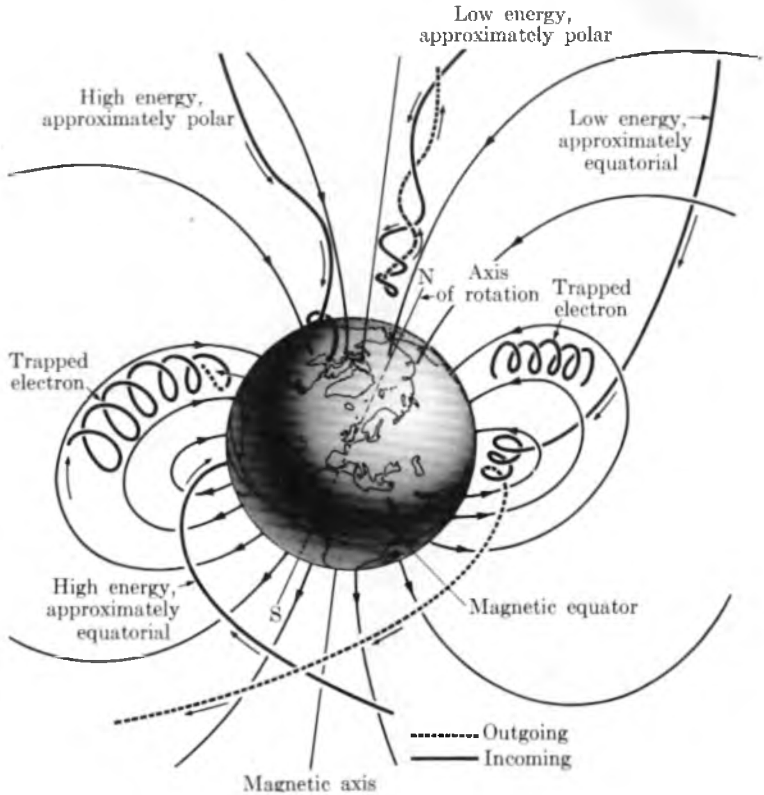


Fig. 4-11. Motion of charged cosmic-ray particles in the earth's magnetic field.

Particles falling at an angle with the magnetic field of the earth describe a helical path, and those moving very slowly may be bent so much that they do not reach the earth's surface. Those arriving on the magnetic equator suffer the largest deflection because they are moving in a plane perpendicular to the magnetic field. Therefore only the most energetic particles at the magnetic equator can reach the earth's surface. In other words the minimum energy that a charged cosmic particle must have to reach the earth's surface increases from the earth's magnetic axis to the earth's magnetic equator.

Another effect caused by the earth's magnetic field is the *east-west asymmetry* of cosmic radiation. Whether the charged cosmic particles are preponderantly positive or negative is not definitely known. However, particles of opposite signs are bent in opposite directions by the earth's magnetic field. If the number of positive particles in the cosmic rays reaching the earth is different from the number of negative particles, the cosmic rays arriving at a given place on the earth's surface in a direction east from the zenith should have a different intensity from those arriving in a direction west from the zenith. The experimental results are highly in favor of a majority of positively charged particles. ▲

Example 4.5. The *Van Allen radiation belts* formed from cosmic charged particles interacting with the earth's magnetic field.

▼ Above the earth's surface the regions commonly called the Van Allen belts [after the American physicist James Van Allen (1914–)] are composed of rapidly moving charged particles, mainly electrons and protons, trapped in the earth's magnetic field. The inner belt extends from about 800 km to about 4000 km above the earth's surface; the outer belt extends up to about 60,000 km from the earth.* They were discovered in 1968 by apparatus carried in an American Explorer satellite and investigated by the lunar probe Pioneer III. To understand better the trapping of charged particles in the Van Allen belts, consider, for example, a free electron produced by a collision between an atom and a cosmic ray many kilometers above the earth's surface. The velocity component perpendicular to the earth's magnetic field causes the electron to travel in a curved path. However, the strength of the field is greater nearer the surface of the earth. The result is a motion similar to that shown in Fig. 4-9, with the electron drifting eastward because of its negative charge (for positive charges the drift is westward). A further effect arises from the component of the electron's velocity parallel to the earth's magnetic field: this component produces a spiraling (like that shown in Fig. 4-7) toward one of the poles along the magnetic lines. Because of the increase in the magnetic field strength toward the north or south, the gyration becomes tighter and tighter while at the same time the parallel component of the velocity decreases as explained in connection with the magnetic mirror effect of Fig. 4-8. Each electron reaches a specific north or south latitude at which the parallel velocity becomes zero; which latitude it is depends on the initial velocity of injection. The electron then retreats toward the opposite pole. The resultant motion is thus an eastward change in longitude and a north-south oscillation in latitude. The motion is repeated continuously, perhaps for several weeks, until the electron is ejected from the Van Allen belt by a collision that ends its trapped condition. A similar situation occurs with the trapped protons. ▲

4.4 Examples of Motion of Charged Particles in a Magnetic Field

In this section several concrete situations in which an ion moves in a magnetic field will be illustrated.

Example 4.6. The mass spectrometer.

▼ Consider the arrangement illustrated in Fig. 4-12. Here I is an ion source (for electrons it may be just a heated filament); and S_1 and S_2 are two narrow slits through which the ions pass, being

*There is good evidence to show that the inner belt is composed of protons and electrons arising from the decay of neutrons that have been produced in the earth's atmosphere by cosmic-ray interactions. The outer belt consists primarily of charged particles that have been ejected by the sun. An increase in the number of these particles is associated with solar activity, and their removal from the radiation belt is the cause for auroral activity and radio-transmission blackouts.

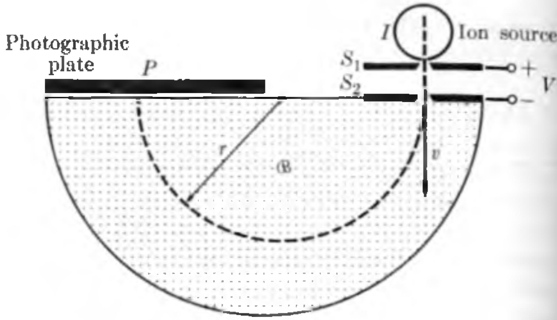


Fig. 4-12. Dempster's mass spectrometer. I is an ion source. Slits S_1 and S_2 serve as collimators of the ion beam. V is the accelerating potential difference applied between S_1 and S_2. P is a photographic plate that registers the arrival of the ions.

accelerated by the potential difference V applied between them. The exit velocity of the ions is calculated from Eq. (1.25), which gives

$$v^2=2\left(\frac{q}{m}\right)V. \tag{4.8}$$

In the region beyond the slits is a uniform magnetic field directed normal to the velocity of the ion (upward from the page in the figure). An ion will then describe a circular orbit, bent in one direction or the other depending on the sign of its charge q. After describing a semicircle the ions fall on a photographic plate P and leave a mark. The radius r of the orbit is given by Eq. (4.5), from which the velocity v is

$$v=\frac{q}{m}\mathscr{B}r. \tag{4.9}$$

Combining Eqs. (4.8) and (4.9) to eliminate v yields

$$\frac{q}{m}=\frac{2V}{\mathscr{B}^2r^2}, \tag{4.10}$$

which gives the ratio q/m in terms of three quantities (V, $\mathscr{B}$, and r), each of which can easily be measured. This technique may be applied to electrons, protons, and any other charged particle, atom or molecule. If the charge q is measured independently, the mass of the particle may be obtained. These are the methods that were referred to previously in Section 1.5.

The arrangement of Fig. 4-12 constitutes a *mass spectrometer* because it separates ions having the same charge q but different mass m since the radius of the path of each ion depends on the ion's q/m value. The particular spectrometer shown in Fig. 4-12 is called *Dempster's* mass spectrometer. Several other types of mass spectrometers, all based on the same principle, have been developed. Scientists using this technique discovered in the 1920's that atoms of the same chemical element do not necessarily have the same mass. The different varieties of atoms of one chemical element, which differ in mass, are called *isotopes*.

The experimental arrangement of Fig. 4-12 may be used also to obtain the ratio q/m for a particle moving with different velocities. It has been found that q/m depends on v as if q remains constant, and m varies with the velocity by the relation $m=m_0/\sqrt{1-v^2/c^2}$. Therefore

the electric charge is an invariant, being the same for all observers in uniform relative motion. ▲

Example 4.7. Discovery of the *electron.*

▼ During the latter part of the nineteenth century there was a great amount of experimental work on electrical discharges. These experiments consisted of producing an electrical discharge through a gas at low pressure by placing two electrodes within the gas and applying a large potential difference to them. Depending on the pressure of the gas in the tube, several luminous effects were observed. When the gas in the tube was kept at a pressure less than 100 Pa, no more visible effects were observed within the tube, but a luminous spot was observed on the tube wall at O directly opposite the cathode C (Fig. 4-13). Therefore the hypothesis was made that the cathode emitted some radiation that moved in a straight line toward O. Accordingly this radiation was called *cathode rays.*

When two parallel plates P and P' were added inside the tube and a potential difference was applied, an electric field $\mathscr{E}$ directed from P to P' was produced. The result of applying this electric field was that the luminous spot moved from O to O'; that is, in the direction corresponding to a negative electric charge. This movement suggested that cathode rays were simply a stream of negatively charged particles. If q is the charge of each particle and v its velocity, the deviation $d = OO'$ can be computed by applying Eq. (1.9), $q\mathscr{E}a/mv^2 = d/L$. The electric force on the particle is $q\mathscr{E}$ and is directed upward. If a magnetic field perpendicular to the electric field and directed into the paper is applied in the same region occupied by the plates, particles accelerating through the field will experience a magnetic force. According to Eq. (4.4), the magnetic force is $qv\mathscr{B}$ and is directed downward because q is a negative charge. By properly adjusting $\mathscr{B}$, we can make the magnetic force equal to the electric force. This adjustment results in a zero net force, and the luminous spot returns from O' to O; that is, there is no deflection of the cathode rays. Then $q\mathscr{E} = qv\mathscr{B}$ or $v = \mathscr{E}/\mathscr{B}$. This equation provides a measurement of the velocity of the charged particle. Substituting this value of v in Eq. (1.19) gives the ratio q/m of the particle constituting the cathode rays:

$$\frac{q}{m} = \frac{\mathscr{E}d}{\mathscr{B}^2 La}.$$

Experiments of this type provided one of the first reliable experiments for measuring q/m, and indirectly gave a proof that cathode rays consist of negatively charged particles, since then called *electrons.*

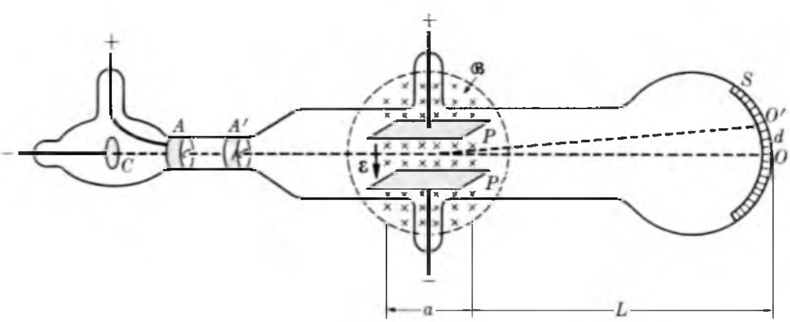


Fig. 4-13. Thomson's experiment for measuring q/m. Cathode rays (electrons) emitted by C and collimated by A and A' arrive at the screen S after passing through a region in which electric and magnetic fields are applied.

These experiments were published in 1897 by the British physicist Sir J. J. Thomson (1856–1940), who expended great effort and time trying to discover the nature of cathode rays. Today it is understood that free electrons present in the metal cathode C are pulled out or evaporated as a result of the strong electric field applied between C and A, and are accelerated along the tube by the same field. ▲

Example 4.8. The cyclotron.

▼ That the path of a charged particle in a magnetic field is circular has permitted the design of particle accelerators that operate cyclically. One difficulty with electrostatic accelerators (described in Section 1.7) is that the acceleration depends on the total potential difference V. Since the electric field within the accelerator is $\mathscr{E}=V/d$, if V is very large, the length d of the accelerator tube must also be very great to prevent the development of strong electric fields that would produce electric breakdown in the materials of the accelerating tube. Also, a very long accelerating tube poses several engineering difficulties. However in a cyclic accelerator an electric charge may receive a series of accelerations by passing many times through a relatively small potential difference. The first instrument working on this principle was the *cyclotron*, designed by the American physicist E. O. Lawrence (1901–1958). The first practical cyclotron started operating in 1932.

Essentially, a cyclotron (Fig. 4-14) consists of a cylindrical cavity that is divided into two halves D_1 and D_2 (each called a "dee" because of its shape), and that is placed in a uniform magnetic field parallel to its axis. The two cavities are electrically insulated from each other. An ion source S is placed in the center of the space between the dees. An alternating potential difference of the order of 10^4 V is applied between the dees. When the ions are positive, they will be accelerated toward the negative dee. Once inside a dee, the ion experiences no electrical force since the electric field is zero in the interior of a conductor. However because of the magnetic field, ions describe a circular orbit with a radius given by Eq. (4.5), $r=mv/q\mathscr{B}$, and an angular velocity, equal to the cyclotron frequency of the particles and given by Eq. (4.6), $\omega=q\mathscr{B}/m$. The potential difference between the dees is made to oscillate with a frequency equal to ω. In this way the potential difference between the dees is in resonance with the circular motion of the ions.

After the particle has described half a revolution, the polarity of the dees is reversed; and when the ion crosses the gap between them, the ion receives another small acceleration. The next half-

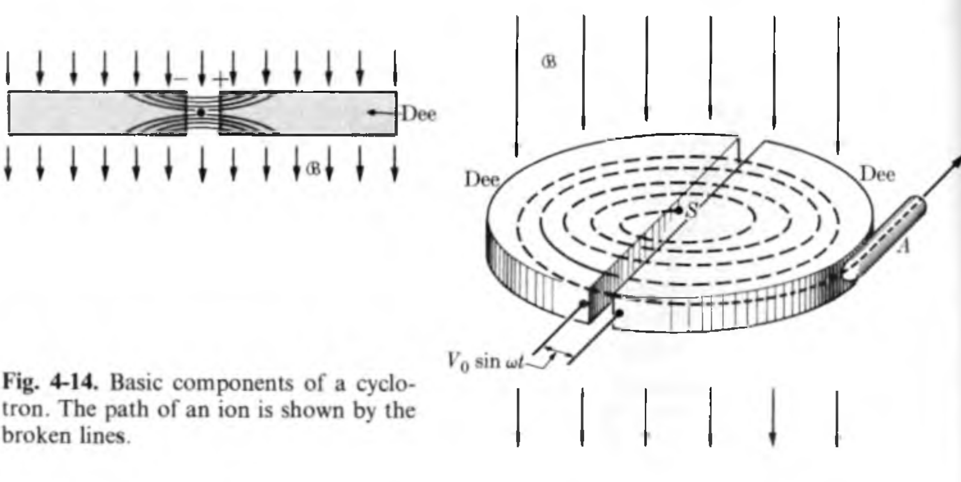


Fig. 4-14. Basic components of a cyclotron. The path of an ion is shown by the broken lines.

circle described then has a larger radius but the same angular velocity. The process repeats itself several times until the radius attains a maximum value R, which is practically equal to the radius of the dees. The magnetic field at the edge of the dees is decreased sharply; and the particle moves tangentially, escaping through a convenient opening. The maximum velocity v_{max} is related to the radius R by Eq. (4.5), namely,

$$R=\frac{mv_{max}}{q\mathscr{B}}$$

or

$$v_{max}=\left(\frac{q}{m}\right)\mathscr{B}R.$$

The kinetic energy of the particles emerging from A is then

$$E_k=\tfrac{1}{2}mv_{max}^2=\tfrac{1}{2}q\left(\frac{q}{m}\right)\mathscr{B}^2R^2, \qquad (4.11)$$

and is determined by the characteristics of the particle, the strength of the magnetic field, and the radius of the cyclotron, but is independent of the accelerating potential. When the potential difference is small, the particle has to make many turns before it picks up the final energy; but when the potential difference is large, only a few turns are required.

The strength of the magnetic field is limited by technological factors, such as availability of materials with the required properties; but, with magnets of a sufficiently large radius, the particle can in principle be accelerated to any desired energy. However, the larger the magnet, the greater the weight and cost. There is also a physical limiting factor to the energy in a cyclotron. As the energy increases, the velocity of the ion also increases and results in a change of mass according to the relativistic relation $m=m_0/\sqrt{1-v^2/c^2}$.* When the energy is very large, the change in mass is appreciable enough to make the cyclotron frequency of the ion change noticeably. Therefore unless the frequency of the electric potential is changed, the orbit of the particle will no longer be in phase with the oscillating potential, and no further acceleration is produced. Thus in a cyclotron the energy is limited by the relativistic mass effect. ▲

Example 4.9. The University of Michigan has a cyclotron that has pole faces with a diameter of 2.1 m and an extraction radius of 36 in. or 0.92 m. The maximum magnetic field is $\mathscr{B}=1.50$ T and the maximum attainable oscillating frequency of the accelerating field is 15×10^6 Hz. Calculate (a) the energy of the protons and alpha particles (doubly charged helium nuclei) when accelerated by this device, and (b) their cyclotron frequency. (c) What is the percentage difference between the cyclotron frequency at the center and that at the edge if the relativistic mass variation* is taken into account?

▼ (a) From Eq. (4.11) with the corresponding values for the charge and mass of the protons and alpha particles, the kinetic energies of both may be expressed as

$$E_k=1.46\times10^{-11}\ \text{J}=91\ \text{MeV}.$$

(b) The cyclotron frequency for the alpha particle is $\omega_\alpha=7.2\times10^7\ \text{s}^{-1}$, or a frequency $\nu_\alpha=\omega_\alpha/2\pi=11.5\times10^6$ Hz, which is within the range of the maximum design frequency. For the protons the

*See the appendix for a discussion of relativistic mechanics.

frequency is twice that, or 23×10^6 Hz. This is the frequency with which the potential applied to the dees must change. However, the maximum design frequency of the cyclotron is 15×10^6 Hz, and therefore this machine cannot accelerate protons to the theoretical value of 91 MeV. If the maximum oscillatory frequency is the maximum cyclotron frequency for protons ($\omega_p = 9.42 \times 10^7$ s^{-1}), the corresponding magnetic field for cyclotron resonance is 0.984 T. Thus the frequency-limited kinetic energy of protons is

$$E_k = \tfrac{1}{2}mv^2 = \tfrac{1}{2}m\omega^2 R^2 = 0.63 \times 10^{-11}\ \text{J} = 39\ \text{MeV}.$$

(c) At an energy $E = m_0c^2 + E_k$, the mass of the particle is

$$m = E/c^2 = m_0 + E_k/c^2;$$

and so E_k/c^2 gives the change in mass. From Eq. (4.6) we see that the cyclotron frequency is inversely proportional to the mass. Therefore if ω and ω_0 are the frequencies corresponding to the masses m and m_0 of the same particle, it follows that $\omega/\omega_0 = m_0/m$ or

$$\frac{\omega - \omega_0}{\omega_0} = -\frac{m - m_0}{m} = -\frac{E_k/c^2}{m_0 + E_k/c^2} = -\frac{E_k}{m_0c^2 + E_k}.$$

The left-hand side gives the percentage change in the cyclotron frequency; and the right-hand side, the percentage change in mass. For relatively low energies the kinetic energy term E_k in the denominator may be neglected in comparison with m_0c^2, so that with $\Delta\omega = \omega - \omega_0$, the above relation becomes

$$\frac{\Delta\omega}{\omega} = -\frac{E_k}{m_0c^2}.$$

Thus so long as the kinetic energy is small compared with the rest energy of the particles, the change in frequency is very small. In our case for alpha particles, $\Delta\omega/\omega = -0.024 = 2.4\%$; and for protons, $\Delta\omega/\omega = -0.042 = 4.2\%$.

The results obtained in this example indicate also that since electrons have a rest mass about 1/1840 that of the proton (Section 1.5), the kinetic energy to which electrons can be accelerated (without appreciably deviating from their cyclotron frequency) is only about 1/1840 that for protons. For this reason, cyclotrons are not used for accelerating electrons.

Experimentally, the relativistic mass effect can be compensated for, either by shaping the magnetic field so that at each radius the value of ω remains constant in spite of the change in mass, or by changing the frequency applied to the dees and keeping the magnetic field constant while the particle is spiraling so that at each instant there is resonance between the particle motion and the applied potential. The first design is called a *synchrotron* and the second is called a *synchro-cyclotron*. A synchrotron may operate continuously, but a synchrocyclotron operates in bursts because of the need for adjusting the frequency. ▲

Example 4.10. Motion of a charged particle in crossed electric and magnetic fields.

▼ Consider the case in which a uniform electric field is parallel to the Y-axis of a coordinate system and a uniform magnetic field is parallel to the Z-axis as shown in Fig. 4-15. The equation of motion of a charged particle produced by these fields is that given by the Lorentz force; that is,

$$\boldsymbol{F} = m\frac{d\boldsymbol{v}}{dt} = q(\boldsymbol{\mathcal{E}} + \boldsymbol{v} \times \boldsymbol{\mathcal{B}}).$$

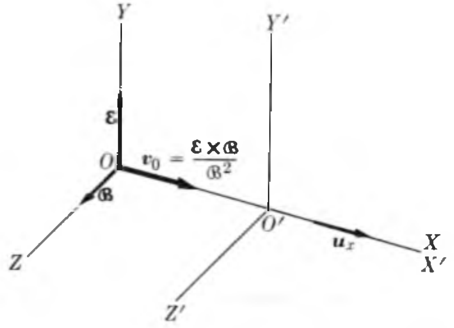


Figure 4-15

By a Galilean transformation from frame XYZ to another frame $X'Y'Z'$, moving relative to the XYZ-axes with the relative velocity

$$\boldsymbol{v}_0 = \frac{\mathcal{E}\times\mathcal{B}}{\mathcal{B}^2} = \boldsymbol{u}_x\frac{\mathcal{E}}{\mathcal{B}},$$

the velocity $\boldsymbol{v}'$ of the particle relative to the $X'Y'Z'$-axes may be written as $\boldsymbol{v}' = \boldsymbol{v} - \boldsymbol{v}_0$ or $\boldsymbol{v} = \boldsymbol{v}' + \boldsymbol{v}_0$ so that $d\boldsymbol{v}/dt = d\boldsymbol{v}'/dt$. Thus the equation of motion may be rewritten as

$$m\frac{d\boldsymbol{v}'}{dt} = q(\mathcal{E} + \boldsymbol{v}'\times\mathcal{B} + \boldsymbol{v}_0\times\mathcal{B}).$$

Because $\boldsymbol{v}_0\times\mathcal{B} = (\boldsymbol{u}_x\mathcal{E}/\mathcal{B})\times\boldsymbol{u}_z\mathcal{B} = -\boldsymbol{u}_y\mathcal{E} = -\mathcal{E}$, the first and last terms in the preceeding equation cancel. The reason for making the Galilean transformation is now clear: relative to the $X'Y'Z'$-axes, the equation of motion is

$$m\frac{d\boldsymbol{v}'}{dt} = q\boldsymbol{v}'\times\mathcal{B}.$$

Note then that the motion relative to $X'Y'Z'$ is as it would be if *no* electric field were present. If the particle moves initially in the XY-plane (that is, if $\boldsymbol{v}$ has no Z-component), the motion of the particle in the $X'Y'Z'$-frame will be a circle of radius $r = mv'/q\mathcal{B}$, described with angular velocity $\omega = -q\mathcal{B}/m$. Relative to XYZ, this circle advances along the X-axis with the velocity v_0; and one of the paths shown in Fig. 4-16 results. The pattern repeats itself in a distance $v_0P = 2\pi v_0/\omega$. If $2\pi v_0/\omega = 2\pi r$ or if $r = v_0/\omega$, the path is the normal cycloid, marked (1); but if $2\pi v_0/\omega \lesseqgtr 2\pi r$ or if $r \lesseqgtr v_0/\omega$, the paths (2) and (3) result, corresponding to curtate and prolate cycloids. If the charged

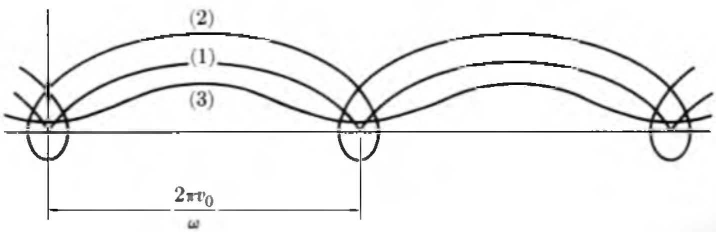


Fig. 4-16. Cycloidal paths of a particle relative to observer O. (1) $r = v_0/\omega$, (2) $r > v_0/\omega$, (3) $r < v_0/\omega$.

particle has an initial velocity component parallel to the Z-axis, the paths illustrated in Fig. 4-16 will move away from the XY-plane at a constant rate.

An interesting aspect revealed by this example is that while the observer who uses frame XYZ observes both an electric and a magnetic field, the observer who uses frame $X'Y'Z'$, in motion relative to XYZ, observes a motion of the charged particle corresponding only to a magnetic field. This point suggests that the value of electric and magnetic fields measured by two observers depends on their relative motion. This is a very important matter, which will be considered in greater detail in Section 4.6. ▲

4.5 Magnetic Field of a Moving Charge (Nonrelativistic)

Up to this point the magnetic field has been discussed without any reference to how magnetic fields are produced; there have been allusions to certain substances, called magnets, that produce magnetic fields in their natural state. A major breakthrough in understanding the origin of magnetism occurred in 1819 when the Danish physicist Hans Christian Oersted (1777–1851) accidently discovered that when placed beneath a long, current-carrying wire, a magnetic compass lined up perpendicular to the direction of the wire. Since an electric current consists of a stream of electric charges in motion, it seems reasonable to assume that *magnetic fields not only are felt by charges in motion but also are produced by charges in motion.* This assumption has been amply verified by analyzing the motion of charged particles.

Accordingly while an electric charge at rest relative to an observer produces only an electric field, a charged particle in motion relative to an observer produces both an electric field and a magnetic field. Thus electric and magnetic fields are simply two aspects of one fundamental property of matter, and the term *electromagnetic field* more appropriately describes the physical situation involving moving charges.

Consider a charge q (Fig. 4-17) moving with a velocity $\boldsymbol{v}$ relative to an observer. (Assume that the velocity is small compared with the velocity of light so that relativistic effects do not have to be taken into account in the calculations.) The electric field of the charge is radial and is given by the same expression found in Chapter 1 when the particle is at rest; that is,

$$\mathscr{E} = \frac{q}{4\pi\epsilon_0 r^2}\,\boldsymbol{u}_r \tag{4.12}$$

The electric lines of force are radial and therefore straight lines passing through the charge. Experiment shows that the magnetic field of the moving charge can be represented by magnetic lines of force that are circles with their centers on the line of motion of the charge. When the charge is positive, the magnetic lines of force are oriented in the direction of the fingers of the right hand when the thumb points in the direction of motion of the charge (see Fig. 4-17). Note in particular that $\mathscr{B}$ has no component in the direction of motion of the charge.

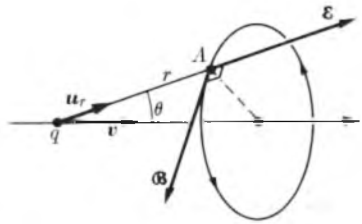


Fig. 4-17. Electric and magnetic fields produced by a moving positive charge.

Therefore at a given point in space, such as A, there is an electric field $\mathscr{E}$ in the radial direction and a magnetic field $\mathscr{B}$ in a direction perpendicular to both $\boldsymbol{r}$ and $\boldsymbol{v}$. Measurement of the magnetic field at several points shows that the magnetic field is given by the expression

$$\mathscr{B}=K_m\frac{q(\boldsymbol{v}\times\boldsymbol{u}_r)}{r^2}. \tag{4.13}$$

The constant K_m in SI units is equal to 10^{-7} m kg C^{-2} as will be explained in Section 5.7. It is customary to write Eq. (4.13) in the form

$$\mathscr{B}=\frac{\mu_0}{4\pi}\frac{q(\boldsymbol{v}\times\boldsymbol{u}_r)}{r^2} \tag{4.14}$$

where

$$\mu_0=4\pi K_m=1.3566\times10^{-6}\text{ m kg C}^{-2} \tag{4.15}$$

is a new constant called the *magnetic permeability of vacuum*. The magnitude of the magnetic field is

$$\mathscr{B}=\frac{\mu_0}{4\pi}\frac{qv\sin\theta}{r^2} \tag{4.16}$$

where θ is the angle between $\boldsymbol{r}$ and $\boldsymbol{v}$. Note that the magnitude of the magnetic field is zero along the line of motion of the charge and has maximum value in the plane that is perpendicular to the line of motion and passes through the charge.

Comparing Eqs. (4.12) and (4.16), note that the electric and magnetic fields of a moving charge at a given point in space are related in the form

$$\mathscr{B}=\mu_0\epsilon_0(\boldsymbol{v}\times\mathscr{E})=\frac{1}{c^2}(\boldsymbol{v}\times\mathscr{E}). \tag{4.17}$$

In this expression the relation

$$c=\frac{1}{\sqrt{\epsilon_0\mu_0}} \tag{4.18}$$

has been used; this constant, as will be shown later, is the velocity of light (or of any electromagnetic signal) in vacuum. After substitution of values for μ_0 and ϵ_0, c is 2.9979×10^8 m s^{-1}. (Usually c is taken as 3.0×10^8 m s^{-1} for most computations.)

In Example 4.12 it will be shown that when the velocity of the charge is comparable to that of light, Eqs. (4.13) and (4.16) for the electric and magnetic fields must be modified. However, the relation (4.17) remains valid at all velocities.

4.6 Electromagnetism and the Principle of Relativity

According to the principle of relativity (see Volume I, Chapter 11),* all laws of nature must be identical to all inertial observers. We must therefore proceed now to obtain the relation between the electric and magnetic fields as measured by two observers in uniform relative motion so that the principle of relativity remains valid.

Suppose that two observers O and O' (Fig. 4-18) are in uniform relative motion with velocity $\boldsymbol{v}$, and that there are two charges q and Q at rest relative to O'. The two charges are then in motion with velocity $\boldsymbol{v}$ relative to O. The values of the two charges are the same for both observers O and O' as previously stated in Example 4.6. For observer O' there is only an electric interaction between Q and q, and the force measured on q is $\boldsymbol{F}' = q\boldsymbol{\mathscr{E}}'$ where $\boldsymbol{\mathscr{E}}'$ is the electric field produced by Q at q as measured by O'.

On the other hand since O sees charge Q in motion, he observes that Q produces both an electric field $\boldsymbol{\mathscr{E}}$ and a magnetic field $\boldsymbol{\mathscr{B}}$; and since q is also observed to be in motion with the velocity $\boldsymbol{v}$, the force exerted by Q on q and measured by O is $\boldsymbol{F} = q(\boldsymbol{\mathscr{E}} + \boldsymbol{v} \times \boldsymbol{\mathscr{B}})$. With a common axis X and X' chosen parallel to the relative velocity of the observers, $\boldsymbol{v} = \boldsymbol{u}_x v$ and $\boldsymbol{v} \times \boldsymbol{\mathscr{B}} = -\boldsymbol{u}_y v \mathscr{B}_z + \boldsymbol{u}_z v \mathscr{B}_y$; and therefore the components of $\boldsymbol{F}$ relative to frame XYZ are

$$F_x = q\mathscr{E}_x, \qquad F_y = q(\mathscr{E}_y - v\mathscr{B}_z), \qquad F_z = q(\mathscr{E}_z + v\mathscr{B}_y). \tag{4.19}$$

The components of $\boldsymbol{F}'$ relative to frame $X'Y'Z'$ are

$$F'_x = q\mathscr{E}'_x, \qquad F'_y = q\mathscr{E}'_y, \qquad F'_z = q\mathscr{E}'_z. \tag{4.20}$$

Since q is at rest relative to O', the relations between the components of $\boldsymbol{F}$ and $\boldsymbol{F}'$ according to the Lorentz transformations of force (see the appendix) are

$$F'_x = F_x, \qquad F'_y = \frac{F_y}{\sqrt{1 - v^2/c^2}}, \qquad F'_z = \frac{F_z}{\sqrt{1 - v^2/c^2}}.$$

Substituting the values of the components as given by Eqs. (4.19) and (4.20), and canceling the common factor q yield

$$\mathscr{E}'_x = \mathscr{E}_x, \qquad \mathscr{E}'_y = \frac{\mathscr{E}_y - v\mathscr{B}_z}{\sqrt{1 - v^2/c^2}}, \qquad \mathscr{E}'_z = \frac{\mathscr{E}_z + v\mathscr{B}_y}{\sqrt{1 - v^2/c^2}}. \tag{4.21}$$

*See the appendix for a review of the special theory of relativity.

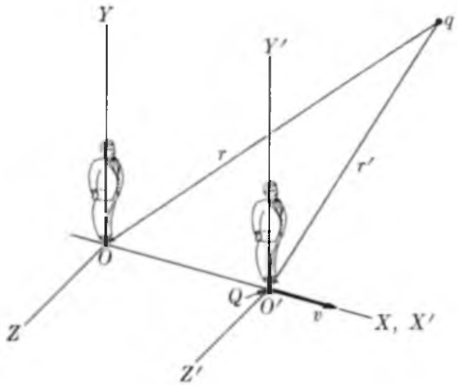


Fig. 4-18. Comparison of electromagnetic measurements by two observers in relative motion.

These expressions relate the electric field measured by observer O' to the electric and magnetic fields measured by observer O in accordance with the special theory of relativity. The inverse transformations of Eq. (4.21) are obtained by exchanging the fields and reversing the sign of v since frame XYZ moves with the velocity $-v$ relative to $X'Y'Z'$. Thus if observer O' measures an electric field $\mathscr{E}'$ and a magnetic field $\mathscr{B}'$, the electric field measured by O is given by

$$\mathscr{E}_x=\mathscr{E}'_x, \qquad \mathscr{E}_y=\frac{\mathscr{E}'_y+v\mathscr{B}'_z}{\sqrt{1-v^2/c^2}}, \qquad \mathscr{E}_z=\frac{\mathscr{E}'_z-v\mathscr{B}'_y}{\sqrt{1-v^2/c^2}}. \tag{4.22}$$

Suppose the charge Q, instead of being at rest in O', is moving relative to O' as well as to O: then O' notes a magnetic field $\mathscr{B}'$ in addition to the electric field $\mathscr{E}'$. A similar but more laborious calculation* then gives

$$\mathscr{B}'_x=\mathscr{B}_x, \qquad \mathscr{B}'_y=\frac{\mathscr{B}_y+v\mathscr{E}_z/c^2}{\sqrt{1-v^2/c^2}}, \qquad \mathscr{B}'_z=\frac{\mathscr{B}_z-v\mathscr{E}_y/c^2}{\sqrt{1-v^2/c^2}}. \tag{4.23}$$

Again as in Eq. (4.21), the inverse transformations of Eq. (4.23) are found by exchanging the fields and replacing v by $-v$, and result in

$$\mathscr{B}_x=\mathscr{B}'_x, \qquad \mathscr{B}_y=\frac{\mathscr{B}'_y-v\mathscr{E}'_z/c^2}{\sqrt{1-v^2/c^2}}, \qquad \mathscr{B}_z=\frac{\mathscr{B}'_z+v\mathscr{E}'_y/c^2}{\sqrt{1-v^2/c^2}}. \tag{4.24}$$

Equations (4.21) and (4.23), or their inverse Eqs. (4.22) and (4.24), constitute the Lorentz transformation for the electromagnetic field. These equations prove once more that the electric and magnetic fields are not separate entities but form a single physical entity called the *electromagnetic field*. The separation of an electromagnetic field into its electric and magnetic components is not an absolute procedure; rather the separation depends on the motion of the charges relative to the observer.

* For example if students wish to obtain the second and third equations in Eq. (4.23), we suggest that they use Eq. (4.21) to eliminate $\mathscr{E}'_y$ and $\mathscr{E}'_z$ from the inverse transformation of Eq. (4.22), and then solve for $\mathscr{B}'_y$ and $\mathscr{B}'_z$.

Example 4.11. Reconsider the situation discussed in Example 4.10: use the Lorentz transformation for the electromagnetic field to relate the fields measured by both observers.

▼ Recall that in Example 4.10 there was an electric field along the Y-axis and a magnetic field along the Z-axis. By a kinematical transformation to a set of axes $X'Y'Z'$ moving in the X-direction with velocity $v=\mathscr{E}/\mathscr{B}$, the motion was reduced to that of a particle under a magnetic field alone. Go one step further in this analysis, and see the implications of this example within the framework of the theory of relativity. In the frame XYZ, $\mathscr{E}_x=0$, $\mathscr{E}_y=\mathscr{E}$, and $\mathscr{E}_z=0$ for the electric field and $\mathscr{B}_x=\mathscr{B}_y=0$, $\mathscr{B}_z=\mathscr{B}$ for the magnetic field. Thus, from Eqs. (4.21) and (4.23) the fields observed in the $X'Y'Z'$ frame are

$$\mathscr{E}'_x=0, \qquad \mathscr{E}'_y=\frac{\mathscr{E}-v\mathscr{B}}{\sqrt{1-v^2/c^2}}, \qquad \mathscr{E}'_z=0,$$

$$\mathscr{B}'_x=0, \qquad \mathscr{B}'_y=0, \qquad \mathscr{B}'_z=\frac{\mathscr{B}-v\mathscr{E}/c^2}{\sqrt{1-v^2/c^2}}.$$

Setting $v=\mathscr{E}/\mathscr{B}$ yields $\mathscr{E}'_y=0$ and thus $\mathscr{E}'=0$; furthermore

$$\mathscr{B}'=\mathscr{B}'_z=\sqrt{1-v^2/c^2}\,\mathscr{B}.$$

Therefore the theory of relativity predicts that observer O' moving with velocity $v=\mathscr{E}/\mathscr{B}$ relative to O will measure no electric field and will measure a magnetic field smaller than the magnetic field measured by O, but in the same direction. ▲

4.7 The Electromagnetic Field of a Moving Charge (Relativistic)

In Chapter 1 a charge at rest was seen to have an electric field $\mathscr{E}=(q/4\pi\epsilon_0 r^2)\boldsymbol{u}_r$; and in Section 4.5 it was shown that a charge in motion produces in addition a magnetic field whose expression was given by $\mathscr{B}=(\mu_0/4\pi)qv\times\boldsymbol{u}_r/r^2$. However according to the preceding section, the fields $\mathscr{E}$ and $\mathscr{B}$ must be related by Eqs. (4.21) and (4.23). Therefore from the very beginning, a relativistic calculation must be used in order to obtain the correct expressions for $\mathscr{E}$ and $\mathscr{B}$ for a moving charge.

Consider a charge q at rest relative to the frame $X'Y'Z'$, which is moving, relative to XYZ, with velocity v parallel to the common X-axis. Observer O' measures no magnetic field but only an electric field as indicated before; therefore $\mathscr{B}'_x=\mathscr{B}'_y=\mathscr{B}'_z=0$, or $\mathscr{B}'=0$. Then the electric field transformations of Eq. (4.22) yield

$$\mathscr{E}_x=\mathscr{E}'_x, \qquad \mathscr{E}_y=\frac{\mathscr{E}'_y}{\sqrt{1-v^2/c^2}}, \qquad \mathscr{E}_z=\frac{\mathscr{E}'_z}{\sqrt{1-v^2/c^2}}. \tag{4.25}$$

Equations (4.25) indicate than when observer O, who sees the the charge moving, and O', who sees the charge at rest, compare their measurements of the electric field of the charge, the observers obtain the same field component parallel to the direction of motion, but O obtains a larger component perpendicular to the direction of motion. Similarly if Eqs. (4.25) are used to write the components of the electric field with respect to O, the magnetic field transformation of Eqs. (4.24) yield

$$\mathscr{B}_x=0, \qquad \mathscr{B}_y=-\frac{v\mathscr{E}_z}{c^2}, \qquad \mathscr{B}_z=\frac{v\mathscr{E}_y}{c^2}, \tag{4.26}$$

which are equivalent to $\mathscr{B}=\boldsymbol{v}\times\boldsymbol{\mathscr{E}}/c^2$. This is identical to Eq. (4.17), which as indicated before is the relation between the electric and magnetic fields of a charge moving with a constant velocity $\boldsymbol{v}$, a relation which is valid at all speeds.

In Fig. 4-19 the observations of O and O' are compared. If the charge is at O', observer O' measures an electric field at P' (in the $X'Y'$-plane) given by

$$\boldsymbol{\mathscr{E}}'=\frac{q}{4\pi\epsilon_0 r'^2}\boldsymbol{u}_{r'}=\frac{q}{4\pi\epsilon_0 r'^3}\boldsymbol{r}'.$$

Observer O sees the same point in the XY plane; but because of the Lorentz contraction, the X-coordinate of the point appears shortened by the factor $\sqrt{1-v^2/c^2}$ while the Y-coordinate remains the same. That is, $x=x'\sqrt{1-v^2/c^2}$; $y=y'$. Thus the angle θ that OP makes with OX is different from the angle θ' that $O'P'$ makes with $O'X'$ (Fig. 4-19). From Eqs. (4.25) the field $\boldsymbol{\mathscr{E}}$ that O measures at P has an x-component that is the same as that measured by O'; but the y-component appears larger by the factor $1/\sqrt{1-v^2/c^2}$. The result is that $\boldsymbol{\mathscr{E}}$ makes the same angle θ with respect to the X-axis as $\boldsymbol{r}$ makes. Thus relative to observer O, the electric field is also along the radial direction. However, the field is no longer spherically symmetric relative to O. A direct calculation (see Example 4.12) shows that

$$\boldsymbol{\mathscr{E}}=\frac{q}{4\pi\epsilon_0 r^2}\frac{1-v^2/c^2}{[1-(v^2/c^2)\sin^2\theta]^{3/2}}\boldsymbol{u}_r. \tag{4.27}$$

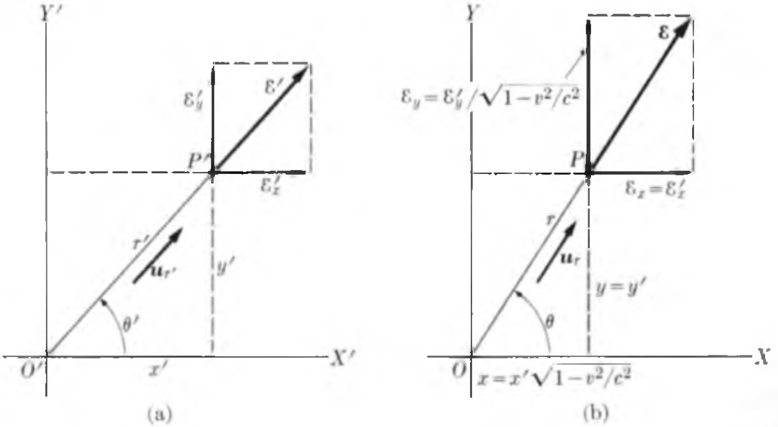


Fig. 4-19. Relativistic transformation of the components of the electric field produced by a charge q at rest relative to O' and located at O'.

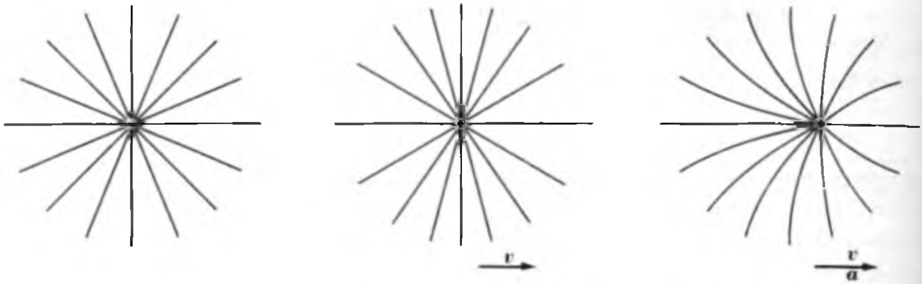


Fig. 4-20. Electric lines of force of a charge at rest and charges in motion.

The factor containing $\sin\theta$ makes the electric field depend on the direction of the position vector $\boldsymbol{r}$. Thus for equal distances from the charge the electric field is stronger in the equatorial plane ($\theta=\pi/2$), perpendicular to the direction of motion, than along the direction of motion ($\theta=0$). This field contrasts with the electric field produced by a charge at rest; the latter field is spherically symmetric. This situation has been illustrated in Fig. 4-20(a) and (b), in which the spacing of the lines indicates the relative strength of the field.

If we apply the relation $\mathscr{B}=\boldsymbol{v}\times\mathscr{E}/c^2$, which has been proven of general validity, and if we use Eq. (4.27) for $\mathscr{E}$, the magnetic field of a moving charge is

$$\mathscr{B}=\frac{\mu_0 q}{4\pi r^2}\frac{1-v^2/c^2}{[1-(v^2/c^2)\sin^2\theta]^{3/2}}\,\boldsymbol{v}\times\boldsymbol{u}_r. \tag{4.28}$$

This expression reduces to the nonrelativistic equation for the magnetic field when v is very small compared with c. Remember that Eqs. (4.27) and (4.28) are valid only for a charge with uniform motion. If the charge is accelerated, the electric field assumes a shape similar to that in Fig. 4-20(c), and the mathematical expressions become more complex.

Example 4.12. The electric field of a uniformly moving charge.

▼ Note from Fig. 4-19(a) that $\mathscr{E}'$ makes an angle θ' with $O'X'$ and that $\cos\theta'=x'/r'$, $\sin\theta'=y'/r'$. Then the components of $\mathscr{E}'$ are

$$\mathscr{E}'_x=\mathscr{E}'\cos\theta'=\frac{q}{4\pi\epsilon_0}\frac{x'}{r'^3}, \qquad \mathscr{E}'_y=\mathscr{E}'\sin\theta'=\frac{q}{4\pi\epsilon_0}\frac{y'}{r'^3}. \tag{4.29}$$

Using Eq. (4.25) and the fact that $x=x'\sqrt{1-v^2/c^2}$ and $y=y'$ according to the Lorentz transformation permits writing the components of the field $\mathscr{E}$ observed by O as

$$\mathscr{E}_x=\mathscr{E}'_x=\frac{q}{4\pi\epsilon_0}\frac{x}{\sqrt{1-v^2/c^2}\,r'^3},$$

$$\mathscr{E}_y = \frac{\mathscr{E}'_y}{\sqrt{1-v^2/c^2}} = \frac{q}{4\pi\epsilon_0} \frac{y}{\sqrt{1-v^2/c^2}\, r'^3}.$$

In vector notation

$$\mathscr{E} = \frac{q\boldsymbol{r}}{4\pi\epsilon_0\sqrt{1-v^2/c^2}\, r'^3}, \tag{4.30}$$

showing that the field $\mathscr{E}$ is along the radial direction in the XYZ frame. Now

$$r'^2 = x'^2 + y'^2 = \frac{x^2}{1-v^2/c^2} + y^2 = \frac{r^2 - (v^2/c^2)y^2}{1-v^2/c^2}$$

and $y^2 = r^2 \sin^2 \theta$. Therefore

$$r' = \frac{r[1-(v^2/c^2)\sin^2\theta]^{1/2}}{\sqrt{1-v^2/c^2}}.$$

If this relation is used to eliminate r' in Eq. (4.30), the electric field is

$$\mathscr{E} = \frac{q}{4\pi\epsilon_0 r^3} \frac{(1-v^2/c^2)\boldsymbol{r}}{[1-(v^2/c^2)\sin^2\theta]^{3/2}} = \frac{q}{4\pi\epsilon_0 r^2} \frac{1-v^2/c^2}{[1-(v^2/c^2)\sin^2\theta]^{3/2}} \boldsymbol{u}_r,$$

which is just the result given previously. ▲

4.8 Electromagnetic Interaction between Two Moving Charges

The discussion of magnetic interactions has departed from the procedure followed for gravitational and electrical interactions. In those cases the analyses began by discussing the interaction between two particles and afterwards introduced the concept of field. However in this chapter the concept of magnetic field was first introduced in an operational form concerned with the force [Eq. (4.1)] exerted on a moving charge. Then the magnetic fields produced by *moving charges* were computed. So far, there has been no expression for the electromagnetic interaction between two moving charges. One reason for this difference in procedure is the following: at velocities small relative to the velocity of light, the gravitational and electrical interactions depend exclusively on the distance between the two interacting particles; that is, the interactions are *static forces*. These interactions can exert forces on particles at rest, and therefore the physical situation can be discussed under steady or *time-independent* conditions. On the other hand the magnetic interaction depends on the motion of the interacting particles, regardless of their relative velocity; that is, the magnetic interaction is a *velocity-dependent force*. At a given point the magnetic field of a charge moving relative to the observer depends on the velocity of the charge as well as on the distance between the charge and the observer; but the distance is changing since the charge is moving, and therefore the magnetic field (as well as the

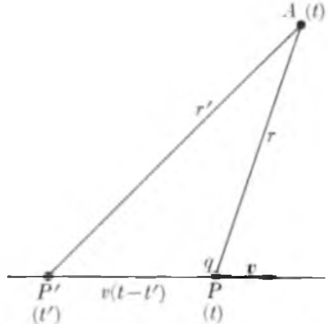


Fig. 4-21. Retardation effect caused by the finite velocity of propagation of electromagnetic fields.

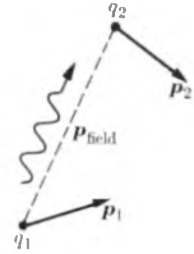


Figure 4-22

electric field) at a particular point is changing with time; i.e., relative to the observer, the magnetic field of the moving charge is *time dependent*.

A new element therefore enters into the physical picture: the *velocity of propagation of an interaction*. One possible approach is to assume that the particles *interact at a distance*. That is, if the charge q (Fig. 4-21) is moving with velocity $\boldsymbol{v}$, the electromagnetic field caused by q at A at some time t is the result of the physical situation in the surrounding space when the charge is at position P at time t, *simultaneously* with the observation at A. In other words, we may assume that the electromagnetic interaction *propagates instantaneously*, or with infinite velocity.

Another reasonable (and perhaps more acceptable) assumption is that the electromagnetic interaction is the result of certain "signals" exchanged between the interacting particles, and that the signal *propagates with a finite velocity* c that requires a certain time to reach a particular point in space. If the charge is at rest, the finite velocity of propagation of the signal is irrelevant because the physical circumstances are not changing with time. However for a moving charge the situation is different, and the field observed at point A at time t does not correspond to the simultaneous position of the charge at P, but to an earlier or *retarded* position P' at time t' such that $t-t'$ is the time required for the signal to travel from P' to A with velocity c. Obviously $P'P = v(t-t')$.

Electromagnetic interactions do propagate with the finite velocity c, given by $1/\sqrt{\epsilon_0\mu_0}$ as will be shown in Chapter 8. This experimental fact rules out action at a distance, and therefore the analysis of the electromagnetic field produced by a moving charge requires the second approach given above. Because c has such a large value, the retardation effect is negligible unless the particles move very fast. For that reason retardation was not considered when the motion of charges was discussed in Chapter 1. Those charges were assumed to move very slowly, and thus PP' is very small compared with PA. It can be verified that the relativistic expressions (4.21) and (4.23) for the electric and magnetic fields of a moving charge already incorporate the effect of retardation. Similar retardation effects should exist for the gravitational interaction between two masses in relative motion. However, the velocity of propagation of gravitational signals has not yet been observed.

Consider two charges q_1 and q_2 moving with velocities $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ relative to an inertial observer O. The force charge q_1 produces on q_2 as measured by O is $\boldsymbol{F}_2 = q_2(\boldsymbol{\mathscr{E}}_1 + \boldsymbol{v}_2 \times \boldsymbol{\mathscr{B}}_1)$ where $\boldsymbol{\mathscr{E}}_1$ and $\boldsymbol{\mathscr{B}}_1$ are the electric and magnetic fields measured by O and produced by q_1 at the position occupied by q_2. On the other hand the force that charge q_2 produces on q_1 as measured by O is $\boldsymbol{F}_1 = q_1(\boldsymbol{\mathscr{E}}_2 + \boldsymbol{v}_1 \times \boldsymbol{\mathscr{B}}_2)$. In general these two have different directions and magnitudes. Therefore it follows that

the forces between two moving charges are neither equal in magnitude nor opposite in direction.

In other words it appears that the law of action and reaction does not hold in the presence of magnetic interactions. This statement in turn implies that the principles of conservation of momentum, of angular momentum, and of energy would not hold for a system of two charged particles in motion. This *apparent* failure of such important laws is due to the following fact. When the law of conservation of momentum for two interacting particles is written as $\boldsymbol{p}_1 + \boldsymbol{p}_2 = \text{const}$, we are considering that $\boldsymbol{p}_1$ and $\boldsymbol{p}_2$ are measured *simultaneously* by O: i.e., at the same time relative to O. However in the presence of an interaction propagating with a finite velocity, the retardation effect requires that the rate of change of momentum of one particle at a given time is related to the change of momentum of the other particle not at the same time but rather *at an earlier time*, and conversely. Therefore it is not reasonable to expect $\boldsymbol{p}_1 + \boldsymbol{p}_2$ to be constant if they are evaluated at the same time.

The student may recall that the result of an interaction may be described as an exchange of momentum between the two particles. To restore the law of conservation of momentum, the momentum that is being exchanged between the two particles and that at a given time is traveling between them with a finite velocity must be taken into account. That is, the momentum "in flight" must be taken into account. For the electromagnetic interaction the electromagnetic field carries this momentum, labeled $\boldsymbol{p}_{\text{field}}$ (Fig. 4-22). Thus the law of conservation of momentum requires that

$$\boldsymbol{p}_1 + \boldsymbol{p}_2 + \boldsymbol{p}_{\text{field}} = \text{const}. \tag{4.31}$$

Similarly a certain angular momentum and energy must be attributed to the electromagnetic field in order to restore these two conservation principles. We shall delay until Chapter 8 a discussion of how the momentum, angular momentum, and energy associated with the electromagnetic field are obtained.

Example 4.13. Comparison of the magnetic interaction between two charges with the electric interaction between them.

▼ Since only orders of magnitude are desired, we shall simplify the writing of the formulas. Thus given charges q and q', the electric force produced by q' on q is $q\mathscr{E}$. The magnetic field produced by q' on q from Eq. (4.17) is of the order of magnitude of $v'\mathscr{E}/c^2$. The magnetic force on charge q from Eq. (4.1) is of the order of magnitude of $qv(v'\mathscr{E}/c^2)=(vv'/c^2)q\mathscr{E}$. Therefore

$$\frac{\text{Magnetic force}}{\text{Electric force}} \approx \frac{vv'}{c^2}.$$

If the velocities of the charges are small compared with the velocity of light c, the magnetic force is negligible compared with the electric force, and in many cases can be ignored. Thus, in a sense, magnetism is a consequence of the finite velocity of propagation of electromagnetic interactions. For example if the charges have a velocity of the order of 10^6 m s^{-1}, corresponding to the orbital speed of electrons in atoms,

$$\frac{\text{Magnetic force}}{\text{Electric force}} \approx 10^{-4}.$$

In spite of its small value relative to the electric force, the magnetic force is the one used in electric motors and many other engineering devices for the following reason. Matter is normally electrically neutral, and the net electric force between two bodies is zero. For example when two wires are placed side by side, the net electric force between them is zero. If the wires are moved as a whole, the positive and negative charges move in the same direction so that the net current in each is zero, and thus the net magnetic field is also zero. This movement results in no force between the wires; but a potential difference applied to the wires results in a motion of the negative charges relative to the positive and produces a net current in each wire and a net magnetic field. Since the number of free electrons in a conductor is very large, their cumulative effect produces a large magnetic field even if their velocities are small; and the result is an appreciable magnetic force between the wires.

Although weak compared with electric force, magnetic force is still very strong compared with gravitational interaction. Recall the discussion of relative strengths of the four forces: the electric interaction is about 10^{36} stronger than the gravitational interaction. Therefore we may say that

$$\frac{\text{Magnetic interaction}}{\text{Gravitational interaction}} \approx 10^{36}\frac{vv'}{c^2}.$$

For velocities comparable to those of orbiting electrons, this ratio is of the order of 10^{31}. ▲

Problems

4.1 Electrons with a velocity of 10^6 m s^{-1} enter a region in which a magnetic field exists. (a) Find the intensity of the magnetic field if the electron describes a path having a radius of 0.10 m. (b) Find the angular velocity of the electron.

4.2 From rest, protons are accelerated through a potential difference of 10^6 V. These are then shot into a region of uniform magnetic field of 2 T; the trajectory is perpendicular to the field. What are (a) the trajectory radius and (b) the angular velocity of the protons?

4.3 In a magnetic field a proton is in motion at an angle of 30° with respect to the field. The velocity is 10^7 m s^{-1} and the field strength is 1.5 T. Compute (a) the radius of the helix of motion, (b) the distance of advance per revolution, or pitch of the helix, and (c) the frequency of rotation in the field.

4.4 A deuteron (an isotope of hydrogen whose mass is very nearly 2 amu) travels in a circular path of radius 0.4 m in a magnetic field of

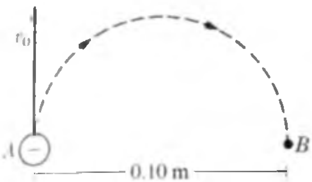


Figure 4-23

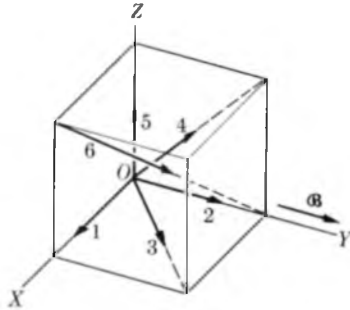


Figure 4-24

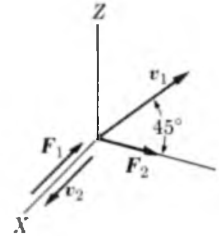


Figure 4-25

1.5 T. (a) Compute the speed of the deuteron. (b) Determine the time required for it to make one-half a revolution. (c) Through what potential difference would the deuteron have to be accelerated to acquire the velocity of part (a)?

4.5 A proton having a kinetic energy of 30 MeV moves transverse to a magnetic field of 1.5 T. Determine (a) the radius of the path and (b) the period of revolution. Note that the proton can be treated classically.

4.6 (a) What is the magnetic field required to force a 30-GeV proton to describe a path 100 m in radius? (b) Find the angular velocity. Note that the calculation must be relativistic.

4.7 A singly charged ^{7}Li ion has a mass of 1.16×10^{-26} kg. The ion is accelerated through a potential difference of 500 V and then enters a magnetic field of 0.4 T; the ion moves perpendicular to the field. What is the radius of the ion's path in the magnetic field?

4.8 An electron at point A in Fig. 4-23 has a velocity v_0 of 10^7 m s^{-1}. Calculate (a) the magnitude and the direction of the magnetic field that will cause the electron to follow the semicircular path from A to B, and (b) the time required for the electron to move from A to B.

4.9 One of the processes for separating the isotopes ^{235}U and ^{238}U is based on the difference of radii of their paths in a magnetic field. Assume that singly ionized atoms of U start from a common source and move perpendicular to a uniform field. Find the maximum spatial separation of the beams when the radius of curvature of the ^{235}U beam is 0.5 m in a field of 1.5 T (a) if the energies of the two isotopes are the same and (b) if the velocities are the same. For the purposes of this problem, the superscript on each isotope may be identified with the mass of the atom in amu.

4.10 A uniform magnetic field $\mathscr{B}$ lies in the Y-direction as shown in Fig. 4-24. Find the magnitude and the direction of the force on a charge q whose instantaneous velocity is $\boldsymbol{v}$ for each of the directions shown in the figure. (The figure is a cube.)

4.11 A particle of mass m and charge q moves with a velocity v_0 perpendicular to a uniform magnetic field. Express as a function of time the components of the velocity and the coordinates of the particle referred to the center of the path.

4.12 Repeat Problem 4.11 for a particle whose velocity makes an angle α with the magnetic field.

4.13 A particle carries a charge of 4×10^{-9} C. When the particle moves with a velocity $\boldsymbol{v}_1$ of 3×10^4 m s^{-1} at 45° above the Y-axis in the YZ-plane, a uniform magnetic field exerts a force $\boldsymbol{F}_1$ along the X-axis. When the particle moves with a velocity $\boldsymbol{v}_2$ of 2×10^4 m s^{-1} along the X-axis, the particle experiences a force $\boldsymbol{F}_2$ of 4×10^{-5} N along the Y-axis. What are the magnitude and the direction of the magnetic field? (See Fig. 4-25.)

4.14 Charged particles are shot into a region of crossed electric and magnetic fields. The incident particle velocity is normal to the plane of the two fields, and the fields are normal to each other. The magnetic field strength is

0.1 T. The electric field is generated between a pair of equal and oppositely charged parallel plates, placed 0.02 m apart. When the potential difference between the plates is 300 V, there is no deflection of the particles. What is the particle velocity?

4.15 (a) What is the velocity of a beam of electrons when the simultaneous influence of an electric field of intensity 3.4×10^5 V m^{-1} and a magnetic field of 2×10^{-2} T, both fields being normal to the beam and to each other, produces no deflection of the electrons? (b) Show in a diagram the relative orientation of the vectors v, $\mathscr{E}$, and $\mathscr{B}$.

4.16 A particle having a mass of 5×10^{-4} kg carries a charge of 2.5×10^{-8} C. The particle is given an initial horizontal velocity of 6×10^4 m s^{-1}. What is the magnitude and the direction of the minimum magnetic field that will keep the particle moving in a horizontal direction and balance the earth's gravitational force?

4.17 In a mass spectrometer such as that illustrated in Fig. 4-12, a potential difference of 1000 V makes the single ionized ions of ^{24}Mg describe a path of radius R. (a) What will be the radius described by ^{25}Mg ions if they are accelerated through the same potential? (b) What potential difference would make the ^{25}Mg ions describe a path of the same radius R? (Assume that the masses in amu are the same as the mass numbers in the superscript to the left of the chemical symbol.)

4.18 A mass spectrometer has an accelerating voltage of 5 keV and a magnetic field of 10^{-2} T. (a) Compute the velocity of the ions to see if it will be necessary to use a relativistic correction. (b) Find the distance between the two isotopes of zinc, ^{68}Zn and ^{70}Zn. By distance we mean the separation of the two spots that appear on the emulsion of the photographic plate after the singly charged ions ^{68}Zn and ^{70}Zn are first accelerated and then turned around in a half circle. See Fig. 4-12. [*Hint:* Do not find the individual radii; rather write an equation to find the separation directly.]

4.19 Protons in a cyclotron, just before they emerge, describe a circle of radius 0.40 m. The frequency of the alternating potential between the dees is 10^7 Hz. Neglecting relativistic effects, compute (a) the magnetic field, (b) the velocity of the protons, (c) the energy of the protons in J and in MeV, and (d) the minimum number of complete turns of the protons if the peak value of the potential between the dees is 20 keV.

4.20 Repeat Problem 4.19 for a deuteron and for an alpha particle (helium nucleus). Their respective masses are 2.014 amu and 4.003 amu.

4.21 The magnetic field in a cyclotron accelerating protons is 1.5 T. (a) How many times per second should the potential across the dees reverse? (b) The maximum radius of the cyclotron is 0.35 m. What is the maximum velocity of the proton? (c) Through what potential difference would the proton have to be accelerated to give it the maximum cyclotron velocity?

4.22 Deuterons in a cyclotron describe a circle of radius 32.0 cm just before emerging from the dees. The frequency of the applied alternating voltage is 10^7 Hz. Find (a) the magnetic field, (b) the energy and (c) the speed of the deuterons when they emerge. The mass of a deuteron is 2.014 amu.

4.23 A cathode ray tube is placed in a uniform magnetic field $\mathscr{B}$ with the axis of the tube parallel to the lines of force. If electrons emerging from the gun with a velocity v make an angle θ with the axis as they pass through the origin O so that their trajectory is a helix, show (a) that they will touch the axis again at the time

$$t = \frac{2\pi m}{\mathscr{B}q},$$

(b) that the coordinate of the point of touching is

$$x = \frac{2\pi m v \cos\theta}{\mathscr{B}q},$$

and (c) that for small values of θ, the coordinate of the point of crossing or touching the axis is independent of θ.

4.24 The arrangement in Problem 4.23 is called a *magnetic lens.* (a) Why? (b) How do the trajectories of the electrons passing through the origin at an angle θ above the axis differ from those directed at an angle θ below the axis?

4.25 Protons with an energy of 3 MeV are injected at a small angle with respect to a uniform magnetic field of 1 T. At what distance will the particles return to a common point of intersection with the axis?

4.26 Relative to observer O, observer O' moves with a velocity v parallel to the common X-axis. Two charges q_1 and q_2 are at rest relative to O', are separated the distance r', and are placed along the X-axis as measured by O'. (a) Find the forces on each charge as recorded by O' and O. (b) Assume that the charges are on the Y'-axis, and repeat the problem.

4.27 Referring to Eq. (4.27), which gives the electric field of a point charge, find the ratio between the electric field in a plane through the charge perpendicular to the direction of motion and the field along the direction of motion for points at the same distance from the charge. Consider values of v/c equal to (a) 0, (b) 0.1, (c) 0.5, and (d) 0.9.

4.28 Evaluate the ratio between the relativistic and the nonrelativistic values of the electric field produced by a moving charge at a point on the plane through the charge perpendicular to the direction of motion. Consider values of v/c equal to (a) 0, (b) 0.1, (c) 0.5, and (d) 0.9.

4.29 Evaluate the ratio between the relativistic and the nonrelativistic values of the magnetic field produced by a moving charge at a point on the plane through the charge perpendicular to the direction of motion. Consider values of v/c equal to (a) 0, (b) 0.1, (c) 0.5, and (d) 0.9.

4.30 Consider two electrons moving in straight parallel paths separated by 10^{-4} m. (a) If the electrons are moving side by side at the same velocity of 10^6 m s^{-1}, find the electric and magnetic forces between them as seen by a laboratory observer (assume that 10^6 m s^{-1} can be considered a nonrelativistic velocity). (b) What is the force according to an observer moving with the electrons? (c) Repeat the inquiry above for the case of velocity 2.4×10^8 m s^{-1}, which is relativistic.

CHALLENGING PROBLEMS

4.31 Two equally charged particles, A and B, with equal kinetic energies enter a constant magnetic field. The angle between the field and the velocity for each particle is 90°. If the mass of A is 4 times the mass of B, find the ratio of the radius of the circular path of A to the radius of the circular path of B. (AP-B: 1971)

4.32 A charged particle accelerated from rest through a potential difference V enters a uniform magnetic field perpendicular to its direction of motion and moves in a circular path of radius R. If, instead, the potential difference were $V/3$, determine the radius of the path. (AP-B: 1972)

4.33 An electron in a vacuum chamber is accelerated through an electric potential difference, and then enters the space between two charged, parallel metal plates as shown in Fig. 4-26. The distance between the plates is 0.010 meter and the potential difference between the plates is 300 volts. (a) If the speed of the electron as it enters the space between the plates is 6.0×10^6 meters per second, through what potential difference V_0 has the electron been accelerated? (b) If between the plates there also exists a constant magnetic field acting into the page as shown in the diagram, what magnitude of the magnetic field $\mathscr{B}$ will allow the electron to move undeflected between the plates? (AP-B; 1973)

4.34 Electrons of various nonrelativistic speeds are moving in a plane perpendicular to a

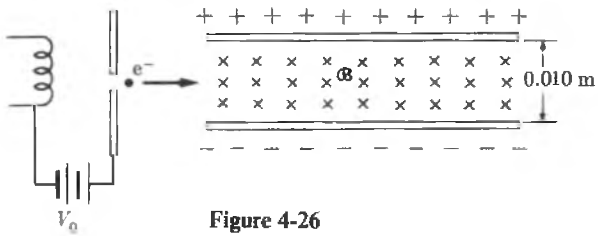


Figure 4-26

uniform magnetic field 𝔅. Because of the magnetic force, the electrons move in circles of various radii. Show that the time required to travel around one complete circle is the same for all electrons, regardless of their speeds. (AP-B; 1974)

4.35 In a mass spectrometer, singly charged ^{16}O ions are first accelerated electrostatically through a voltage V to a speed v_0; they then enter a region of uniform magnetic field 𝔅 directed out of the plane of the paper as shown in Fig. 4-27. (a) If singly charged ^{32}S ions are substituted for the ^{16}O ions, what will be their speed for the same accelerating voltage? (b) When ^{32}S is substituted for ^{16}O in part (a), determine by what factor the radius of curvature of the ions' path in the magnetic field changes. (AP-B; 1975)

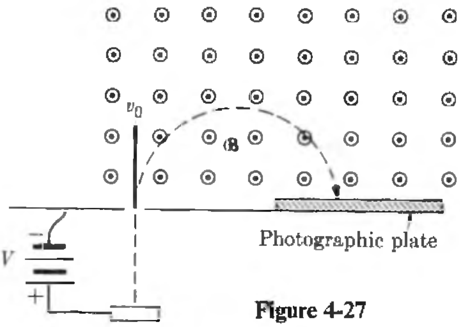


Figure 4-27

4.36 An ion of mass m and charge of known magnitude q is observed to move in a straight line through a region of space in which a uniform magnetic field 𝔅 points out of the paper and a uniform electric field ℰ points toward the top edge of the paper, as in region I (Fig. 4-28a). The particle travels into region II in which the same magnetic field is present, but the electric field is zero. In region II the ion moves in a circular path as shown. (a) Indicate on a diagram, as shown in Fig. 4-28b, the direction of the force on the ion at point P_2 in region II. (b) Is the ion positively or negatively charged? Explain clearly the reasoning on which you base your conclusion. (c) Indicate and label clearly on a diagram, as shown in Fig. 4-28c, the forces which act on the ion at point P_1 in region I. (d) Find an expression for

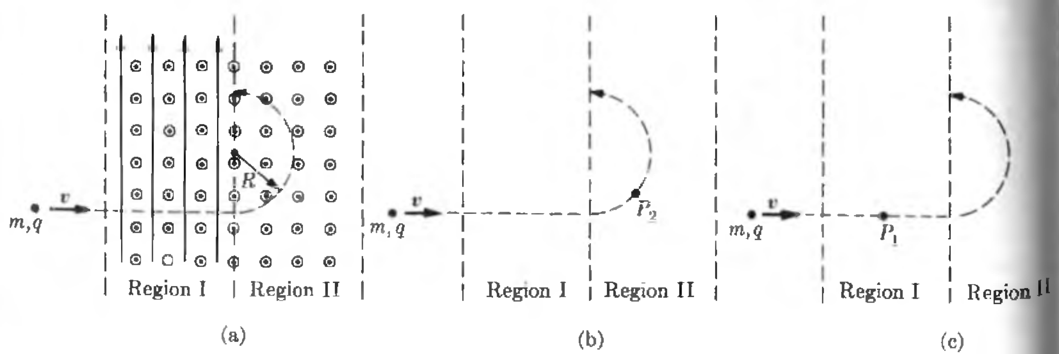


Figure 4-28

the ion's speed v at point P_1 in terms of $\mathscr{E}$ and $\mathscr{B}$. (e) Starting with Newton's law, derive an expression for the mass m of the ion in terms of $\mathscr{B}$, $\mathscr{E}$, q, and R. (AP-B: 1976)

4.37 An electron is accelerated from rest through a potential difference of magnitude V between infinite parallel plates P_1 and P_2. The electron then passes into a region of uniform magnetic field strength $\mathscr{B}$ which exists everywhere to the right of plate P_2. The magnetic field is directed into the page (Fig. 4-29). (a) On a diagram, clearly indicate the direction of the electric field between the plates. (b) In terms of V and the electron's mass and charge, determine the electron's speed at plate P_2. (c) Describe in detail the motion of the electron through the magnetic field and explain why the electron moves this way. (d) If the magnetic field remains unchanged, what could be done to cause the electron to follow a straight-line path to the right of plate P_2? (AP-B: 1977)

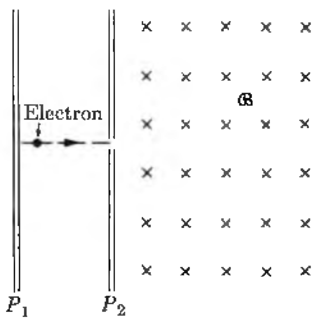


Figure 4-29

4.38 Electrons are accelerated from rest in an electron gun between two plates that have a voltage V_g across them. The electrons then move into the region between two other parallel plates of separation d that have voltage V_p across them. The electrons are projected into this region at an angle θ to the plates as shown in Fig. 4-30. Assume that the entire apparatus is in vacuum and that $V_p > V_g$. Display all results in terms that include d, V_g, V_p, θ, e (the magnitude of the electron charge), and m_0 (the electron mass). (a) Develop an equation for the speed v_g with which the electrons leave the electron gun. (b) Develop an equation for the maximum distance y_{max} that the electrons travel above the lower plate. Suppose that a magnetic field directed into the plane of the paper is introduced in the region between the upper plates. (c) How will the speed with which the electrons strike the lower plate be affected? Explain. (d) Sketch on a diagram a trajectory that an electron might follow with the magnetic field present. Account qualitatively for the difference between the new and old trajectory. (AP-C: 1978)

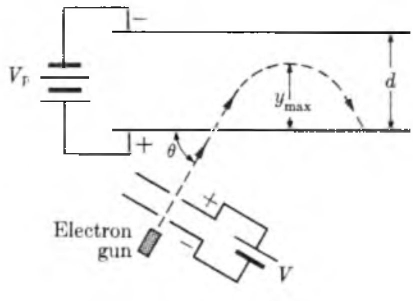


Figure 4-30

4.39 Determine the magnitude and the direction of the force on a proton in each of the following situations. Describe qualitatively the path followed by the proton in each situation and sketch the path on diagrams copied from Fig. 4-31. Neglect gravity. (a) The proton is released from rest at the point P in an electric field $\mathscr{E}$ having intensity 10^4 newtons per coulomb and directed up in the plane of the page as shown in Fig. 4-31a. (b) In the same electric field as in part (a), the proton at point P has velocity $v = 10^5$ meters per second directed to the right as shown in Fig. 4-31b. (c) The proton is released from rest at point P in a magnetic field $\mathscr{B}$ having intensity 10^{-1} tesla and directed into the page as shown in Fig. 4-31c. (d) In the same magnetic field as in

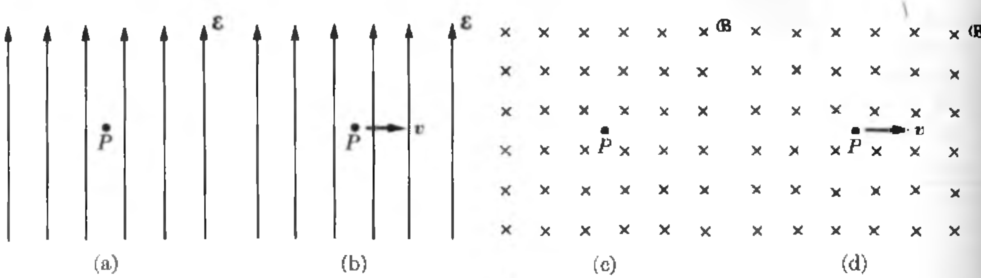


Figure 4-31

part (c), the proton at point P has velocity $v = 10^5$ meters per second directed to the right as shown in Fig. 4-31d. (AP-B: 1979)

4.40 A uniform magnetic field exists in a region of space. Two experiments were done to discover the direction of the field and the following results were obtained. Experiment I: A proton moving to the right with instantaneous velocity v_1 experienced a force F_1 directed into the page as shown in Fig. 4-32a. Experiment II: A proton moving out of the page with instantaneous velocity v_2 experienced a force F_2 in the plane of the page as shown in Fig. 4-32b. (a) State the direction of the magnetic field and show that your choice accounts for the directions of the forces in both experiments. (b) In which experiment did the proton describe a circular orbit? Explain your choice and determine the radius of the circular orbit in terms of the given force and velocity for the proton and the proton mass m. (c) Describe qualitatively the motion of the proton in the other experiment. (AP-C; 1979)

4.41 A proton having a kinetic energy of 30 GeV moves transverse to a magnetic field of 1.5 T. (a) Determine the radius of the path and the period of revolution. Note that the proton must be treated relativistically. Dempster's mass spectrometer, illustrated in Fig. 4-12, uses a magnetic field to separate ions having different masses but the *same energy*. Another arrangement is *Bainbridge's mass spectrometer* (Fig. 4-33), which separates ions having the *same velocity*. The ions, after crossing the slits, pass through a velocity selector composed of an electric field produced by the charged plates P and P', and a magnetic field $\mathscr{B}$ perpendicular to the electric field. Those ions that pass undeviated through the crossed fields enter into a region in which a second magnetic field $\mathscr{B}'$ exists, and are bent into circular orbits. A photographic plate P registers the arrival of the ions. (b) Show that $q/m = \mathscr{E}/r\mathscr{B}\mathscr{B}'$.

4.42 The electric field between the plates of the velocity selector in a Bainbridge mass spectrograph is 1.2×10^5 V m^{-1}, and both magnetic fields are 0.6 T. A stream of singly

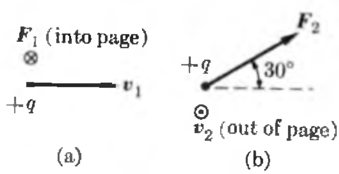


Figure 4-32

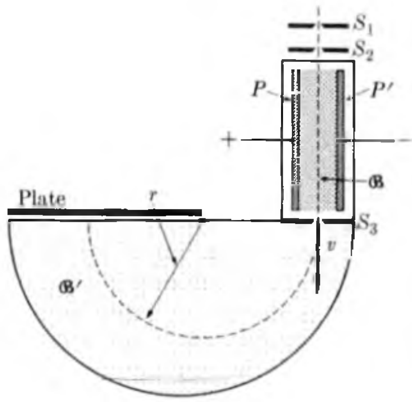


Figure 4-33

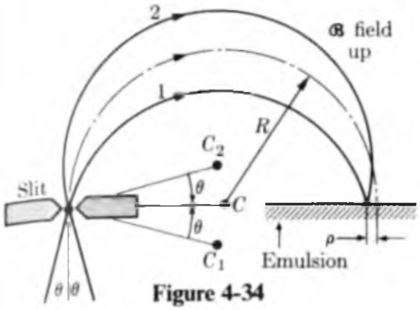


Figure 4-34

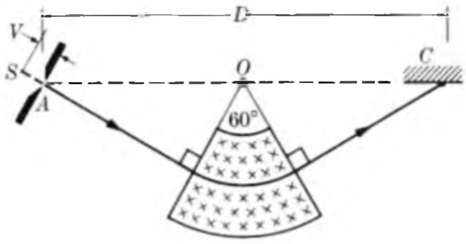


Figure 4-35

charged neon ions moves in a circular path of 7.28×10^{-2} m radius in the magnetic field. Determine the mass of the neon isotope.

4.43 Suppose that the electric intensity between the plates P and P' in Fig. 4-33 is 1.5×10^4 V m^{-1}, and both magnetic fields are 0.5 T. If the source contains the three isotopes of magnesium, ^{24}Mg, ^{25}Mg, and ^{26}Mg, and the ions are singly charged, find the distance between the lines formed by the three isotopes on the photographic plate. Assume that the isotopes' atomic masses in amu are equal to their mass numbers shown at the left of the chemical symbol.

4.44 In a mass spectrometer, such as that shown in Fig. 4-34, it is difficult to ensure that all particles arrive perpendicular to the slit. (a) If R is the radius of their path, show that those particles arriving at the slit making a small angle θ with the normal will arrive at the photographic plate at a distance ρ (approximately equal to $R\theta^2$) from those that fall perpendicularly. (b) What is the value of θ so that this separation is less than 0.1 percent of $2R$? (The situation described in this problem is called *magnetic focusing*.)

4.45 In the mass spectrograph of Fig. 4-35, ions accelerated by a potential difference between S and A fall on the magnetic field covering a sector of 60° and are sent toward a photographic emulsion. (a) Show that

$$\frac{q}{m} = \frac{32V}{\mathscr{B}^2 D^2}.$$

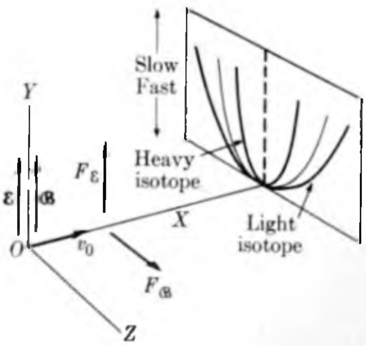
(b) Discuss the change in the position of C for a small deviation in the direction of incidence.

4.46 A particle of charge q and velocity v_0 (along the X-axis) enters a region in which a magnetic field exists (along the Y-axis). Show that if the velocity v_0 is large enough so that its change in direction is negligible and the magnetic force can be considered as constant and parallel to the Z-axis, the equation of the path of the particle is

$$z = \left(\frac{q\mathscr{B}}{2v_0 m}\right) x^2.$$

4.47 A particle of charge q and velocity v_0 (along the X-axis) enters a region (Fig. 4-36) in which uniform electric and magnetic fields exist in the same direction (along the Y-axis). Show that if the velocity v_0 is large enough so that its change in direction is negligible and the magnetic force can be considered as constant and parallel to the Z-axis, (a) the coordinates at time t are

$$x = v_0 t \qquad y = \tfrac{1}{2}\left(\frac{q\mathscr{E}}{m}\right)t^2 \qquad z = \tfrac{1}{2}\left(\frac{q v_0 \mathscr{B}}{m}\right)t^2.$$



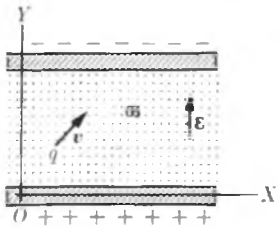


Figure 4-37

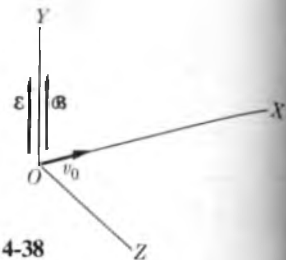


Figure 4-38

(b) By eliminating t and v_0 between these equations, obtain the relation

$$\frac{z^2}{y}=\frac{1}{2}\left(\frac{\mathcal{B}^2}{\mathcal{E}}\right)\left(\frac{q}{m}\right)^2 x^2.$$

The result has an application in one of the earliest mass spectrographs because if we insert a screen perpendicular to the X-axis (Fig. 4-36), all particles having the same ratio q/m will fall along a given parabola, irrespective of their initial velocity. Therefore there will be one parabola for each isotope present in the incoming beam.

4.48 A particle of charge q and mass m moves between two parallel charged plates separated a distance h. A uniform magnetic field is applied parallel to the plates and is directed in the Z-direction (Fig. 4-37). Initially the particle is at rest at the lower plate. (a) Write the equations of motion of the particle. (b) Show that at the distance y from the lower plate,

$$v_x=\left(\frac{q}{m}\right)\mathcal{B}y.$$

(c) Show that the magnitude of the velocity is

$$v^2=2\left(\frac{q}{m}\right)\mathcal{E}y.$$

(d) From the two preceding results, show that

$$v_y=\left(\frac{q}{m}\right)^{\frac{1}{2}}\left[2\mathcal{E}y-\left(\frac{q}{m}\right)\mathcal{B}^2y^2\right]^{\frac{1}{2}},$$

and (e) that the particle will just fail to reach the upper plate if

$$\mathcal{E}=\frac{1}{2}\left(\frac{q}{m}\right)\mathcal{B}^2h.$$

4.49 In a region in which there are uniform electric and magnetic fields in the same direction, a particle of charge q and mass m is injected with a velocity v_0 in a direction perpendicular to the common direction of the two fields. (a) Write the equation of motion in rectangular coordinates. (b) Show by direct substitution in the equation of motion that the components of the velocity at time t are

$$v_x=v_0\cos\left(\frac{q\mathcal{B}}{m}\right)t \qquad v_y=\left(\frac{q\mathcal{E}}{m}\right)t$$

$$v_z=\sin\left(\frac{q\mathcal{B}}{m}\right)t.$$

(c) From the previous result, obtain the coordinates of the particle at time t. (d) Make a plot of the path. (e) What would the motion of the particle be if the initial velocity of the particle were parallel to the fields? [*Hint:* For the answers given, the X-axis is in the direction of v_0, and the Y-axis is in the common direction of the two fields (Fig. 4-38).]

4.50 In a certain region there are uniform electric and magnetic fields perpendicular to each other. A particle is injected with a velocity v_0 parallel to the magnetic field. In rectangular coordinates (a) write the equation of motion of the particle. (b) Show by direct substitution that the components of the velocity at time t are

$$v_y=\left(\frac{\mathcal{E}}{\mathcal{B}}\right)\sin\left(\frac{q\mathcal{B}}{m}\right)t,$$

and

$$v_z=-\left(\frac{\mathcal{E}}{\mathcal{B}}\right)\left[1-\cos\left(\frac{q\mathcal{B}}{m}\right)t\right].$$

(c) From the previous result, derive the coordinates of the particle at time t. (d) Make a plot of the path. [*Hint:* The magnetic field points along the X-axis and the electric field is along the Y-axis.]

4.51 (a) Solve Problem 4.50 for a particle whose initial velocity is parallel to the electric field. (b) Verify that the components of the velocity are

$$v_x = 0,$$
$$v_y = v_0 \cos\left(\frac{q\mathscr{B}}{m}\right)t + \left(\frac{\mathscr{E}}{\mathscr{B}}\right)\sin\left(\frac{q\mathscr{B}}{m}\right)t,$$

and

$$v_z = -\left(\frac{\mathscr{E}}{\mathscr{B}}\right)\left[1-\cos\left(\frac{q\mathscr{B}}{m}\right)t\right] - v_0 \sin\left(\frac{q\mathscr{B}}{m}\right)t.$$

4.52 (a) Solve Problem 4.50 for a particle whose initial velocity is perpendicular to both fields. (b) Verify that the components of the velocity are

$$v_x = 0,$$
$$v_y = \left(\frac{\mathscr{E}}{\mathscr{B}} + v_0\right)\sin\left(\frac{q\mathscr{B}}{m}\right)t,$$

and

$$v_z = -\left(\frac{\mathscr{E}}{\mathscr{B}}\right) + \left(\frac{\mathscr{E}}{\mathscr{B}} + v_0\right)\cos\left(\frac{q\mathscr{B}}{m}\right)t.$$

(c) Show that in order for the particles to move through the field undeflected, it is necessary that $v_0 = -\mathscr{E}/\mathscr{B}$. (d) Compare the result with the statements made in Section 4.4.

4.53 (a) Referring to Problem 4.50, verify that when the velocity has an initial arbitrary direction, the components of the velocity at time t are

$$v_x = v_{0x},$$
$$v_y = \left(\frac{\mathscr{E}}{\mathscr{B}} + v_{0z}\right)\sin\left(\frac{q\mathscr{B}}{m}\right)t + v_{0y}\cos\left(\frac{q\mathscr{B}}{m}\right)t,$$

and

$$v_z = -\frac{\mathscr{E}}{\mathscr{B}} + \left(\frac{\mathscr{E}}{\mathscr{B}} + v_{0z}\right)\cos\left(\frac{q\mathscr{B}}{m}\right)t - v_{0y}\sin\left(\frac{q\mathscr{B}}{m}\right)t.$$

(b) Obtain (by integration) the coordinates of the particle and discuss the path. (c) Compare with the results of Problems 4.51 and 4.52.

4.54 Referring to Problem 4.49 (a) show that when $(q\mathscr{B}/m)t \ll 1$, the coordinates of the particle can be expressed as

$$x = v_0 t \qquad y = \left(\frac{q\mathscr{E}}{2m}\right)t^2 \qquad z = \left(\frac{v_0 q\mathscr{B}}{2m}\right)t^2,$$

in agreement with Problem 4.47.

4.55 A proton having an energy of 30 GeV passes at a distance of 10^{-7} m from an ion. Since the proton must be considered relativistically, (a) find the angle α for which the electric field at the ion is 50 percent of the field when the proton is at its distance of closest approach to the ion. (b) Estimate the duration of the impulse to which the ion is subject and its change in momentum, considering it is essentially the result of the field within the angle found in (a). Repeat if the passing particle is an electron with the same energy, instead of a proton. (See Fig. 4-39.)

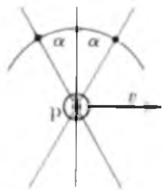


Figure 4-39

4.56 Using the relativistic expression (4.28) for the magnetic field of a moving charge, obtain the expression for the magnetic field of a rectilinear current.

4.57 Using the general rule for the relativistic transformation of force, obtain the relativistic transformations of the electromagnetic fields, Eqs. (4.21) and (4.23).

4.58 Using Eqs. (4.21) and (4.23), prove that the quantities $\mathscr{E}\cdot\mathscr{B}$ and $\mathscr{E}^2 - \mathscr{B}^2$ are invariant with respect to a Lorentz transformation.

4.59 A particle of charge q and mass m moves in a region in which an electric field $\mathscr{E}$ and a magnetic field $\mathscr{B}$ are present. (a) Show that if the motion of the particle is referred to a frame of reference rotating with the Larmor frequency of the particle, $\omega_L = -q\mathscr{B}/2m$ (see Eq. 4.7), its equation of motion becomes

$$ma' = q\left[\mathscr{E} + \left(\frac{m}{q}\right)\omega_L \times (\omega_L \times r)\right].$$

(b) Estimate the value of ω_L for an electron and verify that the last term is negligible. Under this approximation the equation of motion of the particle relative to the rotating frame becomes $ma' = q\mathscr{E}$. Comparing this result with Example 4.10 shows us how to eliminate the effect of a magnetic field. [*Hint:* Express the acceleration and the velocity of the particle relative to the rotating frame.]

CHAPTER FIVE

MAGNETIC FIELDS AND ELECTRIC CURRENTS

5.1 Introduction

Although the discussion in the previous chapter presented an analysis of the forces that magnetic fields exert on moving charges and of the magnetic fields that moving charges produce, electromagnetism evolved historically in an entirely different way. As previously mentioned, the Danish physicist Hans C. Oersted (1770–1851), professor at the University of Copenhagen, discovered in 1819 that electric currents exert forces on magnets; he thereby proved that electric currents produce magnetic fields. Oersted placed a rectilinear conductor directly above and parallel to a compass needle. To his great surprise he observed that when there was a current in the conductor, the compass needle swung and became perpendicular to the current. Immediately after Oersted published his results in 1820, several other scientists began to study the interactions between magnetic fields and electric currents. Among the early investigators were André M. Ampère (1775–1836) and Pierre Laplace (1749–1827), the French scientists who developed the quantitative theory of magnetic interactions of currents and introduced the terminology still used today. Not until the end of the 19th century was the relation between magnetic fields and moving electric charges established, partly as a result of the 1878 experiments of the American physicist H. A. Rowland (1848–1901).

5.2 Magnetic Force on an Electric Current

An electric current is a stream of electric charges moving in vacuum or through a conducting medium. The intensity of the electric current has been defined as the charge passing through a section of the conductor per unit time. Consider a cross section of a conductor through which particles with charge q are moving with velocity $\boldsymbol{v}$. If there are n particles per unit volume, the total number of particles passing through the unit area per unit time is $n\boldsymbol{v}$; and the *current density*, defined as the charge passing through the unit area per unit time, is the vector

$$\boldsymbol{j} = nq\boldsymbol{v}. \tag{5.1}$$

If S is the conductor's cross-sectional area, oriented perpendicular to $\boldsymbol{j}$, the current is the scalar

$$I = \boldsymbol{j} \cdot \boldsymbol{S} = nqvS. \tag{5.2}$$

Suppose now that the conductor is in a magnetic field. The force on each charge is given by $\boldsymbol{F} = q(\boldsymbol{v} \times \mathscr{B})$; and since there are n particles per unit volume, the *force per unit volume* $\boldsymbol{f}$ is

$$\boldsymbol{f} = nq\boldsymbol{v} \times \mathscr{B} = \boldsymbol{j} \times \mathscr{B}. \tag{5.3}$$

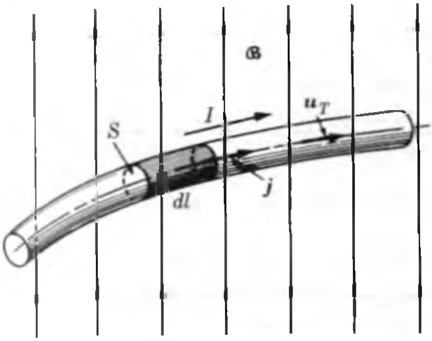


Figure 5-1

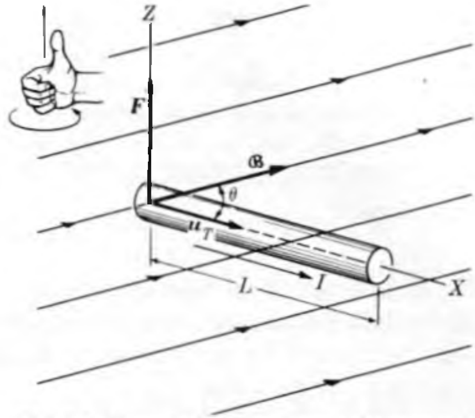


Fig. 5-2. Vector relation between the magnetic force on a current-carrying conductor, the magnetic field, and the current. The force is perpendicular to the plane defined by $\boldsymbol{u}_T$ and $\mathscr{B}$.

The total force on a small volume dV of the medium will be $d\boldsymbol{F} = \boldsymbol{f}\, dV = \boldsymbol{j} \times \mathscr{B}\, dV$, and the total force on a finite volume is obtained by integrating this expression over all the volume. That is,

$$\boldsymbol{F} = \int_{\text{vol}} \boldsymbol{j} \times \mathscr{B}\, dV. \tag{5.4}$$

Consider the case in which there is a current along a wire or filament. A volume element dV is given by $S\, dl$ (Fig. 5-1), and therefore Eq. (5.4) gives

$$\boldsymbol{F} = \int_{\text{filament}} \boldsymbol{j} \times \mathscr{B} S\, dl.$$

Now $\boldsymbol{j} = j\boldsymbol{u}_T$ where $\boldsymbol{u}_T$ is the unit vector tangent to the axis of the filament. Then

$$\boldsymbol{F} = \int (j\boldsymbol{u}_T) \times \mathscr{B} S\, dl = \int (jS)\boldsymbol{u}_T \times \mathscr{B}\, dl. \tag{5.5}$$

However $jS = I$ where the current along the wire is the same at all points of a conductor because of the law of conservation of electric charge. Therefore Eq. (5.5) for the force on a conductor carrying an electric current becomes

$$\boldsymbol{F} = I \int \boldsymbol{u}_T \times \mathscr{B}\, dl. \tag{5.6}$$

This result may be verified by placing conductors of different shapes in a magnetic field and measuring the force on the conductor.

Consider the case of a rectilinear conductor placed in a uniform magnetic field $\mathscr{B}$ (Fig. 5-2). Then both $\boldsymbol{u}_T$ and $\mathscr{B}$ are constant, and Eq. (5.6) becomes

$$\boldsymbol{F} = I\boldsymbol{u}_T \times \mathscr{B} \int dl,$$

or if $L = \int dl$ is the length of the rectilinear conductor,

$$F = IL\boldsymbol{u}_T \times \mathscr{B}. \tag{5.7}$$

The conductor is therefore subject to a force perpendicular to itself and to the magnetic field. This is the principle on which electric motors operate. If θ is the angle between the conductor and the magnetic field, we may write for the magnitude of the force $\boldsymbol{F}$

$$F = I\mathscr{B}L \sin\theta. \tag{5.8}$$

The force is zero if the conductor is parallel to the field ($\theta = 0$) and maximum when the conductor is perpendicular to the field ($\theta = \pi/2$), a result confirmed by experiment. The direction of the force is found by applying the right-hand rule of the cross product as shown in Fig. 3-2.

5.3 Magnetic Torque on a Closed Electric Current

The torque due to the force produced by a magnetic field on a closed electric current can be computed from either Eq. (5.7) or (5.8). For simplicity consider first a current along a rectangular circuit placed so that the normal $\boldsymbol{u}_N$ to the circuit's plane (oriented by the right-hand rule in the sense of the direction of the current) makes an angle θ with the field $\mathscr{B}$, and two sides of the circuit are perpendicular to the field (Fig. 5-3). The forces $\boldsymbol{F}'$ acting on the sides L' have the same magnitude (equal to $I\mathscr{B}L'\sin\theta$) but are in opposite directions. The forces $\boldsymbol{F}'$ tend to deform the circuit but produce no torque. The forces $\boldsymbol{F}$ on the sides L are of magnitude $F = I\mathscr{B}L$, and constitute a couple whose lever arm is $L'\sin\theta$. Therefore the forces $\boldsymbol{F}$ produce on the circuit a

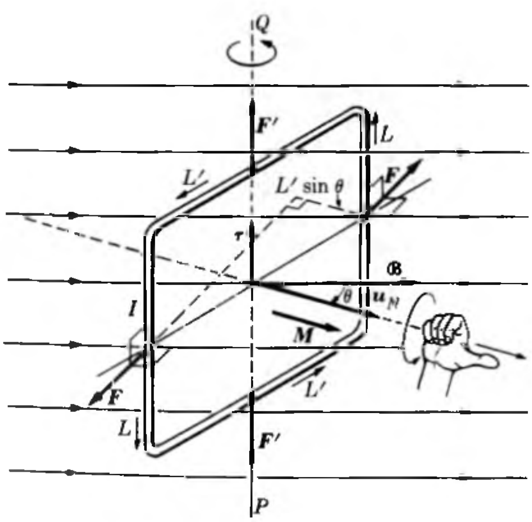


Fig. 5-3. Magnetic torque on a rectangular electric circuit placed in a magnetic field. The torque is zero when the plane of the circuit is perpendicular to the magnetic field.

torque that tends to orient the loop perpendicular to the field and whose magnitude is

$$\tau = (I\mathscr{B}L)(L' \sin\theta).$$

However $LL' = S$ where S is the area of the circuit. Thus $\tau = (IS)\mathscr{B}\sin\theta$. The direction of the torque is perpendicular to the plane of the couple; that is, along the line PQ in Fig. 5-3. Let us define a vector

$$\boldsymbol{M} = IS\boldsymbol{u}_N \tag{5.9}$$

normal to the plane of the circuit so that the torque $\boldsymbol{\tau}$ may be written as

$$\tau = M\mathscr{B}\sin\theta, \tag{5.10}$$

or in vector form

$$\boldsymbol{\tau} = \boldsymbol{M} \times \boldsymbol{\mathscr{B}}. \tag{5.11}$$

Result (5.11) is mathematically similar to Eq. (1.36), which gives the torque an external electric field produces on an electric dipole. The quantity $\boldsymbol{M}$, defined in Eq. (5.9) and equivalent to $\boldsymbol{p}$ in Eq. (1.35), is called the *magnetic dipole moment* of the current loop. Note from Eq. (5.9) that the direction of $\boldsymbol{M}$ is given by the right-hand rule shown in Fig. 5-3.

To obtain the energy of a current in a magnetic field, the logic used for the electric dipole in Section 1.10 to relate Eqs. (1.35) and (1.36) is applied in reverse; therefore the energy of the current loop placed in the magnetic field $\boldsymbol{\mathscr{B}}$ is

$$E_{\text{mag}} = -M\mathscr{B}\cos\theta = -\boldsymbol{M}\cdot\boldsymbol{\mathscr{B}}. \tag{5.12}$$

Although Eqs. (5.11) and (5.12) have been derived for a rectangular current with a special orientation in a uniform magnetic field, a more laborious mathematical discussion indicates that the equations have general validity. For example consider a *small* current loop of *any* shape whose area is S (Fig. 5-4). The magnetic dipole

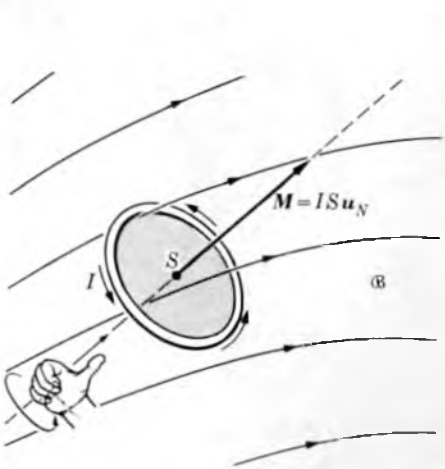


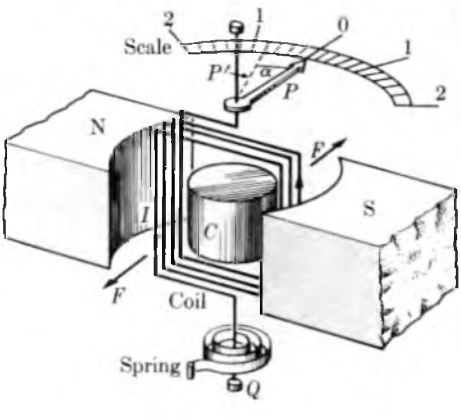
Fig. 5-4. Relation between the magnetic dipole moment of an electric current loop and the direction of the current.

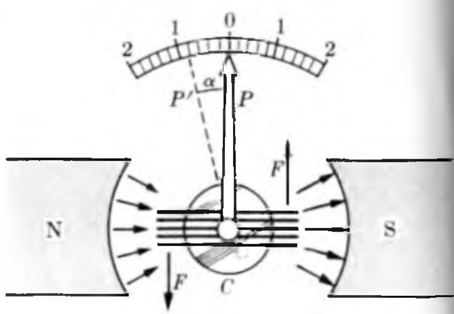
moment M of the current loop is still given by Eq. (5.9), and the torque and potential energy when the current loop is placed in a magnetic field are given by Eqs. (5.11) and (5.12).

The unit of magnetic moment from Eq. (5.12) is usually expressed as joules/tesla or $\mathrm{J\,T^{-1}}$. In terms of the fundamental units the $\mathrm{J\,T^{-1}}$ is $\mathrm{m^2\,s^{-1}\,C}$, in agreement with the definition in Eq. (5.9).

Example 5.1. A current measuring device: the *galvanometer*.

▼ A simple design of a galvanometer is illustrated in Fig. 5-5. The current to be measured passes through a coil suspended between the poles of a magnet. (In some cases the coil is wrapped around an iron cylinder C.) The magnetic field exerts a torque on the coil and rotates it by a certain angle. The angle can then be related to the current passing through the rectangular loop. Let S be the area of the coil. The torque produced by the magnetic field tends to place the coil perpendicular to the field and twist the spring Q. The coil adopts an equilibrium position, rotated an angle α, when the magnetic torque is balanced by the elastic torque $k\alpha$ produced by the spring where k is the spring's elastic constant. The angle α is indicated by a pointer attached to the coil. The pole faces are shaped as indicated in the figure so that the magnetic field between the pole faces and the iron cylinder C is radial as shown in the top view in Fig. 5-5. In this case $\mathscr{B}$ is always in the plane of the circuit, and θ in Eq. (5.10) is $\pi/2$ so that $\sin\theta = 1$. Then the torque is given by $\tau = IS\mathscr{B}$ since $M = IS$. At equilibrium when the torque produced by the magnetic field is balanced by the torque produced by the twisting of the spring, $IS\mathscr{B} = k\alpha$, and therefore $I = k\alpha/S\mathscr{B}$. If k, S, and $\mathscr{B}$ are known, this equation gives the value of the current I in terms of the angle α. Usually the scale is calibrated so that the value of I can be read directly in some convenient units. ▲





(b)

Fig. 5-5. (a) Basic components of a moving coil galvanometer. (b) Top view of galvanometer shown in (a).

5.4 Magnetic Field Produced by a Closed Current Loop

The presence of a magnetic field is recognized from the force the field produces on a moving charge. In addition as already seen in Section 4.5, moving charges produce a magnetic field. The English scientist Michael Faraday (1791–1867) became intrigued with magnetism after reading Oersted's paper in 1820; after some experimentation Faraday came to the conclusion that the magnetic lines set up by the current are closed and encircle the current. This experiment may be done by placing a small compass needle in different positions close to the current. The needle tends to line up perpendicular to both the direction of the current and the radial direction from the wire to the compass position.

After many experiments over a period of years a general expression was obtained for calculating the magnetic field produced by a *closed* current loop of any shape. This expression, called the *Ampère-Laplace law*, is

$$\mathcal{B} = K_m I \oint \frac{\boldsymbol{u}_T \times \boldsymbol{u}_r}{r^2}\, dl \tag{5.13}$$

where the meaning of all symbols is indicated in Fig. 5-6, the integral is extended along the entire closed circuit (therefore the symbol $\oint$ is used), and K_m is a constant whose value in the SI is 10^{-7} T m/A or m kg C^{-2} as was indicated in Section 4.5. [Note that the integral in Eq. (5.13) is expressed in m^{-1} when r and l are given in meters.] Therefore

$$\mathcal{B} = 10^{-7}\, I \oint \frac{\boldsymbol{u}_T \times \boldsymbol{u}_r}{r^2}\, dl. \tag{5.14}$$

It was also mentioned in Section 4.5 that it is customary to write $K_m = \mu_0/4\pi$ where μ_0 is the *magnetic permeability of vacuum*. Thus Eq. (5.13) for the Ampère-Laplace law becomes

$$\mathcal{B} = \frac{\mu_0}{4\pi}\, I \oint \frac{\boldsymbol{u}_T \times \boldsymbol{u}_r}{r^2}\, dl. \tag{5.15}$$

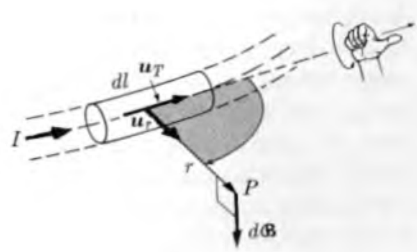


Fig. 5-6. Description for defining terms in Ampère-Laplace law. The small segment dl of the current loop contributes to the magnetic field at point P. This contribution is perpendicular to the plane defined by $\boldsymbol{u}_T$ and $\boldsymbol{u}_R$.

Example 5.2. Magnetic field of a moving charge.

▼ In Section 4.5 it was stated that a moving charge produces a magnetic field. The expression for the magnetic field of a moving charge will now be shown to be compatible with the result for the magnetic field of an electric current. The magnetic field of an electric current as given by Eq. (5.15) may be written as

$$\mathscr{B}=\frac{\mu_0}{4\pi} I \oint \frac{\boldsymbol{u}_T \times \boldsymbol{u}_r}{r^2} dl=\frac{\mu_0}{4\pi} \oint \frac{(I\, dl\, \boldsymbol{u}_T) \times \boldsymbol{u}_r}{r^2}.$$

Recall Eqs. (5.1) and (5.2), and that $dV=S dl$, as well as that $\boldsymbol{j}=j\,\boldsymbol{u}_T=nq\boldsymbol{v}$; then the term in parentheses above may be written

$$I\, dl\, \boldsymbol{u}_T=(jS)\, dl\, \boldsymbol{u}_T=\boldsymbol{j}\, dV=nq\boldsymbol{v}\, dV.$$

Therefore the magnetic field of a moving charge is

$$\mathscr{B}=\frac{\mu_0}{4\pi} \oint \frac{q\boldsymbol{v} \times \boldsymbol{u}_r}{r^2} n\, dV. \tag{5.16}$$

Since $n dV$ is the number of particles in the volume dV, we *may interpret* the result above by saying that each charged particle produces at a distance r a magnetic field given by

$$\mathscr{B}=\frac{\mu_0}{4\pi} \frac{q\boldsymbol{v} \times \boldsymbol{u}_r}{r^2}, \tag{5.17}$$

which is Eq. (4.14) in Section 4.5 for the (nonrelativistic) magnetic field of a moving charge.

Recall that Eq. (4.28) is the expression for the magnetic field of a moving charge and is valid at all velocities of the charge. The question may now arise as to whether Eq. (4.28) is compatible also with Eq. (5.15) for the field of a closed current. A detailed calculation, here omitted, shows that Eq. (5.15) does remain valid, independent of the velocities of the charges composing the closed current. ▲

5.5 Magnetic Field of a Rectilinear Current

As an example of the Ampère-Laplace law, consider a very long and thin rectilinear current as in Fig. 5-7. For any point P and any element dl of the current, the vector $\boldsymbol{u}_T \times \boldsymbol{u}_r$ is perpendicular to the plane determined by P and the current, and therefore the vector's direction is that of the unit vector $\boldsymbol{u}_\theta$. At P the magnetic field produced by dl is then tangent to the circle of radius R centered on the current element that passes through P, and in a plane perpendicular to the current. Therefore when the integration in Eq. (5.15) is performed, the contributions from all terms in the integral have the same direction $\boldsymbol{u}_\theta$; and the resultant magnetic field $\mathscr{B}$ is also tangent to the circle. Thus it is necessary to find only the magnitude of $\mathscr{B}$. The magnitude of $\boldsymbol{u}_T \times \boldsymbol{u}_r$ is $\sin\theta$ since $\boldsymbol{u}_T$ and $\boldsymbol{u}_r$ are unit vectors. Therefore for a rectilinear current Eq. (5.15) is in magnitude

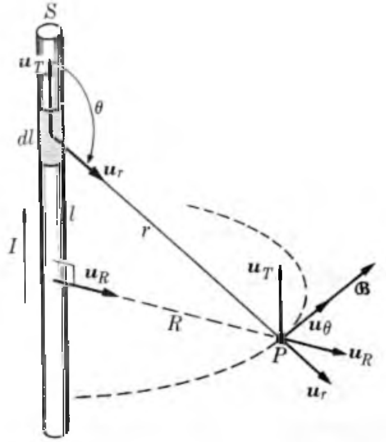


Fig. 5-7. Magnetic field produced by a rectilinear current at point P.

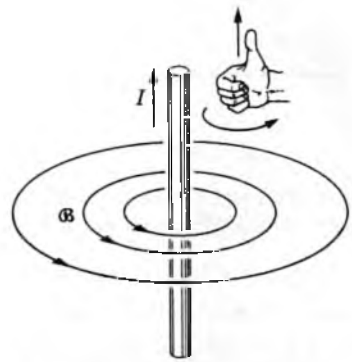


Fig. 5-8. Magnetic lines of force about a rectilinear current.

$$\mathscr{B}=\frac{\mu_0}{4\pi} I\int_{-\infty}^{\infty} \frac{\sin\theta}{r^2}\, dl. \tag{5.18}$$

From Fig. 5-7, $r=R\csc\theta$ and $l=R\cot(180° - \theta)= -R\cot\theta$ so that $dl=R\csc^2\theta\, d\theta$. Substitution in Eq. (5.18) yields

$$\mathscr{B}=\frac{\mu_0}{4\pi} I\int_0^{\pi} \frac{\sin\theta}{R^2\csc^2\theta}(R\csc^2\theta\, d\theta)=\frac{\mu_0 I}{4\pi R}\int_0^{\pi}\sin\theta\, d\theta$$

where $l=-\infty$ corresponds to $\theta=0$ and $l=+\infty$ to $\theta=\pi$. Then

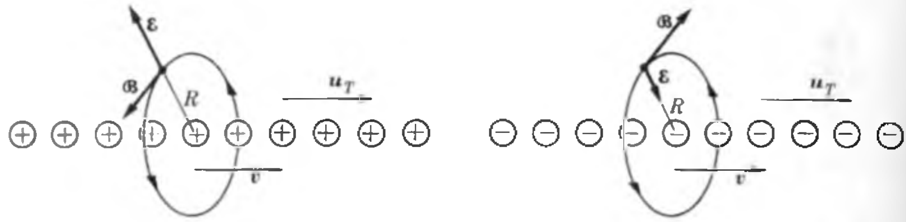
$$\mathscr{B}=\frac{\mu_0 I}{2\pi R}. \tag{5.19}$$

or in vector form

$$\mathscr{B}=\frac{\mu_0 I}{2\pi R}u_\theta. \tag{5.20}$$

The magnetic field of an infinitely long current-carrying conductor is inversely proportional to the distance R, and the lines of force are circles concentric with the current and perpendicular to it as shown in Fig. 5-8. The right-hand rule for determining the direction of the magnetic field relative to the direction of the current is also indicated in the figure. This result is sometimes called the *Biot-Savart formula* after the French experimenters Jean Biot (1774–1862) and Felix Savart (1791–1841) who discovered the relation given by Eq. (5.20).

In the case of a rectilinear current in a conductor, there is a magnetic field $\mathscr{B}$ but no electric field $\mathscr{E}$ because in addition to the moving electrons there are the metal's fixed positive ions that produce an electric field equal and opposite to that of the



(a) q positive (b) q negative

Fig. 5-9. Relation between the electric and magnetic fields produced by a stream of positive (negative) ions moving in a straight line.

electrons. Therefore the net electric field is zero. However for ions moving along the axis of a linear accelerator, there are both a magnetic field and an electric field. The electric field corresponds to the value given in Example 1.5 for the electric field of a charged filament, $\boldsymbol{\mathscr{E}}=\lambda \boldsymbol{u}_r/2\pi\epsilon_0 R$ (Fig. 5-9). Therefore comparing this value with Eq. (5.20) gives the relation of the two fields:

$$\mathscr{B}=\frac{\mu_0\epsilon_0 I}{\lambda}\boldsymbol{u}_T\times\boldsymbol{\mathscr{E}}. \tag{5.21}$$

Example 5.3. Verification that result (5.21) for the magnetic field of a rectilinear current is compatible with Eq. (4.17),

$$\mathscr{B}=\frac{1}{c^2}\boldsymbol{v}\times\boldsymbol{\mathscr{E}}.$$

▼ The magnetic field produced by a rectilinear current is the result of the individual fields produced by all the charges moving along the conductor. According to Eq. (5.2) if S is the cross section of the conductor, $I=nqSv$ where v is the velocity of the charges. Because nq is the charge per unit volume, the charge of a conductor having unit length and cross section S is $nqS=\lambda$. Therefore $I=\lambda v$. Making the substitutions in Eq. (5.21) and noting that $\boldsymbol{v}=v\boldsymbol{u}_T$ give

$$\mathscr{B}=\frac{\mu_0\epsilon_0(\lambda v)}{\lambda}\boldsymbol{u}_T\times\boldsymbol{\mathscr{E}}=\mu_0\epsilon_0\boldsymbol{v}\times\boldsymbol{\mathscr{E}}=\frac{1}{c^2}\boldsymbol{v}\times\boldsymbol{\mathscr{E}},$$

which is just Eq. (4.17). ▲

Example 5.4. Derivation of the magnetic field of a rectilinear current by means of the relativistic transformation for the electromagnetic field.

▼ Although the derivation of Eq. (5.21) for the magnetic field of a rectilinear current used the Ampère–Laplace law, the result can be obtained using the theory of relativity. Consider an infinite row of equally spaced charges moving along the X-axis with velocity $\boldsymbol{v}$ relative to observer O (Fig. 5-10). These moving charges constitute a rectilinear electric current. If λ is the electric charge per unit length, the electric current measured by O is $I=\lambda v$. Now consider an observer

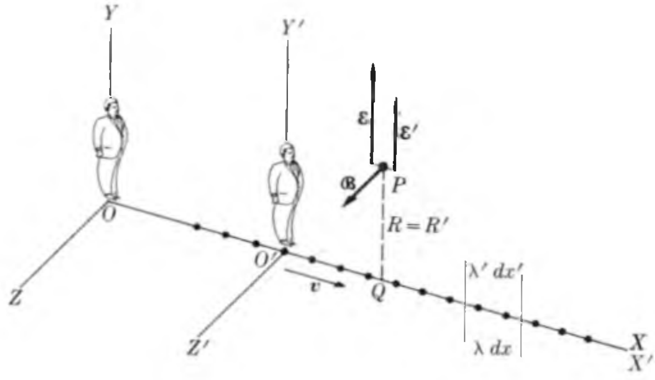


Fig. 5-10. Electromagnetic field produced by a stream of charges moving along the X-axis as observed by two observers in relative motion.

O' moving in the X-direction with velocity v. Relative to O', the charges appear at rest; and O' measures only an electric field. However, O records an electric *and* a magnetic field.

As measured by O, the charge in a segment dx is $dq=\lambda\,dx$. Observer O' measures the same charge; but because of the Lorentz contraction, the segment appears to have a length dx' such that $dx=\sqrt{1-v^2/c^2}\,dx'$. Therefore O' measures a different charge per unit length λ', such that $\lambda'\,dx=\lambda\,dx$ and therefore

$$\lambda'=\lambda\frac{dx}{dx'}=\sqrt{1-v^2/c^2}\,\lambda.$$

The electric field as measured by O' is transverse and at a point P is given by the result of Example 1.5; that is, $\mathscr{E}'=\lambda'/2\pi\epsilon_0 R'$. By placing the Y-axes parallel to the line PQ and noting that $R=R'$ because it is a transverse length, we may write

$$\mathscr{E}'_x=0, \qquad \mathscr{E}'_y=\frac{\lambda'}{2\pi\epsilon_0 R}, \qquad \mathscr{E}'_z=0$$

Then from Eqs. (4.22) with $\mathscr{B}'=0$, the components of the electric field relative to O are

$$\mathscr{E}_x=\mathscr{E}_z=0, \qquad \mathscr{E}_y=\frac{\mathscr{E}'_y}{\sqrt{1-v^2/c^2}}=\frac{\lambda'}{2\pi\epsilon_0 R\sqrt{1-v^2/c^2}}=\frac{\lambda}{2\pi\epsilon_0 R}$$

Similarly Eqs. (4.24) give the components of the magnetic field relative to O as

$$\mathscr{B}_x=\mathscr{B}_y=0, \qquad \mathscr{B}_z=\frac{v\mathscr{E}'_y/c^2}{\sqrt{1-v^2/c^2}}=\frac{\lambda' v/c^2}{2\pi\epsilon_0 R\sqrt{1-v^2/c^2}}=\frac{\mu_0 I}{2\pi R}$$

where the relation $\epsilon_0\mu_0=1/c^2$ has been used. Thus are found not only the correct electric field in frame XYZ for a rectilinear charge distribution, but also the correct expression for the magnetic field produced by a rectilinear current with Eq. (5.15) as the starting point. Hence we may feel confident that the Ampère–Laplace law (5.15) is compatible with the requirements of the principle of relativity, and therefore gives the correct magnetic field associated with a *closed* electric circuit. ▲

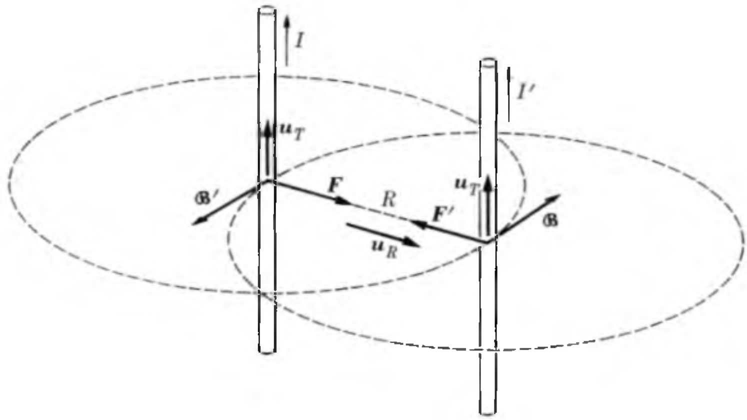


Fig. 5-11. Magnetic interaction between two rectilinear currents.

5.6 Forces between Currents

Equation (5.20) for the magnetic field of an infinitely long current-carrying conductor will now be combined with Eq. (5.2) for the magnetic force on a current to obtain the magnetic interaction between two electric currents. For simplicity consider two parallel currents I and I' (Fig. 5-11) in the same direction and separated by the distance R. The magnetic field $\mathcal{B}$ produced by I at any point of I' is given by Eq. (5.20), and has the direction indicated. The force F' on I' will be from Eq. (5.6)

$$F' = I' \int u'_T \times \mathcal{B} \, dl'.$$

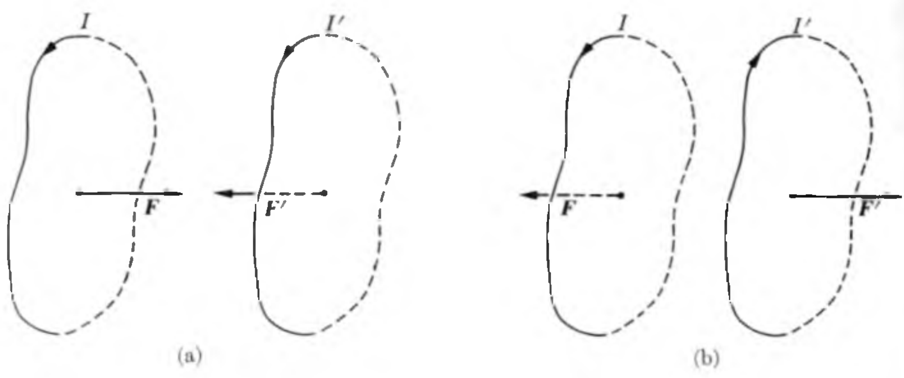


Fig. 5-12. Attraction and repulsion between two current loops.

Now $\boldsymbol{u}'_T \times \mathscr{B} = -\boldsymbol{u}_R \mathscr{B}$ where $\boldsymbol{u}_R$ is defined as the unit vector from I to I'. Therefore, from Eq. (5.20) for $\mathscr{B}$ the force is

$$F = I' \int\left(-\boldsymbol{u}_R \frac{\mu_0 I}{2\pi R}\right) dl' = -\boldsymbol{u}_R \left(\frac{\mu_0 I I'}{2\pi R}\right) \int dl' = -\boldsymbol{u}_R \frac{\mu_0 I I'}{2\pi R} L'. \tag{5.22}$$

This result indicates that current I *attracts* current I'. A similar calculation of the force on I produced by I' gives the same result but with a plus sign so that the force has the same direction as $\boldsymbol{u}_R$ and again represents an attraction. Therefore *two parallel currents in the same direction attract each other with equal forces* as a result of their magnetic interaction. The student should verify that *if parallel currents are in opposite directions, the currents repel each other.*

This result can be extended to currents of any configuration. The circuits of Fig. 5-12(a) attract each other, but those in Fig. 5-12(b) repel each other. Interactions between currents and circuits have great practical importance for electric motors as well as other engineering applications.

5.7 Note on SI Units

For a fourth basic unit to add to those of length, mass, and time, there are two laws from which to choose: Coulomb's law for the electrostatic interaction between two charges is given by Eq. (1.2) as

$$F = K_e \frac{qq'}{r^2},$$

and the law of interaction between two rectilinear currents is given by Eq. (5.22) with $\mu_0/4\pi$ replaced by the magnetic constant K_m as

$$F = K_m \frac{2II'}{R} L'.$$

Although two constants, K_e and K_m, correspond to the electric and magnetic forces, there is only one degree of freedom because only one new physical quantity, the electric charge, has been introduced. The current is related to the charge by the equation *current = charge/time*. Therefore an arbitrary value can be assigned to only one of the constants. When the Eleventh General Conference on Weights and Measures, held in 1960, established the SI, $K_m \equiv 10^{-7}$ was adopted, and the ampere rather than the coulomb was chosen as the fundamental unit. Because of this arbitrary adoption, the ampere is defined as the current that, circulating in two parallel conductors separated a distance of one meter, results in a force on each conductor of 2×10^{-7} N per meter of length of each conductor (Fig. 5-13). Once the ampere is so

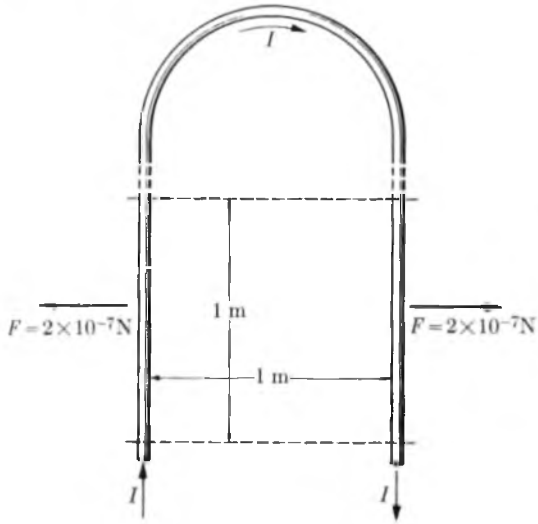


Fig. 5-13. Apparatus for defining the ampere experimentally.

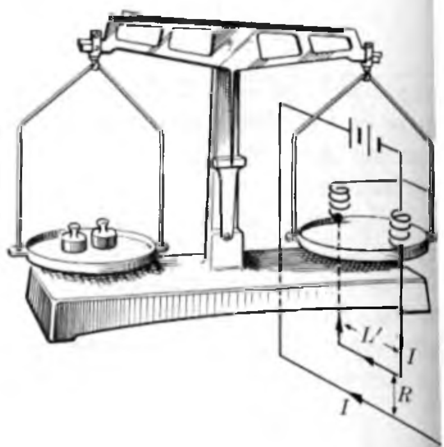


Fig. 5-14. A current balance for measuring a current in terms of the magnetic force between two parallel current-carrying conductors.

defined, the coulomb is defined as the quantity of charge that flows across any cross section of a conductor in one second when the current is one ampere.

A *current balance* is an experimental arrangement for measuring the force between two parallel conductors (Fig. 5-14). The same current passes through the two conductors so that $F=2\times10^{-7}\ I^2L'/R$. The balance is first set in equilibrium with no current in the circuit. When the current is sent through the circuit, additional weights are required on the left pan to bring the balance back to equilibrium. From the known values of F, L', and R, the value of I can be calculated. In practice, two parallel circular coils are used.

Since in terms of the auxiliary constants ϵ_0 and μ_0 we have $K_e=1/4\pi\epsilon_0$ and $K_m=\mu_0/4\pi$, it follows that the ratio of these two constants yields

$$\frac{K_e}{K_m}=\frac{1}{\epsilon_0\mu_0}=c^2$$

where $c=1/\sqrt{\epsilon_0\mu_0}$. This constant is equal to the velocity of light (or of any electromagnetic signal) in vacuum as will be proved in Chapter 11. The constant c has been measured experimentally with very great accuracy. In terms of the velocity of light, $K_e=K_mc^2=10^{-7}\ c^2$. Choosing this value for K_e in Section 1.3 may have appeared somewhat arbitrary at the time but is now explained.

One reason why the Eleventh Conference recommended the use of the ampere as the fourth fundamental unit is that it is easier to prepare a standard of current and to measure the force between two currents than to set up a standard of charge and measure the force between two charges. However from the physical point of view, the concept of charge is more fundamental than that of current. Also from the practical as well as from the theoretical point of view the use of either the coulomb or the ampere as a fundamental unit may be considered as alternate expressions of the SI unit.

5.8 Magnetic Field of a Circular Current Loop

Consider a circular current loop of radius a (Fig. 5-15). Using the Ampère-Laplace law to calculate the magnetic field at an arbitrary point is a somewhat complicated mathematical problem; but at points along the axis of the circle, the computation is a fairly easy task. First recognize that Eq. (5.15) can be interpreted mathematically as saying that at P the resultant magnetic field $\mathscr{B}$ produced by the current is the sum of a large number of very small or elementary contributions $d\mathscr{B}$ by each of the segments or length elements dl composing the circuit. Each elementary contribution is

$$d\mathscr{B}=\frac{\mu_0}{4\pi}I\frac{\boldsymbol{u}_T\times\boldsymbol{u}_r}{r^2}\,dl.$$

However, this equation must be considered only in relation to Eq. (5.15) and not as an independent statement.

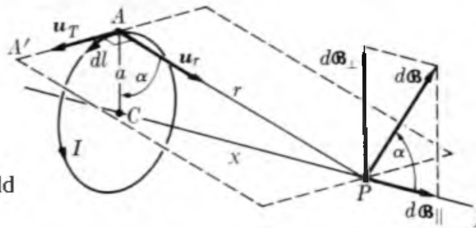


Fig. 5-15. Computation of the magnetic field along the axis of a circular current loop.

In the case of a circular current loop, the vector product $\boldsymbol{u}_T\times\boldsymbol{u}_r$ of Fig. 5-15 is perpendicular to the plane PAA' and has unit magnitude because these two unit vectors are perpendicular. Therefore the field $d\mathscr{B}$ produced by the length element dl at P has the magnitude

$$d\mathscr{B}=\frac{\mu_0}{4\pi}I\frac{dl}{r^2}$$

and is perpendicular to the plane PAA'. However the field is oblique to the X-axis. Decomposing $d\mathscr{B}$ into a component $d\mathscr{B}_{\parallel}$ parallel to the axis and a component $d\mathscr{B}_{\perp}$ perpendicular to it, we see that when we integrate along the circle, for each $d\mathscr{B}_{\perp}$ there is another in the opposite direction from the length element directly opposed

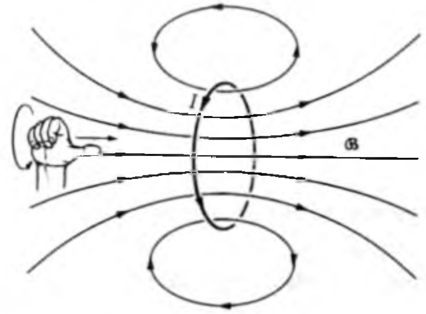


Fig. 5-16. Magnetic lines of force produced by a circular current loop.

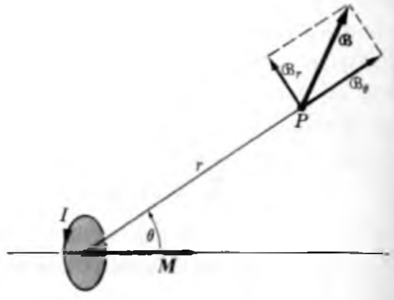


Fig. 5-17. Magnetic field at the point P produced by a magnetic dipole current.

to dl, and therefore all vectors $d\mathscr{B}_\perp$ add to zero. The resultant $\mathscr{B}$ will be the sum of all the $d\mathscr{B}_{||}$, and therefore is parallel to the axis. Now since $\cos\alpha = a/r$,

$$d\mathscr{B}_{||} = (d\mathscr{B})\cos\alpha = \frac{a}{r}\,d\mathscr{B} = \frac{\mu_0 I a}{4\pi r^3}\,dl.$$

The distance r remains constant for the integration around the circle. Then since $\oint dl = 2\pi a$, the magnitude of the resultant magnetic field is given by

$$\mathscr{B} = \oint d\mathscr{B}_{||} = \frac{\mu_0 I a}{4\pi r^3}\oint dl = \frac{\mu_0 I a^2}{2r^3}.$$

Because $r = (a^2 + x^2)^{1/2}$, the magnetic field for points on the axis of a circular current loop is

$$\mathscr{B} = \frac{\mu_0 I a^2}{2(a^2 + x^2)^{3/2}}. \tag{5.23}$$

From the definition (5.9) the magnetic dipole moment of the circuit is $M = I(\pi a^2)$. Then

$$\mathscr{B} = \frac{\mu_0 M}{2\pi(a^2 + x^2)^{3/2}}. \tag{5.24}$$

The magnetic field of a circular current loop has been represented in Fig. 5-16.

An interesting case occurs when the circuit is very small so that the radius a can be neglected in comparison with the distance x. Then Eq. (5.24) reduces to

$$\mathscr{B} = \frac{\mu_0 M}{2\pi x^3} = \frac{\mu_0 (2M)}{4\pi x^3}. \tag{5.25}$$

When Eq. (5.25) is compared with Eq. (1.32) with $\theta = 0$, that is, $\mathscr{E}_r = (1/4\pi\epsilon_0)(2p/r^3)$, it is seen that the magnetic field along the axis of the small current is identical to the electric field along the axis of an electric dipole if $(\mu_0/4\pi)M$ is made to correspond to $p/4\pi\epsilon_0$. For that reason the circuit is called a *magnetic dipole*. Therefore Eqs. (1.32) and (1.33) for an electric dipole can apply to a magnetic dipole so that the magnetic field off-axis may be computed (Fig. 5-17). This application gives

$$\mathscr{B}_r = \frac{\mu_0}{4\pi}\frac{2M\cos\theta}{r^3}, \qquad \mathscr{B}_\theta = \frac{\mu_0}{4\pi}\frac{M\sin\theta}{r^3}. \tag{5.26}$$

In Chapter 1 the lines of force of an electric field were seen to go from the negative to the positive charges or perhaps from or to infinity in some cases. However Figs. 5-8 and 5-16 showed that the lines of force of a magnetic field are *closed* lines, linked about the current. The reason is that the magnetic field does not originate with magnetic poles. This kind of field, which does not have point sources, is called *solenoidal*.

Example 5.5. The *tangent galvanometer.*

A tangent galvanometer consists of a circular coil (Fig. 5-18) having N turns and carrying a current I. The galvanometer is placed in a region in which there is a magnetic field $\mathscr{B}$ so that the plane of the coil is parallel to $\mathscr{B}$. At the center of the coil the current I produces a magnetic field given by $M_0I/2a$ [from Eq. (5.23) with $x=0$]. Because there are N turns, the total magnetic field produced at the center is $\mathscr{B}_c=\mu_0 IN/2a$. Therefore the resultant magnetic field $\mathscr{B}'$ at the center of the coil makes an angle θ with the axis of the coil; angle θ is given by

$$\tan\theta = \frac{\mathscr{B}}{\mathscr{B}_c} = \frac{2a\mathscr{B}}{\mu_0 IN}.$$

A small magnetic needle placed at the center of the coil will turn and rest in equilibrium at an angle θ with the axis. Thus the external field $\mathscr{B}$ can be calculated if the current I is known; conversely the current I can be measured if the field $\mathscr{B}$ is known. Usually $\mathscr{B}$ is the earth's magnetic field. For precise measurements the formula has to be corrected to take into account the finite length of the needle since the field acting on the needle is not exactly the field at the center of the coil. The name "tangent galvanometer" is derived from the trigonometric function appearing above. ▲

Example 5.6. The magnetic field of a solenoidal circuit.

▼ A solenoidal circuit, or simply a *solenoid*, is a circuit composed of several coaxial circular loops of the same radius, which all carry the same current (Fig. 5-19). The magnetic field of a solenoidal

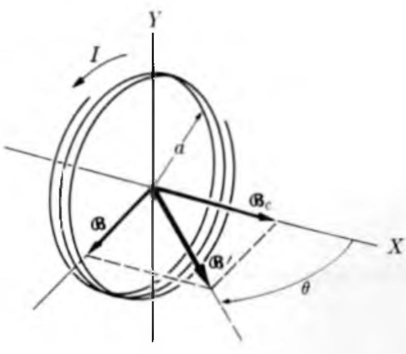


Fig. 5-18. Tangent galvanometer.

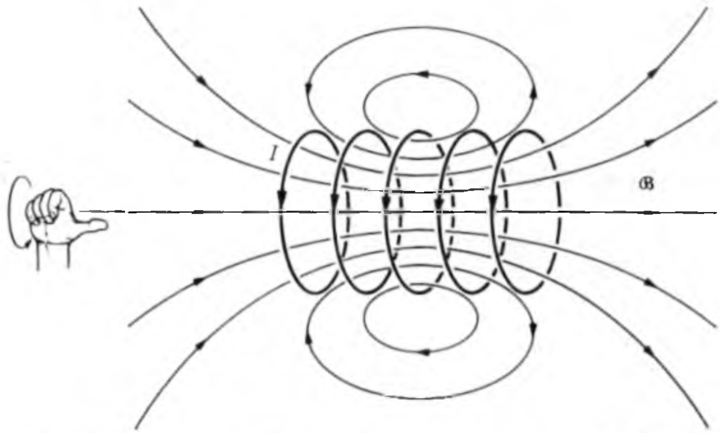


Fig. 5-19. Magnetic lines of force produced by a solenoidal circuit.

circuit is found by adding the magnetic fields of each of the component circular currents. The field is indicated by lines of magnetic force in the figure, in which some fluctuations in the space between loops have been smoothed out. The field of the solenoid will be computed only at points on the axis.

Fig. 5-20 shows a longitudinal cross section of the solenoid. If L is the length and N the number of loops, the number of loops per unit length is N/L, and the number of loops in a section of length dR is $(N/L)\,dR$. The field produced by each loop at a point P on the axis is calculated by using Eq. (5.23), and the field produced by the loops in the section dR can be computed in the following way:

$$d\mathscr{B}=\left[\frac{\mu_0 I a^2}{2(a^2+R^2)^{3/2}}\right]\frac{N}{L}\,dR=\frac{\mu_0 IN}{2L}\,\frac{a^2\,dR}{(a^2+R^2)^{3/2}}. \tag{5.27}$$

From Fig. 5-20 $R=a\cot\beta$, $dR=-a\csc^2\beta\,d\beta$, and $a^2+R^2=a^2\csc^2\beta$. Substitution in Eq. (5.27) yields

$$d\mathscr{B}=\frac{\mu_0 IN}{2L}(-\sin\beta\,d\beta)$$

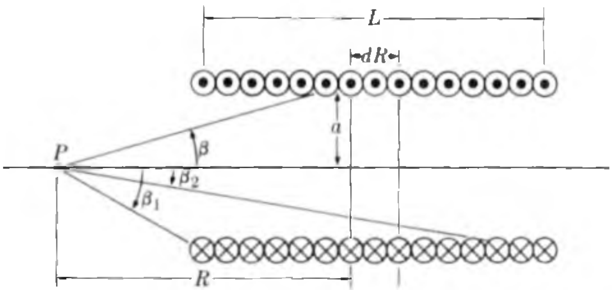


Fig. 5-20. Computation of the magnetic field at a point P located along the axis of a solenoidal circuit.

The resultant field is found by integrating from one end of the solenoid to the other. That is, the resultant field is

$$\mathscr{B}=\frac{\mu_0 IN}{2L}\int_{\beta_1}^{\beta_2} -\sin\beta\, d\beta=\frac{\mu_0 IN}{2L}(\cos\beta_2-\cos\beta_1). \tag{5.28}$$

If the solenoid is very long, a point at the center such that $\beta_1 \approx \pi$ and $\beta_2 \approx 0$ results in

$$\mathscr{B}=\frac{\mu_0 IN}{L}. \tag{5.29}$$

For a point at one end, $\beta_1=\pi/2$ and $\beta_2 \approx 0$, or $\beta_1 \approx \pi$, $\beta_2=\pi/2$. In either case,

$$\mathscr{B}=\frac{\mu_0 IN}{2L}, \tag{5.30}$$

or one-half the value at the center. A long solenoid is used to produce fairly uniform magnetic fields in limited regions around the center. ▲

Example 5.7. The magnetic field of a *magnetic quadrupole*.

▼ The system of Fig. 5-21 is composed of two small identical circuits, carrying equal currents I, but circulating in opposite senses and separated by the distance $2a$. Each circuit is thus a magnetic dipole; but because the currents circulate in opposite senses, the dipole moments are opposed and give a total magnetic dipole moment of zero. However, the resultant magnetic field is not zero, because of the separation of the circuits, and thus the system constitutes a magnetic quadrupole. Note that mathematically the situation is very similar to that of Example 1.8.

Because of the analogy between Eq. (5.26) for a magnetic dipole and Eqs. (1.32) and (1.33) for an electric dipole, we may define a "magnetic" potential V_m associated with the magnetic field of a magnetic dipole given by Eq. (1.31), with $p/4\pi\epsilon_0$ replaced by $\mu_0 M/4\pi$. Therefore

$$V_m=\frac{\mu_0 M\cos\theta}{4\pi r^2}=\frac{\mu_0 \boldsymbol{M}\cdot\boldsymbol{r}}{4\pi r^3}.$$

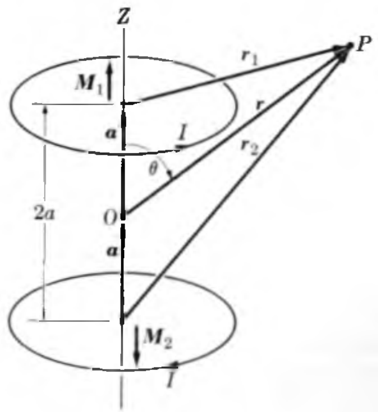


Fig. 5-21. Magnetic quadrupole.

Thus since $M_1=-M_2=M$, the resultant "magnetic" potential at P is

$$V_m=\frac{\mu_0 M_1\cdot r_1}{4\pi r_1^3}+\frac{\mu_0 M_2\cdot r_2}{4\pi r_2^3}=\frac{\mu_0 M}{4\pi}\cdot\left(\frac{r_1}{r_1^3}-\frac{r_2}{r_2^3}\right).$$

From Fig. 5-21, calling $a=u_z a$ where u_z is the unit vector along the Z-axis, we have $r_1=-a+r$, $r_2=a+r$, and also

$$r_1^2=r^2+a^2-2ar\cos\theta,$$

$$r_2^2=r^2+a^2+2ar\cos\theta.$$

Therefore using the binomial expansion up to the first order in a/r gives

$$\frac{1}{r_1^3}=\frac{1}{r^3}\left(1-\frac{2a\cos\theta}{r}+\frac{a^2}{r^2}\right)^{-3/2}=\frac{1}{r^3}\left(1+\frac{3a\cos\theta}{r}+\cdots\right).$$

and similarly,

$$\frac{1}{r_2^3}=\frac{1}{r^3}\left(1-\frac{3a\cos\theta}{r}+\cdots\right).$$

Therefore

$$\frac{r_1}{r_1^3}-\frac{r_2}{r_2^3}=\frac{-a+r}{r^3}\left(1+\frac{3a\cos\theta}{r}+\cdots\right)-\frac{a+r}{r^3}\left(1-\frac{3a\cos\theta}{r}+\cdots\right)=\frac{-2a}{r^3}+\frac{6ra\cos\theta}{r^4}+\cdots.$$

Substituting this value in the expression for V_m gives

$$V_m=\frac{2\mu_0}{4\pi r^3}\left(-M\cdot a+\frac{3M\cdot ra\cos\theta}{r}\right).$$

However $M\cdot a=Ma$ and $M\cdot r=Mr\cos\theta$. Thus

$$V_m=\frac{\mu_0 M(2a)(3\cos^2\theta-1)}{4\pi r^3},$$

which is similar to Eq. (1.44) in angular and radial dependence and confirms the fact that we are dealing with a magnetic quadrupole. The moment of the magnetic quadrupole is $M(2a)$. The magnetic field of the magnetic quadrupole has radial and transverse components given by

$$\mathscr{B}_r=-\frac{\partial V_m}{\partial r}=\frac{3\mu_0 M(2a)(3\cos^2\theta-1)}{4\pi r^4},$$

$$\mathscr{B}_\theta=-\frac{1}{r}\frac{\partial V_m}{\partial\theta}=\frac{6\mu_0 M(2a)\sin\theta\cos\theta}{4\pi r^4}.$$

The student is warned that the "magnetic" potential introduced for mathematical convenience for computing the magnetic field is not related to a magnetic potential energy in the same way as the electric potential is related to an electric potential energy. ▲

Problems

5.1 Find the current density (assumed uniform) required in a horizontal aluminum wire to make it "float" in the earth's magnetic field at the equator. The density of Al is 2.7×10^3 kg m^{-3}. Assume that the earth's field is about 7×10^{-5} T and that the wire is oriented in the east-west direction.

5.2 Find the force on each of the wire segments shown in Fig. 5-22 if the field, $\mathscr{B} = 1.5$ T, is parallel to OZ and $I = 2$ A. An edge of the cube is 0.1 m.

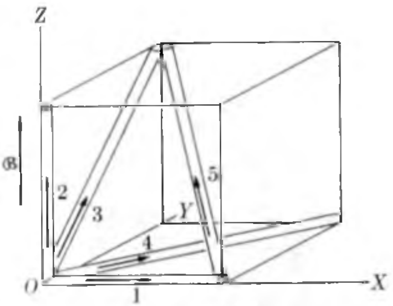


Figure 5-22

5.3 The plane of a rectangular loop of wire .05 m × .08 m is parallel to a magnetic field of 0.15 T. (a) If the loop carries a current of 10 A, what torque acts on the loop? (b) What is the magnetic moment of the loop? (c) What is the maximum torque that can be obtained with the same total length of wire carrying the same current in this magnetic field?

5.4 The rectangular loop in Fig. 5-23 is pivoted about the Y-axis and carries a current of 10 A in the direction indicated. If the loop is in a uniform magnetic field of 0.2 T parallel to the X-axis, calculate (a) the force on each side of the loop, in N, and (b) the torque in N m, required to hold the loop in the position shown. With the loop in a field of 0.2 T parallel to the Z-axis calculate (c) the force on each side of the loop in N and (d) the torque in N m required to hold the loop in that position. (e) What torque would be required if the loop were pivoted about an axis through the loop's center, parallel to the Y-axis?

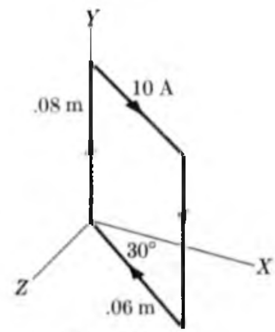


Figure 5-23

5.5 The rectangular loop of wire in Fig. 5-24 has a mass of 0.01 kg per meter of length, and is pivoted about side AB as a frictionless axis. The current in the wire is 10 A in the direction shown. (a) Calculate the magnitude and the sense of the magnetic field, parallel to the Y-axis, that will cause the loop to swing up until its plane makes an angle of 30° with the YZ-plane. (b) Discuss the case in which the field is parallel to the X-axis.

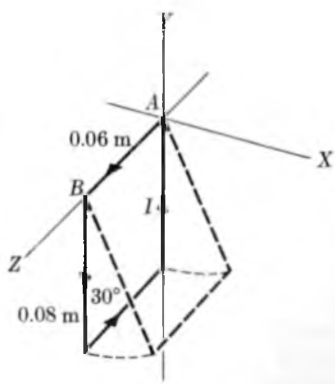


Figure 5-24

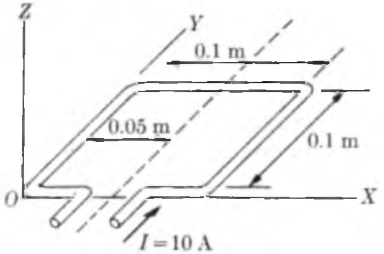


Figure 5-25

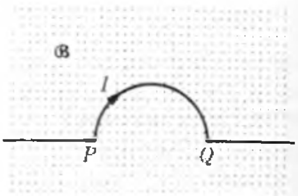


Figure 5-26

5.6 What is the maximum torque on a coil .05 m $\times$ .12 m. composed of 600 turns, when it is carrying a current of 10^{-5} A in a uniform field 0.10 T?

5.7 The coil of a pivoted-coil galvanometer has 50 turns and encloses an area of 6×10^{-4} m^2. The magnetic field in the region in which the coil swings is 10^{-2} T, and is radial. The torsional constant of the hairsprings is $k = 10^{-6}$ N m/deg. Find the angular deflection of the coil for a current of 1 mA.

5.8 A wire loop in the form of a square of side 0.1 m lies in the XY-plane as shown in Fig. 5-25. There is a current of 10 A in the loop as shown. If a magnetic field parallel to the Z-axis and having an intensity $\mathscr{B} = 0.1x$ T (where x is in m) is applied, calculate (a) the resultant force on the loop, and (b) the resultant torque relative to O.

5.9 Repeat Problem 5.8 for a magnetic field applied along the X-axis.

5.10 (a) Find the force on the circular portion of the conductor in Fig. 5-26 if the current is I and the uniform magnetic field $\mathscr{B}$ is directed upward. (b) Show that the force is the same as if the conductor between P and Q were straight.

5.11 Consider a square coil of wire, 0.6 m on a side; it carries a constant current of 0.1 A and is in a uniform magnetic field of strength 10^{-4} T. (a) If the plane of the coil is initially parallel to the magnetic field, is there any torque on the coil? (b) Answer (a) for a coil that is initially perpendicular to the magnetic field. (c) Express the torque as a function of the angle the normal to the coil makes with the magnetic field. (d) Plot the torque as the angle varies from 0 to 2π. (b) If at the point where there is no torque on the coil, the coil has an angular velocity, what happens?

5.12 (a) If the state of affairs is as stated in (b) of Problem 5.11, but the direction of the electric current is instantaneously reversed, what happens? (b) How would you change the direction of the current as mentioned in the first part of this question? (c) Of what use would this change be?

5.13 (a) Compute the intensity of the magnetic field produced by an infinitely long wire of radius 5×10^{-3} m carrying an electric current of 1 A at a distance of 5×10^{-2} m and at 1 m. (b) Compute the electric field at these points.

5.14 A long straight wire carries a current of 1.5 A. An electron travels with a velocity of 5×10^4 m s^{-1} parallel to the wire, 0.1 m from it, and in the same direction as the current. What force does the magnetic field of the current exert on the moving electron?

5.15 A long straight wire carries a current of 10 A along the Y-axis. A uniform magnetic field $\mathscr{B}$, which is 10^{-6} T, is directed parallel to the X-axis. What is the resultant magnetic field at the following points: (a) $x = 0$, $z = 2$ m; (b) $x = 2$ m, $z = 0$; (c) $x = 0$, $z = -0.5$ m?

5.16 Two long straight parallel wires are separated by a distance $2a$. If the wires carry equal currents in opposite directions, what is the magnetic field in the plane of the wires and at a point (a) midway between them, and (b) at a distance a from one and $3a$ from the other? (c) For a case in which the wires carry equal currents in the same direction, answer (a) and (b).

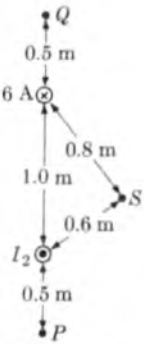


Figure 5-27

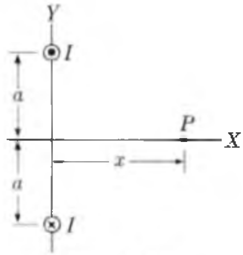


Figure 5-28

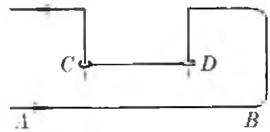


Figure 5-29

5.17 Two long straight parallel wires are 1.0 m apart as shown in Fig. 5-27. The upper wire carries a current I_1 of 6 A into the plane of the paper. (a) What must the magnitude and the direction of the current I_2 be for the resultant field at point P to be zero? (b) What is then the resultant field at Q? (c) at S?

5.18 Figure 5-28 is an end view of two long parallel wires perpendicular to the XY-plane, each carrying a current I, but in opposite directions. (a) Show by vectors the magnetic field at point P. (b) In terms of the coordinate x of the point, derive the expression for the magnitude of $\mathscr{B}$ at any point on the X-axis. (c) Construct a graph of the magnitude of $\mathscr{B}$ at any point on the X-axis. (d) At what value of x is $\mathscr{B}$ a maximum?

5.19 Repeat Problem 5.18 for points on the Y-axis.

5.20 Repeat Problem 5.18 for currents that are in the same direction.

5.21 Repeat Problem 5.19 for currents in the same direction.

5.22 In Fig. 5-28, a third long straight wire, parallel to the other two, passes through point P. Each wire carries a current $I = 20$ A. Let $a = 0.30$ m and $x = 0.4$ m. Find the magnitude and the direction of the force per unit length on the third wire (a) if the current in it is away from the plane of the paper and (b) if the current is toward the plane of the paper.

5.23 A closely wound coil has a diameter of 0.4 m and carries a current of 2.5 A. How many turns does the coil have if the magnetic field at the center of the coil is 2.356×10^{-4} T?

5.24 A solenoid is 0.3 m long and is wound with two layers of wire. The inner layer consists of 300 turns; the outer layer, 250 turns. The current is 3 A in the same direction in both layers. What is the magnetic field at a point near the center of the solenoid?

5.25 A long horizontal wire AB in Fig. 5-29 rests on the surface of a table. Another wire CD, vertically above the first, is 1.00 m long and is free to slide up and down on the two vertical metal guides C and D. The two wires are connected through the sliding contacts and carry a current of 50 A. The mass of the wire CD is 5×10^{-3} kg m^{-1}. To what equilibrium height will the wire CD rise if the magnetic force on it is caused by the current in the wire AB?

5.26 A long straight wire and a rectangular wire frame lie on a table top (Fig. 5-30). The side of the frame parallel to the wire is 0.3 m long; the side perpendicular to the wire is 0.5 m long. The currents are $I_1 = 10$ A and

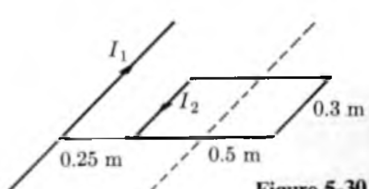


Figure 5-30

$I_2=20$ A. (a) What is the force on the loop? (b) What is the torque on the loop about the straight wire as an axis? about the dashed line as an axis?

5.27 Two long parallel wires are hung from a common axis by cords 0.04 m in length. The wires have a mass of 5×10^{-2} kg m^{-1} and carry the same current in opposite directions. What is the current if the cords hang at an angle of 6° with the vertical?

CHALLENGING PROBLEMS

5.28 A circular loop of radius a and current I is centered on the Z-axis and perpendicular to it. A magnetic field is produced having axial symmetry around the Z-axis and making an angle θ with the Z-axis and the points of the loop (Fig. 5-31). (a) For each of the two possible directions of the current, find the magnitude and the direction of the force. (b) Assume that the circuit is very small and can be considered as a magnetic dipole, and that the magnetic field follows an inverse-square law ($\mathscr{B}=k/r^2$). Show that the force on the circuit is

$$F=\pm M\left(\frac{d\mathscr{B}}{dr}\right)$$

where M is its magnetic dipole moment, which is oriented along the Z-axis. This result is general and shows that a dipole will move in the direction in which the field increases when oriented along the field, and in the opposite direction when oriented opposite to the field. (Compare with the similar result for an electric dipole in Section 1.1.)

5.29 Show that the force on a portion PQ of a conducting wire shown in Fig. 5-32, carrying a current I and placed in a uniform magnetic field, $\mathscr{B}$, is $I(\overrightarrow{PQ})\times\mathscr{B}$ and thus is independent of the shape of the conductor. (b) Apply to Problem 5.10. (c) Conclude from this that the force on a closed current placed in a uniform magnetic field is zero.

5.30 Show that the magnetic field produced by a rectilinear current I of finite length is

$$\mathscr{B}=\left(\frac{\mu_0 I}{2\pi R}\right)(\sin\alpha_1-\sin\alpha_2)$$

where R is the perpendicular distance from the point to the wire, and α_1 and α_2 are the angles between the lines from the point to the ends of the current and the perpendicular to the current (see Fig. 5-33).

5.31 Apply the result of Problem 5.30 to obtain the magnetic field at the center of a square circuit of side L. [Note the signs of the angles.]

5.32 A thin flat conductor of great length has a uniform current density j per unit width. That is, $I_{total}=jw$ where w is the width (Fig. 5-34). (a) Calculate the magnetic field at a point P at a perpendicular distance d above the center of

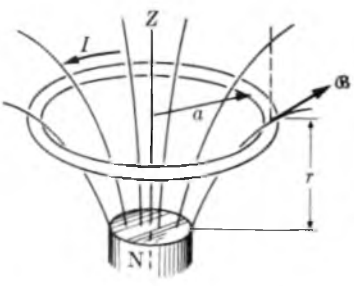


Figure 5-31

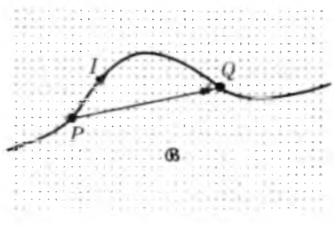


Figure 5-32

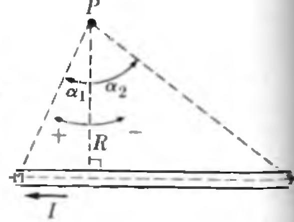


Figure 5-33

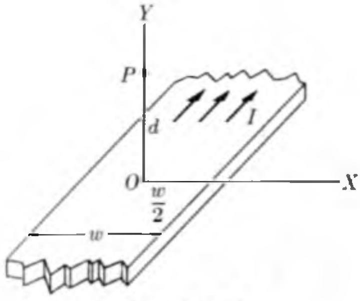


Figure 5-34

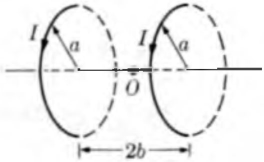


Figure 5-35

the strip as shown. [*Hint:* The expression for the field caused by a long straight strip of width dw is the same as the expression for a long straight wire.] (b) What is the field if $d \ll w$; that is, if the strip becomes an infinite plane?

5.33 Two circular currents of the same strength I and the same radius a are separated the distance $2b$ as shown in Fig. 5-35. (a) Prove that the magnetic field at points along the axis is given by

$$\mathscr{B}=\frac{\mu_0 I a^2}{(a^2+b^2)^{3/2}}\left[1+\frac{3}{2}\frac{(4b^2-a^2)}{(a^2+b^2)^2}x^2 + \frac{15}{8}\frac{(8b^4-12a^2b^2+a^4)}{(a^2+b^2)^4}x^4+\cdots\right]$$

where x is measured from the midpoint between the two currents. (b) Verify that for $a=2b$, the field at the center is independent of x up to the third power. (This arrangement is called *Helmholtz coils* and is widely used in the laboratory to produce a uniform magnetic field over a limited region of space.) (c) Assuming that the condition in (b) is fulfilled, find the value of x in terms of a for which the field differs by 1 percent from the field at the midpoint.

CHAPTER SIX

THE STATIC MAGNETIC FIELD

6.1 Introduction

In Chapters 4 and 5 we discussed how moving charges and accordingly electric currents produce magnetic fields. We also gave specific rules, such as the Ampère-Laplace law, Eq. (5.13), for the computation of such fields. In the present chapter some general properties of the magnetic field will be examined. These properties will allow simplification of the calculation of the magnetic fields produced by currents having certain symmetries. In addition a description of the magnetic properties of matter will be given as well as a discussion of how the magnetic fields produced by electric currents are affected in the vicinity of matter. This analysis will be similar to that carried out in Chapter 2 for the electric field.

6.2 Ampère's Law for the Magnetic Field

Consider first an infinite rectilinear filament carrying a current I (Fig. 6-1). Recalling from Chapter 5 the results of the experiments of Oersted, Ampère, and others, we recognize that the current in the filament produces a magnetic field $\mathcal{B}$ about the wire. At the point A the magnetic field is perpendicular to OA and is given by Eq. (5.20) as

$$\mathcal{B}=\frac{\mu_0 I}{2\pi r}\,u_\theta.$$

The concept of circulation of a vector field along a closed path was defined in Section 2.6: when applied to the magnetic field, circulation becomes

$$\Lambda_{\mathcal{B}}=\oint \mathcal{B}\cdot dl. \tag{6.1}$$

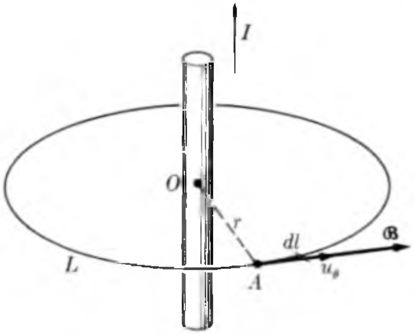


Fig. 6-1. Magnetic field of a rectilinear current.

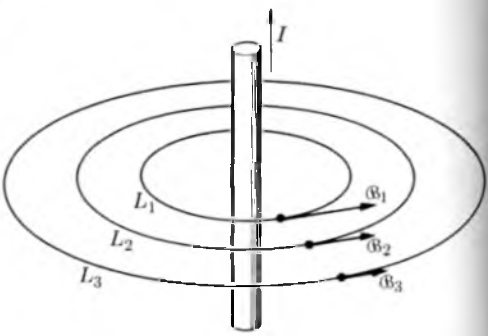


Fig. 6-2. The magnetic circulation along all concentric circular paths around a rectilinear current is the same, and equal to $\mu_0 I$.

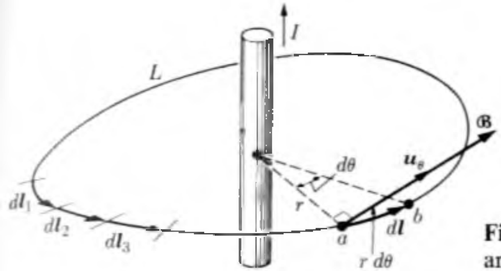


Fig. 6-3. Circulation of the magnetic field along an arbitrary closed curve.

The circulation of the magnetic field* around a circular path that has radius r and is concentric with the filament and in a plane perpendicular to the filament may be easily calculated. The magnetic field $\mathscr{B}$ is tangent to the path so that $\mathscr{B}\cdot dl = \mathscr{B}\, dl$, and is constant in magnitude. Therefore the magnetic circulation (designated by $\Lambda_{\mathscr{B}}$) is

$$\Lambda_{\mathscr{B}} = \oint_L \mathscr{B}\cdot dl = \oint_L \mathscr{B}\, dl = \mathscr{B}\oint_L dl = \mathscr{B}L = \left(\frac{\mu_0 I}{2\pi r}\right)(2\pi r) = \mu_0 I. \tag{6.2}$$

The magnetic circulation is then proportional to the electric current I, and is independent of the radius of the path. Therefore if around the current I several circles $L_1, L_2, L_3, \ldots$ are drawn (Fig. 6-2), the magnetic circulation around all of them is the same and is equal to $\mu_0 I$ according to Eq. (6.2).

Next consider an arbitrary closed path L surrounding the current I (Fig. 6-3). The magnetic circulation along L is

$$\Lambda_{\mathscr{B}} = \oint_L \mathscr{B}\cdot dl = \frac{\mu_0 I}{2\pi}\oint_L \frac{u_\theta \cdot dl}{r}.$$

However $u_\theta \cdot dl$ is the component of dl in the direction of the unit vector u_θ, and therefore is equal to $r\, d\theta$. Hence

$$\Lambda_{\mathscr{B}} = \frac{\mu_0 I}{2\pi}\oint_L d\theta = \frac{\mu_0 I}{2\pi}(2\pi) = \mu_0 I$$

since the total plane angle around a point is 2π. This equation is again the previous result. Therefore Eq. (6.2) is valid for any closed path around the rectilinear current irrespective of the position of the current relative to the path. A more elaborate analysis, which will be omitted, indicates that Eq. (6.2) is correct for *any* shape of the current loops, not necessarily only a rectilinear one. If there are several currents

*The circulation of the magnetic field is called the *magnetomotive force* along the closed circuit.

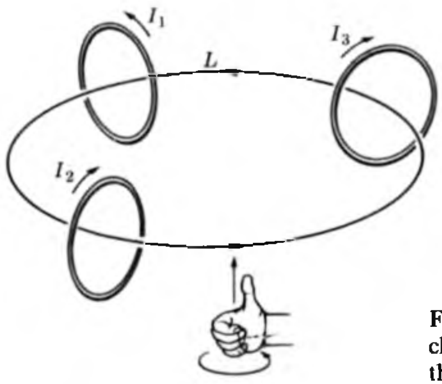


Fig. 6-4. The magnetic circulation along any closed path is proportional to the net current through the path.

$I_1, I_2, I_3, \ldots$ linked by a closed line L (Fig. 6-4), each current makes a contribution to the circulation of the magnetic field along L. Therefore Ampère's law for the magnetic field may be stated as:

the circulation of the magnetic field along a closed path that links currents $I_1, I_2, I_3, \ldots$ is

$$\Lambda_{\mathscr{B}} = \oint \mathscr{B} \cdot dl = \mu_0 I \tag{6.3}$$

where $I = I_1 + I_2 + I_3 + \cdots$ stands for the total current linked by the path L.

To apply Eq. (6.3), a current is taken as positive if it passes through the interior of L in the sense indicated by the thumb of the right hand when the fingers point in the same sense as the orientation of L, and as negative if the current is in the opposite sense. Thus in Fig. 6-4 currents I_1 and I_3 are considered positive; and I_2, negative.

Recalling Example 2.1, we found that the electric current can be expressed as the flux of the current density (that is, $I = \int_S \boldsymbol{j} \cdot \boldsymbol{u}_N \, dS$). Consequently, Ampère's law, Eq. (6.3), may be expressed in the form

$$\Lambda_{\mathscr{B}} = \oint_L \mathscr{B} \cdot dl = \mu_0 \int_S \boldsymbol{j} \cdot \boldsymbol{u}_N \, dS \tag{6.4}$$

where S is any surface bounded by L.

That the circulation of the magnetic field $\mathscr{B}$ is not zero generally indicates that the magnetic field does not have a magnetic potential in the same sense that the electric field has an electric potential.

Ampère's law is particularly useful when we must compute the magnetic field produced by current distributions having certain geometrical symmetries. The following examples will show the effectiveness of Ampère's law when compared with the more difficult procedures necessary to handle the same problems with the Ampère-Laplace law, Eq. (5.13).

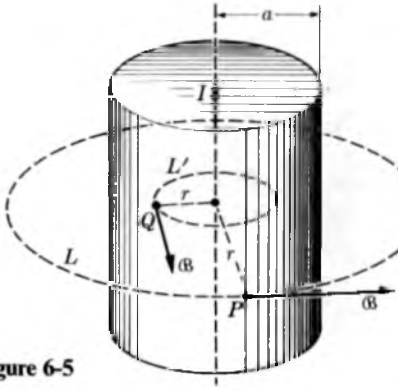


Figure 6-5

Example 6.1. The magnetic field produced by a current along a circular cylinder of infinite length.

▼ Let us consider a current I along a cylinder of radius a (Fig. 6-5). The symmetry of the problem clearly suggests that the lines of force of the magnetic field are circles with their centers along the axis of the cylinder, and that the magnetic field $\mathscr{B}$ at a point depends only on the distance from the point to the axis. Therefore choosing a circle of radius r concentric with the current as the path L gives the magnetic circulation as

$$\Lambda_{\mathscr{B}} = \oint_L \mathscr{B}\, dl = \mathscr{B} \oint_L dl = \mathscr{B}L = 2\pi r \mathscr{B}.$$

If the radius r is larger than the current radius a, all the current I passes through the circle. Therefore applying Eq. (6.3) gives

$$2\pi r \mathscr{B} = \mu_0 I \qquad \text{or} \qquad \mathscr{B} = \frac{\mu_0 I}{2\pi r} \qquad (r > a). \tag{6.5}$$

This result is just that found in Chapter 5 for the magnetic field of a current in a filament. Therefore *at points outside a cylindrical current the magnetic field is the same as if all the current were concentrated along the axis.*

For the case in which r is smaller than a, there are two possibilities. If the current is only along the surface of the cylinder (as may occur if the conductor is a cylindrical sheath of metal), the current through L' is zero; and Ampère's law gives $2\pi r \mathscr{B} = 0$ or $\mathscr{B} = 0$. Therefore *the magnetic field at points inside a cylinder carrying a current on its surface is zero.* However if the current is uniformly distributed throughout the cross section of the conductor, the current through L' is

$$I' = \frac{I}{\pi a^2}(\pi r^2) = \frac{I r^2}{a^2}.$$

Therefore Ampère's law gives $2\pi r \mathscr{B} = \mu_0 I' = \mu_0 I r^2/a^2$ or

$$\mathscr{B} = \frac{\mu_0 I r}{2\pi a^2} \qquad (r < a). \tag{6.6}$$

Thus *the magnetic field at a point inside a cylinder carrying a current uniformly distributed throughout its cross section is proportional to the distance from the point to the axis of the cylinder.* ▲

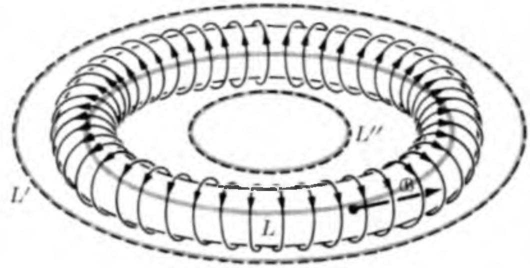


Fig. 6-6. A toroidal coil.

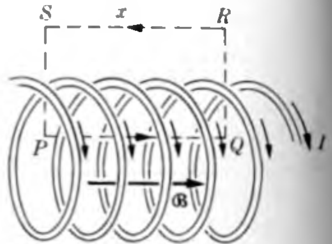


Fig. 6-7. A solenoid.

Example 6.2. The magnetic field produced by a toroidal coil.

▼ A *toroidal coil* consists of a wire uniformly wound on a torus, or doughnut-shaped surface as in Fig. 6-6. Let N be the number of turns, all equally and closely spaced, and I be the electric current in them. The symmetry of the problem suggests that the lines of force of the magnetic field are circles concentric with the torus. First take as the path of integration a circle L within the torus. The magnetic circulation is then $\Lambda_{\mathscr{B}} = \mathscr{B}L$. Path L links with all the turns around the torus, and therefore the total current through the path is NI. Thus Ampère's law gives $\mathscr{B}L = \mu_0 NI$ or

$$\mathscr{B} = \mu_0 NI/L.$$

If the cross-sectional radius of the torus is small compared with the radius of the torus, we may assume that L is practically the same for all interior paths. Given that $n = N/L$ is the number of turns per unit length, the magnetic field inside the torus is nearly uniform and has the constant value

$$\mathscr{B} = \mu_0 nI. \tag{6.7}$$

For any path, such as L' or L'', outside the torus, the total current linking with the path is zero. Therefore $\mathscr{B} = 0$. In other words *the magnetic field of a toroidal coil is entirely confined to its interior.* This situation applies only for toroidal coils on which the turns of the wire are very closely spaced. ▲

Example 6.3. The magnetic field at the center of a very long solenoid.

▼ Consider a solenoid (Fig. 6-7), having n turns per unit length, each turn carrying a current I. If the turns are very closely spaced and the solenoid is very long, it can be verified that the magnetic field is entirely confined to the interior and is nearly uniform. For the path of integration choose the rectangle $PQRS$ of Fig. 6-7. The contribution by sides QR and SP to the magnetic circulation is zero because the field is perpendicular to them; also the contribution of side RS is zero because there is no field there. Therefore only side PQ contributes the amount $\mathscr{B}x$ so that $\Lambda_{\mathscr{B}} = \mathscr{B}x$. The total current linking with the integration path is nxI since nx gives the number of turns within the length x. Therefore Ampère's law gives $\mathscr{B}x = \mu_0 nxI$, or $\mathscr{B} = \mu_0 nI$, in agreement with the result in Example 5.6 for the field at the center of a long solenoid. ▲

6.3 Ampère's Law in Differential Form

Since Ampère's law can be applied to a path of any shape, the law will now be applied to a very small or infinitesimal rectangular path $PQRS$ in the XY-plane; the path has sides dx and dy and an area $dx\,dy$ (Fig. 6-8). The sense of circulation around $PQRS$ is as indicated by the arrows. The circulation $\Lambda_{\mathscr{B}}$ consists of four terms, one for each side; that is,

$$\Lambda_{\mathscr{B}}=\oint_{PQRS}\mathscr{B}\cdot dl=\int_{PQ}+\int_{QR}+\int_{RS}+\int_{SP}\mathscr{B}\cdot dl. \tag{6.8}$$

Now along the path QR, which is oriented parallel to the $+Y$-direction, $dl=u_y\,dy$ and

$$\int_{QR}\mathscr{B}\cdot dl=\mathscr{B}\cdot u_y\,dy=\mathscr{B}_y\,dy.$$

Similarly for side SP, which is oriented in the $-Y$-direction, $dl=-u_y\,dy$ and thus

$$\int_{SP}\mathscr{B}\cdot dl=-\mathscr{B}'\cdot u_y\,dy=-\mathscr{B}'_y\,dy$$

so that

$$\int_{QR}+\int_{SP}=(\mathscr{B}_y-\mathscr{B}'_y)\,dy.$$

However since $PQ=dx$ is the distance between the points where $\mathscr{B}_y$ and $\mathscr{B}'_y$ are calculated and because $\mathscr{B}_y-\mathscr{B}'_y=d\mathscr{B}_y$, the partial differential notation may be used to write $\mathscr{B}_y-\mathscr{B}'_y$ as $(\partial\mathscr{B}_y/\partial x)\,dx$. Therefore,

$$\int_{QR}+\int_{SP}\mathscr{B}\cdot dl=\frac{\partial\mathscr{B}_y}{\partial x}\,dx\,dy.$$

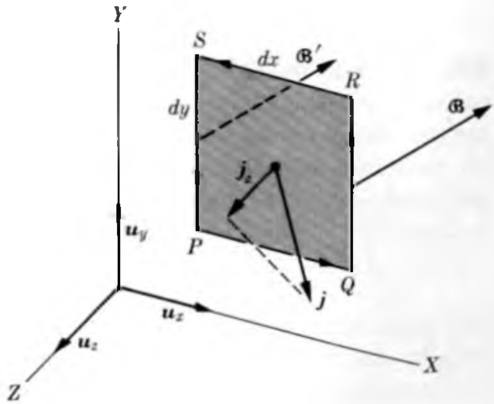


Fig. 6-8. Elementary path to evaluate Ampère's law in differential form.

By similar reasoning the remaining two integrals in Eq. (6.8) are

$$\int_{PS} + \int_{RS} \mathscr{B}\cdot dl = -\frac{\partial \mathscr{B}_x}{\partial y}\,dx\,dy.$$

Adding the two results gives the magnetic circulation over the infinitesimal path:

$$\Lambda_{\mathscr{B}} = \oint_{PQRS} \mathscr{B}\cdot dl = \left(\frac{\partial \mathscr{B}_y}{\partial x} - \frac{\partial \mathscr{B}_x}{\partial y}\right) dx\,dy. \tag{6.9}$$

Given that dI is the current passing through $PQRS$, this infinitesimal current may be related to the current density $\boldsymbol{j}$ by writing

$$dI = \boldsymbol{j}\cdot d\boldsymbol{S} = j_z\,dS = j_z\,dx\,dy \tag{6.10}$$

We write j_z because only the Z-component of the current density contributes to the current dI through $PQRS$. The components j_x and j_y correspond to current densities parallel to the surface and not through it. Substituting Eqs. (6.9) and (6.10) into Ampère's law, Eq. (6.3), yields

$$\left(\frac{\partial \mathscr{B}_y}{\partial x} - \frac{\partial \mathscr{B}_x}{\partial y}\right) dx\,dy = \mu_0\,dI = \mu_0 j_z\,dx\,dy.$$

Canceling the common factor $dx\,dy$ on both sides gives Ampère's law in its differential form for the small area:

$$\frac{\partial \mathscr{B}_y}{\partial x} - \frac{\partial \mathscr{B}_x}{\partial y} = \mu_0 j_z. \tag{6.11}$$

The surface $PQRS$ may be placed in the YZ- or the ZX-planes and results in the equivalent expressions

$$\frac{\partial \mathscr{B}_z}{\partial y} - \frac{\partial \mathscr{B}_y}{\partial z} = \mu_0 j_x, \tag{6.12}$$

$$\frac{\partial \mathscr{B}_x}{\partial z} - \frac{\partial \mathscr{B}_z}{\partial x} = \mu_0 j_y. \tag{6.13}$$

The three equations (6.11), (6.12), and (6.13) can be combined into one vector equation. Note that the right-hand sides are the components of the vector $\boldsymbol{j}$, the current density, multiplied by μ_0. Similarly the left-hand sides can be considered as the components of a vector obtained from $\mathscr{B}$ by combining derivatives in the form indicated, and this combination is called the *curl* of $\mathscr{B}$. Then the three equations can be consolidated in the vector equation

$$\text{curl}\,\mathscr{B} = \mu_0 \boldsymbol{j}. \tag{6.14}$$

This equation is the expression of Ampère's law in differential form. Equation (6.14) may be used to obtain the magnetic field when the current distribution is known, and conversely. In a region in which there are no electric currents, curl $\mathscr{B} = 0$.

Ampère's law in differential form establishes a *local* relation between the magnetic field $\mathscr{B}$ at a point and the current density $\boldsymbol{j}$ at the *same* point of space, much as Gauss's law in differential form [Eq. (2.5)] relates the electric field and the charges at the same point of space. By considering both Eq. (2.5) and Eq. (6.14), we may thus say that electric currents are the sources of the magnetic field.

The expression equivalent to Eq. (6.14) for a static electric field is

$$\operatorname{curl} \mathscr{E} = 0 \tag{6.15}$$

since Eq. (3.17) showed that for such a field the circulation is zero ($\oint_L \mathscr{E} \cdot d\boldsymbol{l} = 0$).

6.4 Magnetic Flux

The *magnetic flux* across any surface, closed or not, placed in a magnetic field is

$$\Phi_{\mathscr{B}} = \int_S \mathscr{B} \cdot \boldsymbol{u}_N \, dS. \tag{6.16}$$

The concept of magnetic flux across an arbitrary surface is of great importance especially when the surface is not closed. The magnetic flux, being magnetic field times area, must be expressed in SI units as T m^2, a unit called the *weber* (Wb) in honor of the German physicist Wilhelm E. Weber (1804–1891). Note that since T = kg s^{-1} C^{-1}, Wb = T m^2 = m^2 kg s^{-1} C^{-1}.

Since there are no magnetic masses or poles (or at least none have yet been observed), the lines of force of the magnetic field $\mathscr{B}$ are closed, as indicated in the examples discussed in Chapter 4. Therefore

the flux of the magnetic field through a closed surface is always zero.

That is, the inward flux through a closed surface is equal to the outward flux. Thus

$$\oint_S \mathscr{B} \cdot \boldsymbol{u}_N \, dS = 0, \tag{6.17}$$

a result that can also be verified mathematically from the general expression for $\mathscr{B}$ given in Eq. (5.13). (The proof will be omitted.) This result constitutes *Gauss's law for the magnetic field*. In differential form by similarity with Eq. (2.4) for the electric field,

$$\frac{\partial \mathscr{B}_x}{\partial x} + \frac{\partial \mathscr{B}_y}{\partial y} + \frac{\partial \mathscr{B}_z}{\partial z} = 0 \qquad \text{or} \qquad \operatorname{div} \mathscr{B} = 0. \tag{6.18}$$

6.5 Magnetization of Matter

In Section 5.8 a small current was shown to constitute a magnetic dipole. We may assume that the electrons in their motion around the nucleus of an atom constitute small magnetic dipoles. Depending on their symmetry or on the relative orientation of their electronic orbits, atoms may or may not exhibit a net magnetic dipole moment. Since most molecules are not spherically symmetric, they may exhibit a permanent magnetic dipole moment because of special orientation of the electronic orbits. For example diatomic molecules have axial symmetry and may possess a magnetic dipole moment parallel to the molecular axis. Even so, with the exception of ferromagnetic materials, matter in bulk does not exhibit a net magnetic moment because of the random orientation of the molecules, a situation similar to that found in the electric polarization of matter. However, the presence of an external magnetic field distorts the electronic motion and gives rise to a net magnetic polarization or *magnetization* of the material. What happens essentially is that the magnetic field produces on all the electrons a precessional or rotational motion about the direction of the local magnetic field as explained in Section 5.3. Each electron contributes a magnetic dipole moment as given by Eq. (5.9).

For simplicity consider a substance in the form of a cylinder that is magnetized uniformly parallel to the axis of the cylinder (Fig. 6-9). Thus the molecular magnetic dipoles are oriented parallel to the axis of the cylinder, and therefore the molecular electronic currents are oriented perpendicular to the axis of the cylinder. Figures 6-9 and the more detailed 6-10 show that the internal currents tend to cancel each other because of the contrary effects of adjacent currents so that no net current is observed

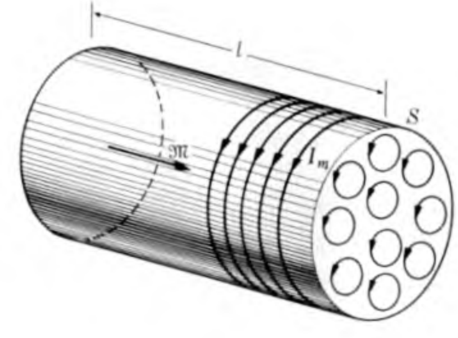


Fig. 6-9. Magnetization surface current on a magnetized cylinder.

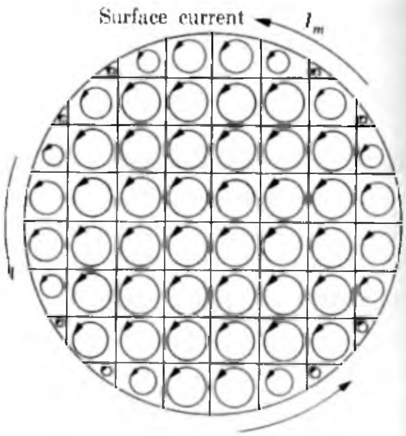


Fig. 6-10. Elementary currents in a magnetized cylinder.

inside the substance. However, the magnetization gives rise to a net current I_m on the surface of the material, which therefore behaves as if it were a solenoid.

The *magnetization vector* $\mathcal{M}$ of a material is defined as the magnetic dipole moment of the medium per unit volume. If $\boldsymbol{m}$ is the magnetic dipole moment contributed by each atom or molecule, and n is the number of atoms or molecules per unit volume, the magnetization is $\mathcal{M} = n\boldsymbol{m}$. The magnetic moment of an elementary current is expressed in A m^2; and therefore the magnetization $\mathcal{M}$ is expressed in A m^2/m^3 = A m^{-1} or m^{-1} s^{-1} C, and is equivalent to a current per unit length.

There is a very important relation between the surface current on the magnetized body and the magnetization $\mathcal{M}$. Note from Fig. 6-9 that I_m has a direction perpendicular to $\mathcal{M}$. The cylinder itself behaves like a large magnetic dipole resulting from the superposition of all individual dipoles. If S is the area of the cross section of the cylinder and l is its length, its volume is lS, and therefore its total magnetic dipole moment is $\mathcal{M}(lS) = (\mathcal{M}l)S$. However S is just the cross-sectional area of the circuits formed by the surface current. Since magnetic dipole moment = current × area, the total magnetization current that appears on the surface of the cylinder is $\mathcal{M}l$; and therefore the current per unit length I_m on the surface of the magnetized cylinder is $\mathcal{M}$, or $I_m = \mathcal{M}$. Although obtained for a particular geometrical arrangement, this result has more general validity. Thus we can say that

> *the current per unit length on the surface of a magnetized piece of matter is equal to the component of the magnetization vector $\mathcal{M}$ parallel to a plane tangent to the surface of the body and has a direction perpendicular to $\mathcal{M}$.*

6.6 The Magnetizing Field

In the preceding section a magnetized substance was shown to have certain currents on its surface (and throughout its volume if the magnetization is not uniform). These magnetization currents are "frozen" in that they are produced by electrons bound to specific atoms or molecules and are not free to move through the substance. On the other hand in some substances such as metals, there are electric charges capable of moving through the substance. The electric current produced by these free charges will be called *free* current. In many instances it is necessary to distinguish explicitly between free currents and magnetization currents.

Again consider a cylindrical piece of matter placed inside a long solenoid that is carrying a current I (Fig. 6-11). This current produces a magnetic field that magnetizes the cylinder and gives rise to a magnetization surface current on the cylinder in the same direction as I. The magnetization surface current per unit length is equal to $\mathcal{M}$. If the solenoid has n turns per unit length, the system of solenoid plus magnetized cylinder is equivalent to a single solenoid carrying a current per unit length equal to $nI + \mathcal{M}$. This effective solenoidal current gives rise to a resultant magnetic field $\mathcal{B}$

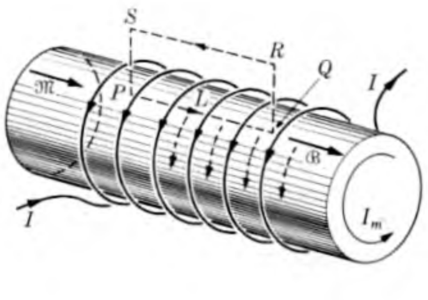


Figure 6-11

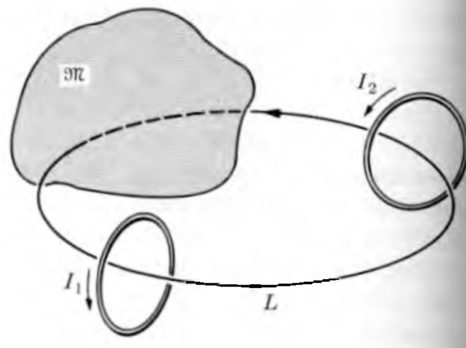


Figure 6-12

parallel to the axis of the cylinder, a field whose magnitude is given by Eq. (6.7) with nI replaced by the total current per unit length $nI+\mathcal{M}$. That is, $\mathcal{B}=\mu_0(nI+\mathcal{M})$ or

$$\frac{1}{\mu_0}\mathcal{B}-\mathcal{M}=nI.$$

This expression gives the conduction or free current per unit length, nI, on the surface of the cylinder in terms of the magnetic field $\mathcal{B}$ in the medium and the magnetization $\mathcal{M}$ of the medium. The result above suggests the introduction of a new vector field, called the *magnetizing field*, defined by

$$\mathcal{H}=\frac{1}{\mu_0}\mathcal{B}-\mathcal{M}. \tag{6.19}$$

The magnetizing field is expressed in A m^{-1} or m^{-1} s^{-1} C, which are the SI units of the two terms that appear on the right-hand side.

In this special example $\mathcal{H}=n\boldsymbol{I}$, which relates $\mathcal{H}$ to the conduction or free currents per unit length of the solenoid. When a length $PQ=L$ along the surface is considered, then

$$\mathcal{H}L=Ln\boldsymbol{I}=\boldsymbol{I}_{\text{free}} \tag{6.20}$$

where $\boldsymbol{I}_{\text{free}}=Ln\boldsymbol{I}$ is the total free current of the solenoid corresponding to the length L. Computing the circulation of $\mathcal{H}$ around the rectangle $PQRS$, gives $\Lambda_{\mathcal{H}}=\mathcal{H}L$ since $\mathcal{H}$ is zero outside the solenoid (both $\mathcal{B}$ and $\mathcal{M}$ are) and the sides QR and SP do not contribute to the circulation since they are perpendicular to the magnetic field. Thus Eq. (6.20) may be written in the form $\Lambda_{\mathcal{H}}=I_{\text{free}}$ where I_{free} is the total free current across the rectangle $PQRS$. This result has more general validity than this simplified proof may suggest. In fact it can be verified that *the circulation of the magnetizing field along a closed line is equal to the total free current through the path*. That is,

$$\Lambda_{\mathcal{H}}=\oint_L \mathcal{H}\cdot d\boldsymbol{l}=\boldsymbol{I}_{\text{free}} \tag{6.21}$$

where I_{free} is the total current linking with the path L caused by free-flowing charges in the medium or in an electric circuit, but excluding currents caused by the magnetization of matter. For example if the path L (Fig. 6-12) links with circuits I_1 and I_2 and a body with magnetization $\mathcal{M}$, Eq. (6.21) includes only the currents I_1 and I_2 while in Ampère's law, Eq. (6.3), for the magnetic field $\mathcal{B}$, all currents must be included; that is, I_1 and I_2, produced by freely moving charges, as well as those currents produced by the magnetization $\mathcal{M}$ of the body resulting from bound electrons.

Equation (6.19) may be written in the form

$$\mathcal{B}=\mu_0(\mathcal{H}+\mathcal{M}). \tag{6.22}$$

Since the magnetization $\mathcal{M}$ of the body is physically related to the resultant magnetic field $\mathcal{B}$, a relation between $\mathcal{M}$ and $\mathcal{B}$ could be introduced similar to the relation [Eq. (2.10)] between $\mathcal{P}$ and $\mathcal{E}$ in the electrical case. However for historical reasons it is customary to proceed differently and express a relation between $\mathcal{M}$ and $\mathcal{H}$* instead, by writing

$$\mathcal{M}=\chi_m\mathcal{H}. \tag{6.23}$$

The quantity χ_m is called the *magnetic susceptibility* of the material, and is a pure number independent of the units chosen for $\mathcal{M}$ and $\mathcal{H}$. Substituting Eq. (6.23) into Eq. (6.22) gives

$$\mathcal{B}=\mu_0(\mathcal{H}+\chi_m\mathcal{H})=\mu_0(1+\chi_m)\mathcal{H}=\mu\mathcal{H} \tag{6.24}$$

where

$$\mu=\mathcal{B}/\mathcal{H}=\mu_0(1+\chi_m) \tag{6.25}$$

is called the *permeability* of the medium and is expressed in the same units as μ_0; that is, m kg^{-2} C. The relative permeability is defined by

$$\mu_r=\mu/\mu_0=1+\chi_m \tag{6.26}$$

and is a pure number independent of the system of units.

When the relation $\mathcal{B}=\mu\mathcal{H}$ holds, Eq. (6.21) may be rewritten as

$$\oint_L \frac{1}{\mu}\mathcal{B}\cdot dl=I_{\text{free}}.$$

If the medium is homogeneous so that μ is constant, the circulation of the magnetic field is

$$\Lambda_{\mathcal{B}}=\oint_L \mathcal{B}\cdot dl=\mu I_{\text{free}}. \tag{6.27}$$

This result is similar to Ampère's law, Eq. (6.3), but with the total current replaced by the free current and μ instead of μ_0. Thus the effect of magnetized matter on the magnetic field $\mathcal{B}$ is to replace μ_0 by μ. For example the magnetic field of a rectilinear

*This relationship is valid only for certain magnetic materials, such as isotropic diamagnets and ordinary paramagnetic materials.

current I embedded in a magnetized medium is

$$\mathscr{B}=\frac{\mu I}{2\pi r} \tag{6.28}$$

instead of the formula given by Eq. (5.20).

6.7 Calculation of Magnetic Susceptibility

Section 6.6 indicated that the magnetic susceptibility χ_m expresses the response of a medium to an external magnetic field; like the electric susceptibility χ_e, the magnetic susceptibility is related to the properties of the atoms and molecules of the medium. Two effects enter into the phenomenon of the magnetization of matter by an external magnetic field. One is a *distortion* of the electronic motion caused by the magnetic field. The other is an *orientation effect* when the atom or molecule has a permanent magnetic dipole moment. Both effects contribute to the value of χ_m and will be discussed separately.

Distortion effect

A magnetic field exerts a force on a moving charge. Therefore if an external magnetic field is applied to a substance, the electrons moving in the atoms or molecules are subject to an additional force caused by the applied magnetic field. This force results in a perturbation of the electronic motion. To evaluate this perturbation precisely would require the methods of quantum mechanics. Thus we shall limit ourselves to stating the main results and providing a simplified illustration in Example 6.4.

The effect of a magnetic field on the electronic motion in an atom is equivalent to an additional current induced in the atom. This current is oriented in a direction such that the magnetic dipole moment associated with the current is in the direction *opposite* to that of the magnetic field. Since this effect is independent of the orientation of the atom and is the same for all atoms, *the substance has acquired a magnetization $\mathscr{M}$ opposed to the magnetic field*, a result in contrast to that found in the electric field case. This behavior, called *diamagnetism*, is common to all substances although in many materials it is masked by the paramagnetic effect described below. The resulting magnetization is given by

$$\mathscr{M}=-\frac{ne^2\mu_0}{6m_e}\left(\sum_i r_i^2\right)_{\text{ave}}\mathscr{H} \tag{6.29}$$

where $\mathscr{H}$ is the magnetizing field in the substance, n is the number of atoms per unit volume, and r_i is the distance of the ith electron from the nucleus in an atom. The summation extends over all the electrons in the atom, and the average must be computed according to the prescriptions of quantum mechanics. The other quantities have their usual meaning. The negative sign is due to the fact that $\mathscr{M}$ is opposed to $\mathscr{H}$.

Then according to Eq. (6.23), the magnetic susceptibility is

$$\chi_m = -\frac{ne^2\mu_0}{6m_e}\left(\sum_i r_i^2\right)_{\text{ave}}; \tag{6.30}$$

and since χ_m is negative, the relative permeability $\mu_r = 1 + \chi_m$ is smaller than one. If the values of the known constants are placed in Eq. (6.30) and n is assumed to be about 10^{28} atoms per m^3 in a solid and r_i about 10^{-10} m (which is the order of magnitude of an electronic orbit), then χ_m is of the order of magnitude of 10^{-5} for solids, in agreement with the values listed in Table 6-1.

Orientation effect

An atom or molecule may have a permanent magnetic dipole moment, associated with the angular momentum of its electrons. In this case the presence of an external magnetic field produces a torque that tends to align all the magnetic dipoles along the magnetic field, and results in an additional magnetization called *paramagnetism*. The magnetism acquired by a paramagnetic substance is therefore in the direction of the magnetic field. This effect is much stronger than diamagnetism, and in the case of paramagnetic substances the diamagnetic effects are in general completely screened by the paramagnetic effects.

The *paramagnetic susceptibility* of gases is given approximately by an expression similar to Eq. (2.26) for the electric susceptibility produced by polar molecules:

$$\chi_m = \frac{nm_0^2\mu_0}{3kT} \tag{6.31}$$

where m_0 is the permanent atomic or molecular magnetic moment, T is the absolute temperature of the substance, and k is the Boltzmann constant. Equation (6.31) is valid only when $m_0\mathscr{B}/kT \ll 1$. Otherwise the relation between $\mathscr{M}$ and $\mathscr{B}$ is more complex. As in the electric case, χ_m decreases if the temperature of the substance increases. This temperature dependence is due to the molecular motion, which

Table 6-1. Magnetic Susceptibilities at Room Temperature

Diamagnetic substances	χ_m	Paramagnetic substances	χ_m
Hydrogen*	-2.1×10^{-9}	Oxygen*	2.1×10^{-6}
Nitrogen*	-5.0×10^{-9}	Magnesium	1.2×10^{-5}
Sodium	-2.4×10^{-6}	Aluminum	2.3×10^{-5}
Copper	-1.0×10^{-5}	Tungsten	6.8×10^{-5}
Bismuth	-1.7×10^{-5}	Titanium	7.1×10^{-5}
Diamond	-2.2×10^{-5}	Platinum	3.0×10^{-4}
Mercury	-3.2×10^{-5}	Gadolinium chloride ($GdCl_3$)	2.8×10^{-3}

*Gases at 100 kPa.

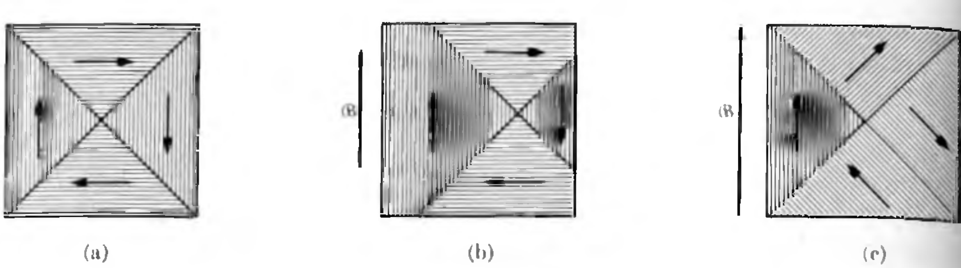


Fig. 6-13. Magnetic domains. (a) Unmagnetized substances, (b) magnetization by domain growth, (c) magnetization by domain orientation. The arrows indicate the direction of magnetization of each domain.

increases with temperature and therefore tends to offset the aligning effect of the magnetic field. The order of magnitude of the atomic magnetic dipole moment is 10^{-23} J T^{-1}. Thus, when the values of the other constants are placed in Eq. (6.31), the paramagnetic susceptibility at room temperature (298° K) has an order of magnitude of 10^{-4} for solids and 10^{-7} for gases at STP. This result is in satisfactory agreement with the values given in Table 6-1 for paramagnetic substances.

An important conclusion is that for both paramagnetic and diamagnetic substances χ_m is very small compared with unity, and in many instances $\mu_r = 1 + \chi_m$ may be replaced by one.

Other effects

A third class of magnetic substances is called *ferromagnetic*. The chief characteristic of ferromagnetic substances is that they exhibit a permanent magnetization, which suggests a natural tendency of the magnetic dipole moments of the atoms or molecules to align under their mutual interactions. The lodestone and other natural magnets mentioned at the beginning of Chapter 4 are examples of ferromagnetic substances. For these substances no linear relationship between $\mathcal{M}$ and $\mathcal{B}$ exists. Ferromagnetism is thus similar to ferroelectricity in overall behavior although their origins differ. Ferromagnetism is associated with an interaction between the spins $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$ of two electrons in the same atom. The interaction is basically of the form $-J\boldsymbol{S}_1 \cdot \boldsymbol{S}_2$ where the quantity J, called the *exchange integral*, depends on the distance between the electrons. When J is positive, equilibrium is attained if $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$ are parallel, resulting in a parallel orientation of electronic spins in microscopic regions called *domains* (Figs. 6-13a and 6-14a), which have dimensions of the order of 10^{-8} to 10^{-12} m^3 and which contain from 10^{21} to 10^{17} atoms. The direction of magnetization of a domain depends on the crystal structure of the substance. For example for iron, which crystalizes with a cubic structure, the directions of *easy* magnetization are along the three axes of the cube. In a piece of matter the domains themselves may be oriented in different directions; thus the net, macroscopic effect can be zero or negligible. In the presence of an external magnetic field those domains oriented favorably

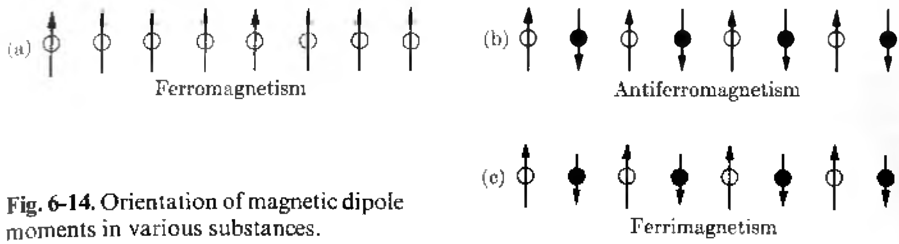


Fig. 6-14. Orientation of magnetic dipole moments in various substances.

with respect to the magnetic field grow at the expense of those oriented less favorably (Fig. 6-13b); as the strength of the external magnetic field increases, the magnetization of the domains tends to align in the direction of the field (Fig. 16-13c), and the piece of matter becomes a *magnet*. Ferromagnetism is a property that depends on temperature; and for each ferromagnetic substance there is a temperature, called the *Curie temperature*, above which the substance becomes paramagnetic. This phenomenon occurs when thermal motion is great enough to offset the aligning forces. Substances that are ferromagnetic at room temperature are iron, nickel, cobalt, and gadolinium. Their Curie temperatures are 1043 K, 638 K, 1348 K, and 288 K, respectively.

However, for some substances J is negative. Equilibrium is then attained if the electronic spins are antiparallel; and a zero net magnetization (Fig. 6-14b) results. In this case the substance is called *antiferromagnetic*. Some antiferromagnetic substances are MnO, FeO, CoO, and NiO.

Another type of magnetization is called *ferrimagnetism*. It is similar to antiferromagnetism, but the atomic or ionic magnetic dipole moments in one direction are different from those oriented in the opposite direction, and the result is a net magnetization (Fig. 6-14c). These substances are called *ferrites* and can be generally represented by the chemical formula $MOFe_2O_3$ where M stands for Mn, Co, Ni, Cu, Mg, Zn, Cd, etc. Note that if M is Fe, the compound Fe_3O_4, or magnetite, results.

Example 6.4. The atomic magnetic dipole moment induced by an external magnetic field.

▼ That an external magnetic field produces in an atom a magnetic dipole moment in a direction opposite to the field will be justified by a very simple model. Consider an electron having a charge of $-e$ and revolving about a nucleus N. For simplicity assume that the orbit of the electron is circular, has radius ρ, and is in the XY-plane. If ω_0 is the electron's angular velocity and $\boldsymbol{F}$ the force on the electron caused by the nucleus, the equation of motion of the electron is then

$$m_e\omega_0^2\rho = F. \tag{6.32}$$

If now a magnetic field $\mathscr{B}$ is applied along the Z-axis (that is, perpendicular to the plane of the orbit), an additional force $\boldsymbol{F}' = -e\boldsymbol{v} \times \mathscr{B}$ is exerted on the electron. This force will be either in the same direction as $\boldsymbol{F}$ or in the opposite direction, depending on the relative orientation of ω_0

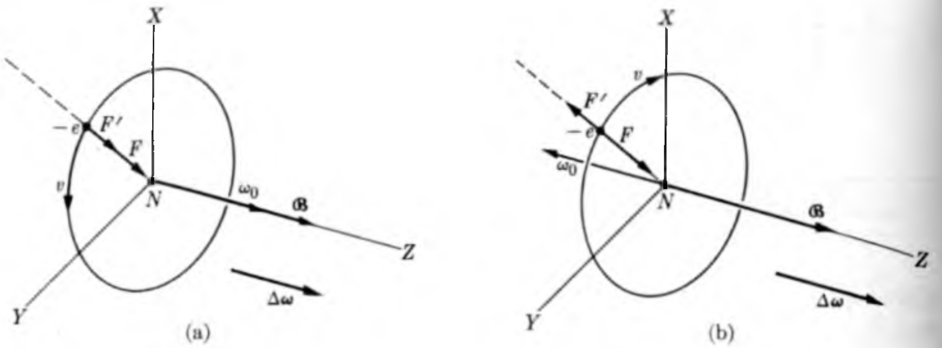


Fig. 6-15. Explanation of diamagnetism.

and $\mathcal{B}$ as indicated in Fig. 6-15. Since the radial force on the electron has changed, the angular frequency (assuming the radius remains the same) will also change and become ω. From $v=\omega\rho$ and the fact that the magnitude of F' is $ev\mathcal{B}$, the equation of motion of the electron is now

$$m_e\omega^2\rho = F \pm e\omega\rho\mathcal{B} \tag{6.33}$$

where the plus sign holds for case (a) of Fig. 6-15 and the minus sign for case (b). Substituting Eq. (6.32) into Eq. (6.33) to eliminate F yields

$$m_e(\omega^2-\omega_0^2)\rho = \pm e\omega\mathcal{B} \qquad \text{or} \qquad m_e(\omega+\omega_0)(\omega-\omega_0) = \pm e\omega\mathcal{B}.$$

Now if the change in frequency, $\Delta\omega=\omega-\omega_0$, is very small, $\omega+\omega_0$ may be replaced by 2ω without great error to give

$$2m_e\Delta\omega = \pm e\mathcal{B} \qquad \text{or} \qquad \Delta\omega = \pm\frac{e}{2m_e}\mathcal{B}.$$

This change in frequency is equal to the Larmor frequency Ω_L, which will be discussed in detail in Section 8.6. The plus sign holding in case (a) means an increase in ω_0, and $\Delta\omega$ is then pointing to the right. The minus sign holding in case (b) means a decrease in ω_0, and $\Delta\omega$ is then also pointing to the right. Thus in both cases the vector relation is

$$\Delta\boldsymbol{\omega} = \frac{e}{2m_e}\mathcal{B}.$$

The change in frequency of the electronic motion produces a net current $-e(\Delta\omega/2\pi)$ and therefore from definition (5.9) a magnetic dipole moment:

$$\boldsymbol{M} = -e\left(\frac{\Delta\omega}{2\pi}\right)(\pi\rho^2) = -\frac{e^2\rho^2}{4m_e}\mathcal{B}. \tag{6.34}$$

The magnetic moment of the atom is in the direction opposite to that of the magnetic field $\mathcal{B}$, and the substance as a whole will acquire a magnetization opposed to the applied magnetic field. This calculation has been oversimplified in order to obtain a more general result; in reality the random distribution in space of the electronic orbits must be taken into account and the nature of the local magnetic field $\mathcal{B}$ acting on the electron must be analyzed in greater detail. However, the simplified calculation basically coincides with the result quoted in Eq. (6.29). ▲

Table 6-2. Equations of the Static Electromagnetic Field

Law	Integral form	Differential form
I. Gauss's law for the electric field [Eqs. (2.3) and (2.5)]	$\oint_S \mathscr{E} \cdot \boldsymbol{u}_N \, dS = \dfrac{q}{\epsilon_0}$	$\operatorname{div} \mathscr{E} = \dfrac{\rho}{\epsilon_0}$
II. Gauss's law for the magnetic field [Eqs. (6.17) and (6.18)]	$\oint \mathscr{B} \cdot \boldsymbol{u}_N \, dS = 0$	$\operatorname{div} \mathscr{B} = 0$
III. Circulation of the electric field [Eqs. (3.17) and (6.15)]	$\oint \mathscr{E} \cdot d\boldsymbol{l} = 0$	$\operatorname{curl} \mathscr{E} = 0$
IV. Circulation of the magnetic field (Ampère's law) [Eqs. (6.3) and (6.14)]	$\oint \mathscr{B} \cdot d\boldsymbol{l} = \mu_0 I$	$\operatorname{curl} \mathscr{B} = \mu_0 \boldsymbol{j}$

6.8 Summary of the Laws for Static Fields

In the last three chapters static electric and magnetic fields have been discussed as two separate entities with no relation whatsoever between them except that the sources of the electric field are electric charges and the sources of the magnetic field are electric currents. Hence two separate sets of equations were derived; these appear in both integral and differential form in Table 6-2. These equations allow computing the static electric field $\mathscr{E}$ and the static magnetic field $\mathscr{B}$ if the charges and currents are known, and conversely. It thus appears that static electric and magnetic fields can be considered as two independent fields. It is well known, however, that this is not true. In Chapter 4 the rules for relating the electric and magnetic fields as measured by two observers in uniform relative motion were derived, using the Lorentz transformation. It was noted that $\mathscr{E}$ and $\mathscr{B}$ are intimately related. Thus we may expect that in time-dependent cases the preceding equations will require some modifications. How to make these modifications is the subject of Chapter 8, in which a new set of equations that are based on experimental evidence and that are extensions of the preceding equations will be obtained.

Problems

6.1 An infinitely long rectilinear wire contains a uniformly continuous current of 10 A. The radius of the wire is 4×10^{-2} m. (a) Calculate the magnetic field at the center of the wire, at 2×10^{-2} m from the center, and at the surface of the wire. (b) Calculate the magnetic field at the surface of the wire, at 8×10^{-2} m, and at 1.5 m from the center of the wire. (c) Calculate the point at which the field is 10^{-2} as strong as that at the wire's surface.

6.2 A toroidal coil has a radius of 0.5 m and contains a constant current of 7 A. If the coil has been wound with 600 turns, calculate the magnitude of the magnetic field within the toroid.

6.3 A very long solenoid with 1400 turns per meter has a constant current of 25 A in its coils. Determine the magnetic field within the solenoid.

6.4 A hollow cylindrical conductor of radii R_1 and R_2 carries a current I uniformly distributed over its cross section (Fig. 6-16). Using Ampère's law, show (a) that the magnetic field at $r > R_2$ is $\mu_0 I/2\pi r$, (b) that the field for $R_1 < r < R_2$ is

$$\frac{\mu_0 I(r^2 - R_1^2)}{2\pi(R_2^2 - R_1^2)r},$$

and (c) that the field is zero for $r < R_1$.

6.5 A coaxial cable is formed by surrounding a solid cylindrical conductor of radius R_1 with a concentric conducting shell of inner radius R_2 and outer radius R_3 (Fig. 6-17). In usual practice a current I is sent down the inner wire and is returned via the outer shell. (a) Using Ampère's law, determine the magnetic field for all points about and within the conductor. (b) Plot $\mathscr{B}$ as a function of r. Assume uniform current density.

6.6 A solenoid is to be constructed with a magnetic field of 0.25 T in its interior. The radius of the solenoid is to be 0.1 m and the wire may carry a maximum current of 7 A. (a) How many turns per meter are needed? (b) If the solenoid is 1 m long, what length of wire is needed?

6.7 Show that in a medium in which a uniform electric current of constant density exists, the magnetic field is $\mathscr{B} = \frac{1}{2}\mu_0 \boldsymbol{j} \times \boldsymbol{r}$. (*Hint:* Verify that the relation curl $\mathscr{B} = \mu_0 \boldsymbol{j}$ holds.)

6.8 Show that there is no electric current density at a point in space at which the magnetic field is constant. (*Hint:* Let $\mathscr{B} = \boldsymbol{u}_x \mathscr{B}$ and verify that curl $\mathscr{B} = 0$.)

6.9 The magnetic field $\mathscr{B}$ in a certain region is 2 T and has the direction of the positive X-axis in Fig. 6-18. (a) What is the magnetic flux across the surface *abcd* in the figure? (b) What is the magnetic flux across the surface *befc*? (c) What is the magnetic flux across the surface *aefd*?

6.10 Determine the magnetic flux through the rectangular circuit of Fig. 6-19 when there is a current I along the straight wire.

6.11 Consider an infinitely long rectilinear wire carrying a current I. By use of Eq. (5.20) to define the magnetic field of a current-carrying wire along with a closed cylindrical

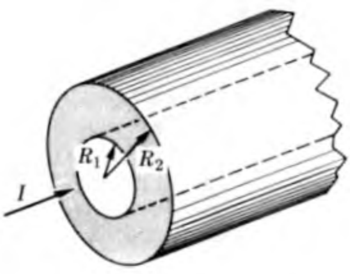


Figure 6-16

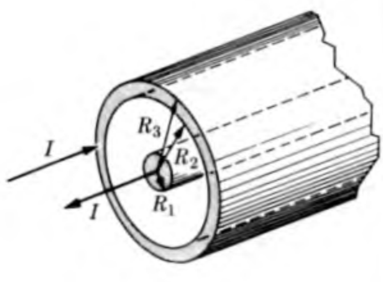


Figure 6-17

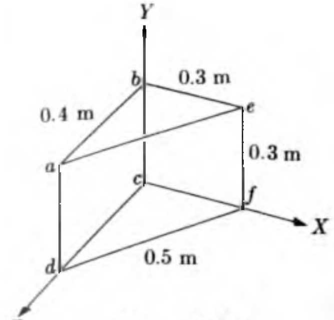


Figure 6-18

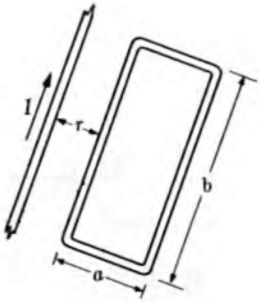


Figure 6-19

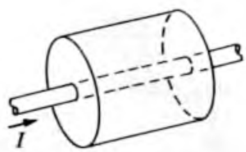


Figure 6-20

surface to enclose a portion of the wire (see Fig. 6-20), calculate the magnetic flux through the surface. (*Hint:* There are three surfaces to consider: two circular areas and the cylindrical surface.)

6.12 Show that for a uniform magnetic field div $\mathcal{B}=0$.

6.13 Show that the magnetic field given in Problem 6.7 results in div $\mathcal{B}=0$.

6.14 Introduce into Eq. (6.14) the value of $\mathcal{H}$ given by Eq. (6.19), and show that

$$\text{curl } \mathcal{B}=\mu_0(j_{\text{free}}+\text{curl } \mathfrak{M}).$$

(This result indicates that the effect of the magnetization of a medium is equivalent to the addition of a magnetization-current density, $j_{\mathfrak{M}}=\text{curl } \mathfrak{M}$, to the free-current density.]

CHALLENGING PROBLEMS

6.15 A slab of infinite length and infinite width has a thickness d. Point P_1 is a point inside the slab at $x=a$, and point P_2 is a point inside the slab at $x=-a$ as shown in Fig. 6-21. For parts (a) and (b), consider the slab to be non-conducting with a uniform charge per unit volume ρ as shown in Fig. 6-21a. (a) On a diagram identical to Fig. 6-21b, sketch vectors representing the electric field $\mathcal{E}$ at points P_1 and P_2. (b) Use Gauss's law and symmetry arguments to determine the magnitude of $\mathcal{E}$ at point P_1. For parts (c) and (d), consider the slab to be conducting and uncharged but with a uniform current density j directed out of the page as shown in Fig. 6-21c. (c) On a diagram identical to Fig. 6-21d, sketch vectors representing the magnetic field $\mathcal{B}$ at points P_1 and P_2. (d) Use Ampère's law and symmetry arguments to determine the magnitude of $\mathcal{B}$ at point P_1. (AP-C; 1979)

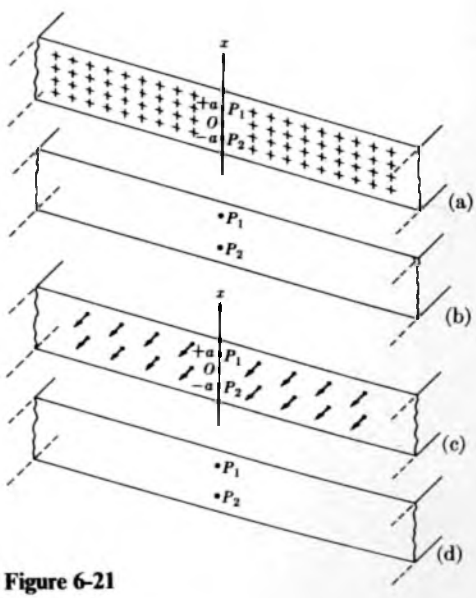


Figure 6-21

6.16 Using the components of the magnetic field of a magnetic dipole given in Eq. (5.26), verify that curl $\mathscr{B}=0$ and div $\mathscr{B}=0$.

6.17 Using the operator called *del* (∇),

$$\nabla = \boldsymbol{u}_x\left(\frac{\partial}{\partial x}\right)+\boldsymbol{u}_y\left(\frac{\partial}{\partial y}\right)+\boldsymbol{u}_z\left(\frac{\partial}{\partial z}\right),$$

show that the following identities hold

$$\text{div}\ \boldsymbol{A}=\nabla\times\boldsymbol{A}$$

$$\text{curl}\ \boldsymbol{A}=\nabla\times\boldsymbol{A},$$

and

$$\text{grad}\ V=\nabla V$$

where V is a scalar and A is a vector.

6.18 Using the operator ∇, rewrite the differential equations of the electromagnetic field that appear in Table 6-2.

6.19 Using the result of Problem 6.17 show that

$$\text{curl grad}\ V=\nabla\times(\nabla V)=0$$

and

$$\text{div curl}\ A=\nabla\cdot(\nabla\times A)=0.$$

Two important results derive from these identities. One result is that since for a static electric field $\mathscr{E}=-\text{grad}\ V$, then curl $\mathscr{E}=\nabla\times\mathscr{E}=0$; this result was stated in Eq. (6.15). The other is that since for the magnetic field div $\mathscr{B}=\nabla\cdot\mathscr{B}=0$, then there is a vector field $\mathscr{A}$ such that $\mathscr{B}=\nabla\times\mathscr{A}$. The vector field $\mathscr{A}$ is called the *vector potential* of the electromagnetic field.

6.20 Show that the vector potential of a uniform magnetic field $\mathscr{B}$ is $\mathscr{A}=\frac{1}{2}\mathscr{B}\times\mathbf{r}$. (*Hint*: Assume that $\mathscr{B}$ is along the Z-axis, obtain the rectangular components of $\mathscr{A}$, and then find $\nabla\times\mathscr{A}$.)

6.21 Write the operator $\nabla^2=\nabla\cdot\nabla$. Then show that Laplace's equation (2.7) and Poisson's equation (2.6) can be written as $\nabla^2 V=0$ and $\nabla^2 V=-\rho/\epsilon_0$, respectively.

CHAPTER SEVEN

THE ELECTRICAL STRUCTURE OF MATTER

7.1 Introduction

Until the discovery of the electron, the proton, and the neutron, a satisfactory model for the structure of atoms and molecules was not available. Understanding how atoms and molecules are composed progressed very rapidly once the motion of electrons around the positively charged nucleus of the atom was understood in terms of the electric and magnetic interactions just discussed. To a first approximation the electron motion about a nucleus can be analyzed with the methods of newtonian mechanics; however for a precise description of electron motion in atoms and molecules, quantum mechanics must be used. In this chapter we shall discuss the fundamental aspects of atomic structure in terms of newtonian mechanics and refer to quantum mechanics only when absolutely necessary for an understanding of certain aspects. Detailed descriptions using quantum mechanics will be found in Chapter 18 and in Volume III.

7.2 Electric Interactions in Atoms and Molecules

The student has been reminded on several occasions that matter is composed of charged particles. One manifestation of this composition is the frequently observed fact that bodies of certain substances can be electrified by rubbing them with cloth or fur. Many other laboratory experiments point to the fact that the basic constituents of all atoms are charged particles. For example when a metallic filament is heated, it emits electrons just as molecules are vaporized when a liquid is heated. This phenomenon is called *thermionic* emission.

Another interesting phenomenon related to the electric structure of matter is that of *electrolysis*. Suppose that an electric field $\mathscr{E}$ is produced (Fig. 7-1) in a molten salt (such as KHF_2) or in a solution containing an acid (such as HCl), a base (such as NaOH), or a salt (such as NaCl). The electric field can be produced by immersing in the solution two oppositely charged bars or plates called *electrodes*. Under the influence of the electric field electric charges flow; certain kinds of charged atoms move toward the positive electrode or *anode*, and others move toward the negative electrode or *cathode*. This phenomenon suggests that the molecules of the dissolved substance have separated (or *dissociated*) into two different kinds of charged parts, or *ions*. Some ions are positively charged and move in the direction of the electric field; others are negatively charged and move in the direction opposite to the electric field. For example in the case of NaCl, Na ions move toward the cathode and therefore are positive ions, called *cations*, while the Cl ions go to the anode and are negative ions, called *anions*. The dissociation may be written in the form

$$NaCl \rightarrow Na^+ + Cl^-.$$

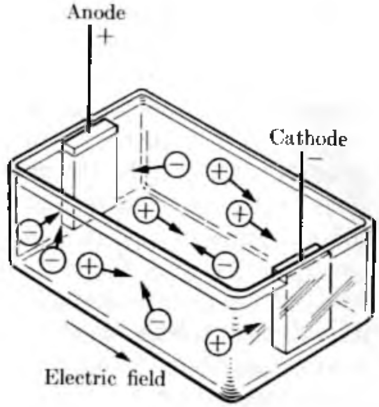


Fig. 7-1. Electrolysis. Ions move under the action of the electric field produced by the charged electrodes.

Since normal molecules of NaCl do not exhibit any obvious electrical charge, it may be assumed that they are composed of equal amounts of positive and negative charges. When the NaCl molecules dissociate, the charges are no longer balanced. One of the ions carries an excess of negative electricity; and the other part, an excess of positive electricity. Since it has been shown that all charges are multiples of a fundamental unit charge e, suppose that the positive ions carry a charge $+ve$; and the negative ions, a charge $-ve$ where v is an integer to be determined later. When the ions arrive at each electrode, they become neutralized by exchanging their charge with the charge available at the electrodes. Usually there follows a series of chemical reactions that are of no concern here; these reactions serve to identify the nature of the ions that move to each electrode.

After a certain time t, a number N of each kind of ion has gone to each electrode. The total charge Q transferred at each electrode is then, in absolute value, $Q=Nve$. If m is the mass of each molecule, the total mass M deposited at both electrodes is $M=Nm$. Dividing the first relation by the second, gives

$$\frac{Q}{M}=\frac{ve}{m}. \tag{7.1}$$

If N_A is *Avogadro's constant* (the number of molecules in one mole of any substance), the mass of one mole of the substance is $M_A=N_A m$. Therefore Eq. (7.1) can be written in the form

$$\frac{Q}{M}=\frac{ve}{m}=\frac{N_A ve}{N_A m}=\frac{Fv}{M_A}. \tag{7.2}$$

The quantity

$$F=N_A e \tag{7.3}$$

is a universal constant called the *Faraday constant*. It represents the charge of one mole of ions having $v=1$. The experimental value of the Faraday constant is found to be

$$F=9.6487\times 10^4 \text{ C mole}^{-1}. \tag{7.4}$$

From this value and the one previously found for e, Avogadro's constant is then

$$N_A = 6.0225 \times 10^{23} \text{ mole}^{-1} \tag{7.5}$$

in agreement with other calculations of this constant.

Equation (7.2) has been verified experimentally, and v has been found equal to the *chemical valence* of the ion concerned. That v is the chemical valence suggests that when two atoms bind together to make a molecule, they exchange the charge ve, and one becomes a positive ion and the other a negative ion. The resultant electric interaction between the two ions holds them together. A safe assumption is that electrons are the particles that are exchanged since they are much lighter than the protons and more easily moved. This picture of chemical binding, called *ionic* binding, must be considered as only a preliminary discussion and subject to further revision and criticism.

Gravitational forces were shown in Chapter 13 of Volume I to be too weak to produce the attraction necessary to bind two atoms together to make a molecule, or two molecules together to form a piece of matter, and that gravity was too small by a factor of 10^{35}. Let us now compare the order of magnitude of the electrical and gravitational forces. If the distances are the same, the strength of the electrical interaction is determined by the coupling constant $q_1 q_2/4\pi\epsilon_0$, and that of the gravitational interaction by $\gamma m_1 m_2$. Therefore

$$\frac{\text{Electrical interaction}}{\text{Gravitational interaction}} = \frac{q_1 q_2}{4\pi\epsilon_0 \gamma m_1 m_2}.$$

To obtain the order of magnitude, set $q_1 = q_2 = e$ and $m_1 = m_2 = m_p$ so that for two protons or two hydrogen ions

$$\frac{\text{Electrical interaction}}{\text{Gravitational interaction}} = \frac{e^2}{4\pi\epsilon_0 \gamma m_p^2} = 1.5 \times 10^{36}.$$

For the interaction between a proton and an electron ($m_1 = m_p$, $m_2 = m_e$), the ratio above is even larger: 2.76×10^{40}. Therefore we conclude that

> *the electrical interaction is of the order of magnitude required to produce the binding between atoms to form molecules, or the binding between electrons and protons to form atoms.*

The conclusion now is obvious: chemical processes (and in general the behavior of matter in bulk) are due to electrical interactions between atoms and molecules. A thorough understanding of the electrical structure of atoms and molecules is thus essential for explaining chemical processes and in general for explaining all the phenomena currently observed around us in both inert and living matter. The reader is reminded that the aim of physics is to understand the structure of the fundamental constituents of matter and to explain the behavior of matter in bulk in terms of their interactions.

Whenever electrically charged bodies are present, gravitational forces are in general negligible. Gravitational forces are important only in dealing with massive bodies with no net electrical charge or very small charge compared with their masses as is the case of planetary motion or the motion of bodies near the earth's surface.

7.3 Atomic Structure

Understanding atomic structure is one of the basic problems of physics. Some preliminary ideas will now be presented and a satisfactory model of the atoms will be developed. Atoms are normally electrically neutral since matter in bulk does not exhibit gross electrical forces. Therefore atoms must contain equal amounts of positive and negative electricity, or in other words equal numbers of protons and electrons. The equal number of protons and electrons is called the *atomic number* and is designated by Z. Therefore the atom consists of a positive charge $+Ze$ associated with the protons and an equal negative charge from the electrons.

The distribution of electrons and protons in an atom may be determined by experimentally probing the interior of the atom. In a *scattering* experiment a stream of fast charged particles such as hydrogen ions (that is, protons) or helium ions (called *alpha particles*) is sent against the atom, and the interactions produced are observed. Symmetry suggests that atoms may be considered spherical with a radius of the order of 10^{-10} m. As proved in Example 2.4, a charged sphere of radius a (Fig. 7-2) with the charge Q uniformly distributed throughout all its volume produces at all exterior points ($r > a$) an electric field given by

$$\mathcal{E} = \frac{Q}{4\pi\epsilon_0 r^2}, \quad r > a, \tag{7.6}$$

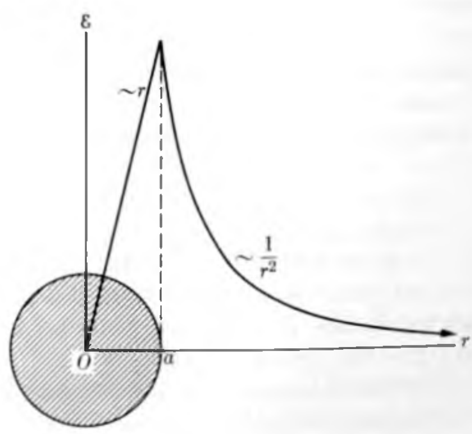


Fig. 7-2. Electric field of a charged sphere of radius a.

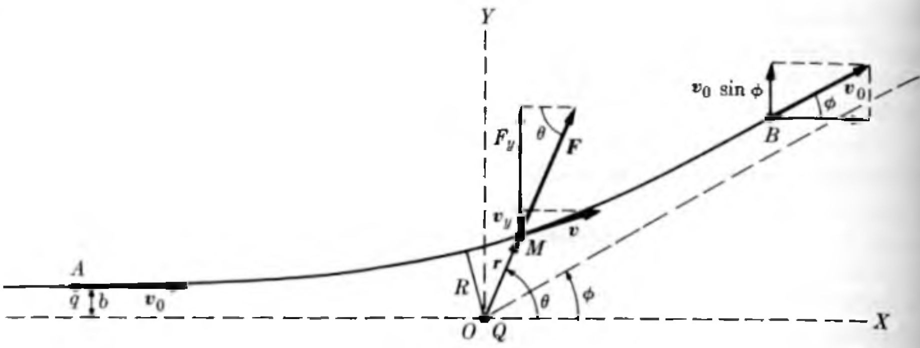


Figure 7-3

and at all interior points ($r<a$) an electric field given by

$$\mathscr{E}=\frac{Qr}{4\pi\epsilon_0 a^3}, \quad r<a. \tag{7.7}$$

The deflection experienced by the particles of charge q, mass m, and velocity v_0, approaching a uniformly charged sphere of radius a, but not passing through it, is computed in Example 7.1; the result is

$$\cot\tfrac{1}{2}\phi=\frac{4\pi\epsilon_0 m v_0^2}{Qq}\,b \tag{7.8}$$

where b is the *impact parameter*, defined as the perpendicular distance from the scattering center to the initial path of the incident particle (see Fig. 7-3). Scattering experiments with protons and α-particles show that many particles are deflected at large angles and even through 180° in some cases.

Assume a *nuclear model* of the atom (Fig. 7-4): all protons are clustered at the center of the atom in a very small region called the *nucleus*; the nucleus therefore carries a charge $+Ze$; and the electrons orbit around the nucleus as planets do around the sun. In this model, $Q=Ze$; set $q=\nu e$ for the bombarding particle ($\nu=1$ for protons, $\nu=2$ for α-particles); and, from Eq. (7.8) the impact parameter is related to the deflection angle by the expression

$$b=\frac{\nu Ze^2}{4\pi\epsilon_0 m v_0^2}\cot\tfrac{1}{2}\phi.$$

In the experimental arrangement several particles are directed against a very thin foil and the deflections are observed (Fig. 7-3). Since b cannot be controlled because aiming directly at a particular atom is impossible, a statistical analysis is made to interpret the experimental results.

Consider a thin metallic foil with thickness t and with n atoms of the scattering material per unit volume. If N particles per unit area of the foil impinge on the foil,

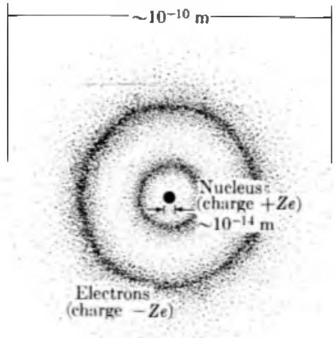


Fig. 7-4. Electron distribution in an atom.

some will pass close to a nucleus in the foil (small impact parameter) and thereby suffer a large deflection. Some will pass at a relatively large distance from the nuclei in the foil (large impact parameter) and suffer a small deflection. The result of a statistical analysis (see Example 7.2) shows that the number of particles dN deviated within the solid angle $d\Omega$ (corresponding to the scattering angles ϕ and $\phi+d\phi$ relative to the direction of incidence) is given by

$$\frac{dN}{d\Omega}=-\frac{Nnv^2Z^2e^4}{4(4\pi\epsilon_0)^2m^2v_0^4}\csc^4\tfrac{1}{2}\phi. \tag{7.9}$$

The sign is negative because dN represents the particles removed from the incident beam as a consequence of scattering, and this removal corresponds to a decrease in N.

The result predicted by Eq. (7.9) is that the particles scattered per unit solid angle must be distributed statistically according to a $\csc^4\frac{1}{2}\phi$ law. The verification of this prediction for all angles is thus an indirect proof that all positive charge is concentrated near the center of the atom. This proof was obtained by experiments performed for the first time during the period 1911–1913 by H. Geiger and E. Marsden under the direction of the British physicist Ernest Rutherford (1871–1937). These experiments were the foundation for the nuclear model of the atom; this model has been accepted since then as the correct one.

For each value of the impact parameter b, there is a distance R (Fig. 7-3) of closest approach; at this distance the bombarding particle is closest to the center. The minimum distance occurs for a head-on collision; that is when $b=0$. Then the particle is reflected directly back and suffers a deflection of 180°. This minimum value for the distance of closest approach can be obtained by equating the initial kinetic energy of the particle, $\frac{1}{2}mv_0^2$, to the potential energy, $vZe^2/4\pi\epsilon_0 R$, at the point at which the particle stops momentarily before it is sent back. Then

$$R=\frac{vZe^2}{4\pi\epsilon_0(\frac{1}{2}mv_0^2)}. \tag{7.10}$$

Calculation of this distance for different experimental conditions (see Example 7.3) indicates that this distance is of the order of 10^{-14} m for energies of the order of

10^{-13} J (or one MeV). This distance gives an upper limit for the radius of the atomic nucleus. Therefore the protons are concentrated in a region whose dimensions are of the order of 10^{-14} m. Because the atomic radius is of the order of 10^{-10} m, it is seen that the nucleus occupies only a small fraction of the atomic volume.

For very small values of the impact parameter and sufficiently high energy, when the incoming particle comes very close to the nucleus, we observe that the $\csc^4 \frac{1}{2}\phi$ law is not followed. This observation indicates the presence of other interactions, the *nuclear forces*. Analysis of the discrepancies from the pure coulomb scattering given by Eq. (7.9) gives valuable information about nuclear forces.

The simplest and lightest of all atoms are hydrogen atoms. A hydrogen atom is composed of one electron orbiting around a single proton. Then $Z = 1$, and the nucleus of the hydrogen atom is just one proton. Since the electron is subject to a central $1/r^2$ attractive force, the orbits should be ellipses with the proton at one focus for the same reason as that for planetary motion. Analysis of electron orbits, however, requires special techniques since some special features make electron orbits different from planetary orbits. These techniques rely on quantum mechanical principles.

For atoms heavier than hydrogen, the atomic mass is greater than the mass of the Z protons they contain. The difference may be attributed to the presence of *neutrons* in the nucleus. The total number of particles in a nucleus is called the *mass number* and is designated by A. Therefore an atom has Z electrons, Z protons, and $A-Z$ neutrons. Neutrons are apparently necessary to stabilize the nucleus. If protons were subject to their own electrical interaction alone, they would repel each other since they are all charged positively. That they stay together in a nucleus indicates that besides their electrical interaction other very strong interactions, corresponding to the so-called *nuclear forces*, counterbalance the electrical repulsion. The neutrons contribute to the nuclear force without adding electrical repulsion, and thus produce a stabilizing effect.

Since it is an electrical effect, the chemical behavior of an atom is determined by the atomic number Z. Thus each chemical element is composed of atoms having the same number Z. However for a given Z there may be several values of the mass number A. In other words to a given number of protons in a nucleus there may correspond different numbers of neutrons. Atoms having the same atomic number but different mass number are called *isotopes*. They all correspond to the same chemical element. Different isotopes of a chemical element are designated by the symbol of the chemical element (which also identifies the atomic number) along with a superscript to the left to indicate the mass number. For example hydrogen ($Z=1$) has three isotopes: ^{1}H, ^{2}H or deuterium, and ^{3}H or tritium. Similarly two of the most important isotopes of carbon ($Z=6$) are ^{12}C and ^{14}C. The isotope ^{12}C is the one used to define the atomic mass unit.

Example 7.1. Scattering of a charged particle by the Coulomb repulsion of another charged particle.

▼ Consider the deviation (or *scattering*) that a charged particle suffers when subjected to the repulsive force of another charged particle with a mass so much larger that the larger mass may be considered at rest during the interaction. (This problem is especially interesting because of its application to a number of different situations in atomic and nuclear physics.) For example when a proton, accelerated by a machine such as a cyclotron, passes near a nucleus of the target material, the proton is deflected (or scattered) under the action of the electrostatic repulsion of the nucleus.

Let O be the origin of a coordinate system in which a particle of charge Q may be considered at rest. From the point A a particle of charge q and mass m is propelled toward the origin with velocity v_0. The point A is at a large distance from O (Fig. 7-3). The distance b, called the *impact parameter*, is the perpendicular distance between the line of action of v_0 and a line (defined here as the X-axis) drawn through O parallel to the line of action. The moving particle will follow the path AMB, which is a branch of a hyperbola, whenever the force is repulsive and varies inversely as the square of the distance; for the Coulomb force

$$F=\left(\frac{1}{4\pi\epsilon_0}\right)\frac{qQ}{r^2}.$$

When the particle is at A, the magnitude of the particle's angular momentum relative to O is mv_0b. At any position, such as M, the angular momentum relative to O is given by $mr^2(d\theta/dt)$. Because the Coulomb force is central, the angular momentum must remain constant; that is,

$$mr^2\frac{d\theta}{dt}=mv_0b.$$

The equation of motion in the Y-direction is given by

$$m\frac{dv_y}{dt}=F_y=F\sin\theta=\left(\frac{1}{4\pi\epsilon_0}\right)\frac{qQ\sin\theta}{r^2}.$$

From the previous equation, r^2 may be eliminated to give an equation of motion in the Y-direction as

$$\frac{dv_y}{dt}=\left(\frac{1}{4\pi\epsilon_0}\right)\frac{qQ}{mv_0b}\sin\theta\frac{d\theta}{dt}.$$

This equation must be integrated from one extreme of the path to the other in order to find the total deflection of the particle. At A the value of v_y is zero because the initial motion is parallel to the X-axis and also $\theta=\pi$. At B $v_y=v_0\sin\phi$ and $\theta=\phi$. Note that at B the velocity is again v_0 because by symmetry the velocity lost as the particle approaches must be regained as it recedes. The principle of conservation of energy also verifies this assertion. Therefore

$$\int_0^{v_0\sin\phi}dv_y=\left(\frac{1}{4\pi\epsilon_0}\right)\frac{qQ}{mv_0b}\int_\pi^\phi\sin\theta\,d\theta$$

or

$$v_0\sin\phi=\left(\frac{1}{4\pi\epsilon_0}\right)\frac{qQ}{mv_0b}(1+\cos\phi).$$

From the trigonometric identity, $\cot\frac{1}{2}\phi=(1+\cos\phi)/\sin\phi$, the result above may be written

$$\cot\tfrac{1}{2}\phi=(4\pi\epsilon_0)\frac{mv_0^2}{qQ}b. \tag{7.11}$$

This relation gives the scattering angle ϕ in terms of the impact parameter b and may be rearranged to give Eq. (7.8) for the case $q=\nu e$ and $Q=Ze$.

Note that Eq. (7.11) is valid only for an inverse-square force. If the force depends on the distance in a different manner, the angle of scattering satisfies a different equation. Therefore scattering experiments may also be very useful to determine the law of force in interactions between particles.

In nuclear physics laboratories, scattering experiments are performed by accelerating electrons, protons, or other particles by means of a cyclotron, a Van de Graaff accelerator, or some other similar device. The observed angular distribution of the scattered particles is then used to understand the forces between the particles. ▲

Example 7.2. Derivation of Eq. (7.9) for Coulomb scattering.

▼ Let n be the number of atoms per unit volume of the scatterer. Then nt will be the number of scattering atoms in a thin foil of thickness t and unit area. The number of atoms in a ring of radius b and width db (and therefore of area $2\pi b\,db$) will be $(nt)(2\pi b\,db)$, as shown in Fig. 7-5. If N particles impinge on a unit area of the foil, the number of particles whose impact parameter is between b and $b+db$ is $dN=N(nt)(2\pi b\,db)$. However, differentiating the expression for the impact parameter b given by Eq. (7.11) yields $db=-\frac{1}{2}(qQ/4\pi\epsilon_0 mv_0^2)\csc^2\frac{1}{2}\phi\,d\phi$. Returning to the notation $q=\nu e$ and $Q=Ze$, dN may be written as

$$dN=-\frac{\pi N n\nu^2 Z^2 e^4 t}{(4\pi\epsilon_0)^2 m^2 v_0^4}\cot\tfrac{1}{2}\phi\csc^2\tfrac{1}{2}\phi\,d\phi. \tag{7.12}$$

For light target atoms the mass m of the particle is replaced by the reduced mass of the system of particle plus atom.

If two cones of angles ϕ and $\phi+d\phi$ are drawn around the nucleus (Fig. 7-6), all particles given by Eq. (7.12) will be deflected through the solid angle between the two conical surfaces. The solid angle is measured by dividing the shaded area by the square of its radius. The shaded area is $(2\pi r\sin\phi)(r\,d\phi)=2\pi r^2\sin\phi\,d\phi$. Therefore in view of the definition of a solid angle as area divided by radius squared, the solid angle is $d\Omega=2\pi\sin\phi\,d\phi=4\pi\sin\frac{1}{2}\phi\cos\frac{1}{2}\phi\,d\phi$ where the relation $\sin\phi=2\sin\frac{1}{2}\phi\cos\frac{1}{2}\phi$ has been used. The angular distribution is given by the number of particles scattered per unit solid angle. Then

$$\frac{dN}{d\Omega}=-\frac{Nn\nu^2Z^2e^4t}{4(4\pi\epsilon_0)^2m^2v_0^4}\csc^4\tfrac{1}{2}\phi,$$

which is Eq. (7.9).

Sometimes the results of scattering experiments are better expressed by using the concept of *cross section*. The differential cross section for a process is defined by the ratio of the fraction of particles scattered per unit solid angle to the total number of atoms in the scattering foil per unit

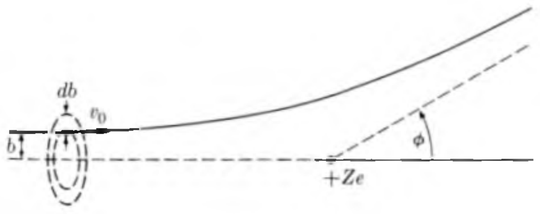


Fig. 7-5. Deflection of a positive ion by the Coulomb repulsion of the nucleus.

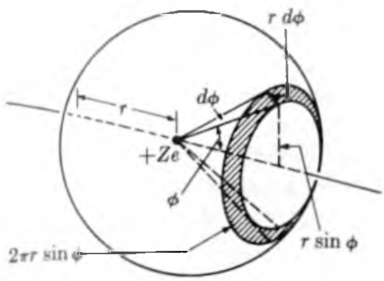


Figure 7-6

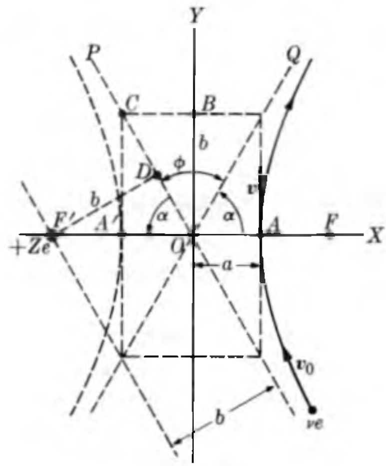


Figure 7-7

area; that is,

$$\sigma(\phi)=\frac{1}{Nnt}\left|\frac{dN}{d\Omega}\right|. \tag{7.13}$$

The vertical bars indicate that the absolute value of $dN/d\Omega$ is to be used. The quantity $\sigma(\phi)$ represents the probability that an incident particle will be deflected through an angle between ϕ and $\phi+d\phi$, and is expressed in units of area (m^2) since n is a density (m^{-3}) and t is a distance (m) (note that the units of N cancel out). Therefore substituting Eq. (7.9) in Eq. (7.13) gives the *differential cross section* for coulomb scattering:

$$\sigma(\phi)=\frac{\nu^2 Z^2 e^4}{4(4\pi\epsilon_0)^2 m^2 v_0^4}\csc^4 \tfrac{1}{2}\phi. \quad \blacktriangle \tag{7.14}$$

Example 7.3. The distance of closest approach of a particle having charge νe and directed with a velocity v_0 against an atom whose atomic number is Z.

▼ Figure 7-7 shows the geometry of the problem. According to Example 7.1, the particle describes a branch of a hyperbola with the nucleus $+Ze$ at the most distant focus F'. The distance of closest approach is $R=F'A$. Let $b=F'D$ be the impact parameter. We shall first prove that b is equal to the vertical axis OB of the hyperbola. The angle $\phi=POQ$ between the two asymptotes is the angle by which the particle has been deviated by the Coulomb repulsion of the nucleus. The distance $OA=OA'=a$ is the horizontal axis, and from the properties of the hyperbola it follows that $OF'=OC$. Therefore triangles $OF'D$ and OCA' are equal so that $b=F'D=CA'=OB$. From the geometry of the figure $OF'=b\csc\alpha$ and $OA=a=b\cot\alpha$. Therefore $R=F'A=b\,(\csc\alpha+\cot\alpha)$; but $2\alpha+\phi=\pi$ so that $\alpha=\frac{1}{2}\pi-\frac{1}{2}\phi$. Therefore

$$R=b(\sec\tfrac{1}{2}\phi+\tan\tfrac{1}{2}\phi)$$
$$=\frac{b(1+\csc\frac{1}{2}\phi)}{\cot\frac{1}{2}\phi}.$$

From result (7.8) with $Q = Ze$ and $q = \nu e$, the distance of closest approach is

$$R = \frac{\nu Z e^2}{4\pi\epsilon_0 (m v_0^2)} (1 + \csc \tfrac{1}{2}\phi). \tag{7.15}$$

The distance of closest approach is given in terms of the initial energy of the particle, $\frac{1}{2}mv_0^2$, and the angle of scattering ϕ. For a head-on collision the particle bounces back so that it is scattered through an angle equal to π. Therefore $\csc \frac{1}{2}\phi = 1$ and

$$R = \frac{\nu Z e^2}{4\pi\epsilon_0 (\frac{1}{2} m v_0^2)}.$$

For example substituting numerical values with $\nu = 1$, $Z = 6$ (corresponding to carbon), and $\frac{1}{2}mv_0^2 = 1.6 \times 10^{-13}$ J or 1 MeV gives the distance of closest approach as approximately 10^{-14} m. This is the order of magnitude quoted previously for nuclear dimensions. ▲

Example 7.4. Use of the principle of conservation of energy to compute the distance of closest approach of a charged particle directed against an atomic nucleus in terms of the angular momentum of the particle.

▼ If the charges are Ze for the nucleus and νe for the projectile, the potential energy of the system of projectile plus nucleus is

$$E_p = \frac{\nu Z e^2}{4\pi\epsilon_0 r}.$$

If the mass M of the nucleus is much greater than the mass m of the projectile, the total energy of the system relative to the nucleus is

$$E = \tfrac{1}{2}mv^2 + \frac{\nu Z e^2}{4\pi\epsilon_0 r}.$$

However if, for example, protons are directed against protons ($\nu = Z = 1$), the reduced mass, $\mu = \frac{1}{2}m_p$, must be used. When the particle is very far away, all its energy is kinetic and equal to $\frac{1}{2}mv_0^2$. Call v the particle's velocity at the point of closest approach A (Fig. 7-7) when $r = R$. The conservation of energy requires that

$$\tfrac{1}{2}mv^2 + \frac{\nu Z e^2}{4\pi\epsilon_0 R} = \tfrac{1}{2}mv_0^2.$$

At the point of closest approach A, the velocity is all transverse, and therefore the angular momentum is $L = mvR$. This relation may be used to eliminate the velocity v at A since L is a constant of the motion. Therefore

$$\frac{L^2}{2mR^2} + \frac{\nu Z e^2}{4\pi\epsilon_0 R} = \tfrac{1}{2}mv_0^2. \tag{7.16}$$

This equation of second degree in $1/R$ allows obtaining R in terms of the energy and the angular momentum of the particle. For a head-on collision, $L = 0$ and

$$R = \frac{\nu Z e^2}{4\pi\epsilon_0 (\frac{1}{2}mv_0^2)}. \tag{7.17}$$

which is in agreement with the result previously obtained in Example 7.3. Note that for a head-on collision, $v=0$ at the point of closest approach, and all the kinetic energy has been transformed into potential energy. ▲

7.4 Electron Energy Levels; The Bohr Theory

If the motion of an electron in an atom can be described by the laws of newtonian mechanics, the possible orbits of a single electron about a nucleus having a nuclear charge Ze may be predicted. The case $Z=1$ corresponds to the hydrogen atom; $Z=2$, to a singly ionized helium atom He^+ (i.e., a helium atom that has lost one electron); $Z=3$, to a doubly ionized lithium atom Li^{++} (i.e., a lithium atom that has lost two electrons); and so on.

The inverse-square electrical interaction involved in the motion of an electron around a nucleus is dynamically identical to the gravitational interaction involved in the motion of a planet around the sun: and therefore the results derived in the gravitational case (Chapter 13 of Volume I) are directly applicable if $\gamma mm'$ is replaced by $Ze^2/4\pi\epsilon_0$ in the corresponding expressions. For example the orbits will have to be ellipses (or circles) with the nucleus at one focus.

Consider two charges. q_1 and q_2, separated a distance r and moving with velocities v_1 and v_2. The electric potential energy of the two charges is $E_p=q_1q_2/4\pi\epsilon_0 r$ and their total energy is

$$E=\tfrac{1}{2}m_1v_1^2+\tfrac{1}{2}m_2v_2^2+\frac{q_1q_2}{4\pi\epsilon_0 r}.$$

In the case of two particles referred to their center of mass, the energy can be written in the form

$$E=\tfrac{1}{2}\mu v^2+\frac{q_1q_2}{4\pi\epsilon_0 r} \tag{7.18}$$

where μ is the reduced mass of the system of two particles and v is their relative velocity.

For an electron moving around a nucleus, $q_1=-e$ and $q_2=Ze$. Also since the mass of the nucleus is much greater than the mass of the electron, the reduced mass of the electron–nucleus system may be approximated by the electron mass m_e. Only in the lightest atoms, such as hydrogen and helium, can the effect of the reduced mass be detected. Therefore within this approximation the total energy of the atom is

$$E=\tfrac{1}{2}m_ev^2-\frac{Ze^2}{4\pi\epsilon_0 r}. \tag{7.19}$$

If we assume that the orbit is circular, the equation of motion of the electron is $m_ev^2/r=F$. If we recognize that the Coulomb force, $Ze^2/4\pi\epsilon_0 r^2$, is the centripetal force, the equation of motion of an electron in a circular orbit is

$$\frac{m_e v^2}{r}=\frac{Ze^2}{4\pi\epsilon_0 r^2} \quad \text{or} \quad m_e v^2=\frac{Ze^2}{4\pi\epsilon_0 r}. \tag{7.20}$$

When this result is inserted in Eq. (7.19), the total energy may be written

$$E=-\frac{Ze^2}{4\pi\epsilon_0(2r)}=-9\times 10^9\,\frac{Ze^2}{2r} \tag{7.21}$$

where the constant has been given in SI units. With this value, E is in J, r is in m, and e is in C.

Expression (7.21) for the energy of the electron–nucleus system will be revised later to take into account relativistic and magnetic effects. For the hydrogen atom ($Z=1$), E represents the energy required to separate the electron from the proton; that is, the ionization energy of the hydrogen atom. The experimental value for this ionization energy is 2.177×10^{-18} J or 13.6 eV, from which the radius of the electron orbit of a hydrogen atom is $r=0.53\times 10^{-10}$ m. This quantity is called the *Bohr radius* and is designated by a_0. That this radius is of the same order of magnitude as the estimate of atomic dimensions gives confidence in the nuclear model of the atom.

The next question to investigate is whether the energy of the electron in an atom can have any value or is restricted to certain values. The best way to answer this question is to perform experiments that excite an atom and thus increase the energy of the orbiting electron. The excitation can be produced by bombarding the atom with fast electrons; through inelastic collisions these electrons transfer part of their kinetic energy to the electron in the atom. The excitation may also be accomplished by letting the atom absorb energy from electromagnetic radiation.

These experiments show that *the energy of electronic motion is quantized*: that is, that the energy of the electrons can have only certain values $E_1, E_2, E_3, \ldots, E_n, \ldots$. The states corresponding to these energies are called *stationary states*. The state having the lowest possible energy is the *ground state*. To determine the energies of the stationary states is one of the tasks of quantum mechanics. Since the energy (in a classical sense) determines the "size" of the orbit, only certain regions of space are available for the electronic motion as was indicated schematically by the shaded region in Fig. 7-4.

In the case of atoms with only one electron, such as the hydrogen atom, or ions such as He^+, Li^{++}, etc., the experimental values of the energy levels E_n with $n=1, 2, 3, \ldots$, are inversely proportional to n^2; that is,

$$E_n=\frac{\text{constant}}{n^2}.$$

Since for $n=1$ the ground state energy is $E_1=-2.177\times 10^{-8}\,Z$ J, the general quantized energies for hydrogen are

$$E_n = -\frac{2.177\times10^{-18}}{n^2}Z\,\mathrm{J} = -\frac{13.598}{n^2}Z\,\mathrm{eV}. \tag{7.22}$$

Comparing Eq. (7.22) with Eq. (7.21) shows that the radius of the allowed electron orbits must be proportional to n^2; that is, the radius corresponding to the energy E_n is $r_n = \text{constant}\times n^2$.

If the energy and the radius of the electron orbits are limited to certain values (i.e., are quantized), one may suspect that other quantities may also be quantized. Consider for example the electron's angular momentum, defined by $L = m_e vr$. Then from Eq. (7.20)

$$L^2 = m_e^2 v^2 r^2 = m_e(m_e v^2)r^2 = m_e\left(\frac{Ze^2}{4\pi\epsilon_0}\right)r; \tag{7.23}$$

and since r is proportional to n^2, L must then be proportional to n, or $L = \text{const}\times n$. This quantization of angular momentum was suggested for the first time in 1913 by the Danish scientist Niels Bohr (1885–1962), who was the founder of the quantum theory of atomic orbits. Therefore it is assumed that *the angular momentum of the electronic motion is quantized*; that is, that the magnitude of the angular momentum of an electron may have only certain discrete values and that since angular momentum is a vector, it can point along only certain directions in space. This last property is sometimes referred to as *space quantization.*

Bohr wrote the relation between L and n in the form

$$L = n\left(\frac{h}{2\pi}\right) = n\hbar \tag{7.24}$$

where $h = 6.6256\times10^{-34}$ J s is a constant called *Planck's constant* and $\hbar = h/2\pi = 1.0545\times10^{-34}$ J s. Then Eq. (7.24) applied to Eq. (7.23) allows the radii of the electron orbits to be written as

$$r_n = \frac{n^2h^2\epsilon_0}{\pi Ze^2m_e} = \frac{n^2a_0}{Z}, \tag{7.25}$$

where a_0 is the *Bohr radius*, given by

$$a_0 = \frac{h^2\epsilon_0}{\pi e^2 m_e}.$$

The value of r_n must not be taken too literally because of the previous implication that electronic motion, unlike planetary motion, does not correspond to well-defined orbits. Instead, r_n can be considered as only an indication of the order of magnitude of the region in which the electron is likely to be found.

Substituting Eq. (7.25) into Eq. (7.21) gives the allowed energy values of the hydrogen atom:

$$E_n = -\frac{m_e e^2 Z^2}{8\epsilon_0^2 h^2 n^2}.$$

Comparison with the experimental values for E_n confirms the value of h as equal to that of Planck's constant, which the German physicist Max Planck (1858–1947) introduced in 1901 in connection with the radiation emitted from a cavity held at a constant temperature; the radiation is known as *blackbody radiation*.

Example 7.5. Correction to the energy produced in an orbiting electron in an atom because of relativistic effects.

▼ Whenever motion under an inverse-square law has been discussed previously, newtonian mechanics has been used and all relativistic effects neglected. This method is correct for almost all planetary motion; but for the electrons in an atom, ignoring relativity is not justified in many cases. The inner electrons in atoms move with velocities large enough so that the relativistic correction can be measured experimentally. The order of magnitude of the relativistic effect will now be estimated.

The total energy of a fast electron in an atom (subtracting its rest-mass energy) is, according to Eq. (11.15) of Volume I,*

$$E = c\sqrt{m_e^2c^2 + p^2} + (-eV) - m_ec^2$$

where $(-eV)$ is the electric potential energy of the system. If the momentum p is much smaller than m_ec, the radical may be expanded up to the second-order term; and the expansion results in

$$\begin{aligned} E &= \frac{1}{2m_e}p^2 - \frac{1}{8m_e^3c^2}p^4 + \cdots + (-eV) \\ &= \left[\frac{1}{2m_e}p^2 + (-eV)\right] - \frac{1}{8m_e^3c^2}p^4 + \cdots. \end{aligned}$$

The two terms inside the brackets give the nonrelativistic approximation for the energy, which for circular orbits is given by Eq. (7.21). Therefore the last term is the first-order relativistic correction to the total energy of the electron, and can be designated by ΔE_r. Thus

$$\Delta E_r = -\frac{1}{8m_e^3c^2}p^4 = -\frac{1}{2m_ec^2}\left(\frac{p^2}{2m_e}\right)\left(\frac{p^2}{2m_e}\right).$$

The two terms inside the parentheses correspond to the nonrelativistic kinetic energy of the electron. As a reasonable approximation, the first of the two bracketed terms, using Eq. (7.21) for the total energy, may be written as

$$\frac{p^2}{2m_e} \cong E - E_p = -\frac{Ze^2}{4\pi\epsilon_0(2r)} + \frac{Ze^2}{4\pi\epsilon_0 r} = \frac{Ze^2}{4\pi\epsilon_0(2r)} = -E.$$

The second term may be written $p^2/2m_e = \frac{1}{2}m_ev^2$. Therefore

$$\Delta E_r = -\frac{1}{2m_ec^2}(-E)(\tfrac{1}{2}m_ev^2) = \frac{1}{4}\frac{v^2}{c^2}E. \tag{7.26}$$

Thus the relativistic correction is of the order of $(v/c)^2$ times the energy of the electron. In the hydrogen atom for example, v/c is of the order of 10^{-2}; and therefore $\Delta E_r \sim 10^{-5}E$, or about

---

*See also the appendix.

0.001% of E, a quantity that can easily be detected in the laboratory with experimental techniques now in use. ▲

7.5 Magnetic Dipole Moment Caused by the Orbital Motion of a Charged Particle

Consider a charge q describing a closed orbit like that of an electron in an atom. For simplicity let the orbit be circular. If $\nu=\omega/2\pi$ is the frequency of the charge's motion, then the current at any point of the charge's path is $I=q\nu$ since ν gives the number of times per second that the charge q passes the same point of the orbit, and therefore $q\nu$ gives the total charge that passes through the point per unit time. The current is either in the same direction as the velocity or in the opposite direction, depending on whether q is positive or negative. Then, applying Eq. (5.9) gives the magnetic dipole moment of the orbiting charge:

$$M=(q\nu)(\pi r^2)=\left(\frac{q\omega}{2\pi}\right)(\pi r^2)=\tfrac{1}{2}q\omega r^2. \tag{7.27}$$

According to the rule previously given, the direction of the current depends on the sign of q as indicated in Fig. 7-8. If m is the mass of the particle, its orbital angular momentum L is

$$L=mvr=m\omega r^2. \tag{7.28}$$

Comparing Eqs. (7.27) and (7.28) shows that

$$M=\frac{q}{2m}L. \tag{7.29}$$

In vector form, Eq. (7.29) is

$$\boldsymbol{M}=\frac{q}{2m}\boldsymbol{L}. \tag{7.30}$$

Therefore $\boldsymbol{M}$ and $\boldsymbol{L}$ are either in the same direction or in opposite directions, depending on whether the charge q is positive or negative. For an electron $q=-e$ and $m=m_e$, and the result is

$$\boldsymbol{M}_e=-\frac{e}{2m_e}\boldsymbol{L}. \tag{7.31}$$

For a proton $q=+e$ and $m=m_p$, and thus we obtain

$$\boldsymbol{M}_p=\frac{e}{2m_p}\boldsymbol{L}. \tag{7.32}$$

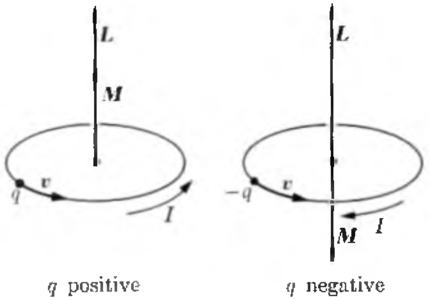


Fig. 7-8. Vector relation between the magnetic dipole moment and the angular momentum of an orbiting charge.

If the charged particle can be assumed also to rotate about a diameter in the same way that the earth spins about its NS axis, the particle will have, in addition to its orbital angular momentum $\boldsymbol{L}$, some internal angular momentum $\boldsymbol{S}$, called *spin*. Associated with the spin $\boldsymbol{S}$ will be a magnetic dipole moment since each volume element of the rotating charged particle behaves in the same way as the charge q in Fig. 7-8. However, the relation between the magnetic dipole moment and the spin is not the same as the relation of Eq. (7.30) because the coefficient by which one has to multiply the spin angular momentum $\boldsymbol{S}$ to obtain the corresponding magnetic moment depends on the internal structure of the particle. Consequently it is useful to write the magnetic dipole moment caused by the spin in the form

$$\boldsymbol{M}_S = \gamma \frac{e}{2m} \boldsymbol{S} \tag{7.33}$$

where the coefficient γ, called the *gyromagnetic ratio*, depends on the structure of the particle and the sign of its charge. Combining Eqs. (7.30) and (7.33) gives the total magnetic dipole moment of an orbiting and spinning particle carrying a charge $\pm e$:

$$\boldsymbol{M} = \frac{e}{2m}(\pm \boldsymbol{L} + \gamma \boldsymbol{S}). \tag{7.34}$$

The plus (minus) sign before $\boldsymbol{L}$ corresponds to a positively (negatively) charged particle. Although the neutron has no net electric charge and therefore no *orbital* magnetic dipole moment, as given by (7.30), the neutron does have a *spin* magnetic dipole moment, which is opposite to the spin $\boldsymbol{S}$. The total magnetic dipole moment of the neutron is given not by Eq. (7.34), but by Eq. (7.33). The nonvanishing value of $\boldsymbol{M}_S$ suggests some complex internal structure of the neutron. Similarly that the magnitude of γ for the proton is different from the magnitude of γ for the electron indicates that the internal structure of the proton is different from that of the electron.

The experimental values of γ for the electron, the proton, and the neutron are given in Table 7-1.

Table 7-1. Gyromagnetic Ratios

Particle	γ
Electron	-2.0024
Proton	5.5851
Neutron	-3.8256

7.6 Torque and Energy of a Charged Particle Moving in a Magnetic Field; Space Quantization

Suppose that an orbiting particle without spin is placed in a uniform magnetic field (Fig. 7-9). Using Eqs. (5.11) and (7.30) gives the torque exerted on the particle:

$$\boldsymbol{\tau} = \frac{q}{2m} \boldsymbol{L} \times \boldsymbol{\mathscr{B}} = -\frac{q}{2m} \boldsymbol{\mathscr{B}} \times \boldsymbol{L} \tag{7.35}$$

in a direction perpendicular to $\boldsymbol{L}$ and $\boldsymbol{\mathscr{B}}$. This torque tends to change the orbital angular momentum $\boldsymbol{L}$ of the particle according to the relation $d\boldsymbol{L}/dt = \boldsymbol{\tau}$. Using the *Larmor frequency* $\boldsymbol{\Omega} = -(q/2m)\boldsymbol{\mathscr{B}}$, which is one-half the cyclotron frequency given in Eq. (4.7), the torque given in Eq. (7.35) is

$$\boldsymbol{\tau} = \boldsymbol{\Omega} \times \boldsymbol{L}. \tag{7.36}$$

This equation is similar to that for gyroscopic motion so that the plane of motion of the charged particle may be expected to precess about the direction of the field. The precession of a gyroscope is due to the torque produced by the gravitational interaction. The precession in the case of an orbiting charge is due to the torque produced by the magnetic interaction. The angular momentum $\boldsymbol{L}$ precesses around $\boldsymbol{\mathscr{B}}$ and produces a rotation of the orbit of the particle. In Fig. 7-10 the direction of $\boldsymbol{\Omega}$ and the sense of precession for a positive and a negative charge have been indicated.

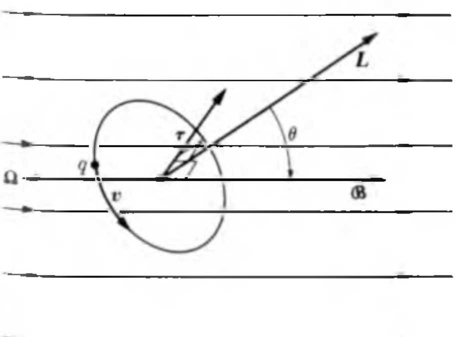


Fig. 7-9. The magnetic torque τ on a moving charged particle is perpendicular to the angular momentum $\boldsymbol{L}$ of the particle and the magnetic field $\boldsymbol{\mathscr{B}}$.

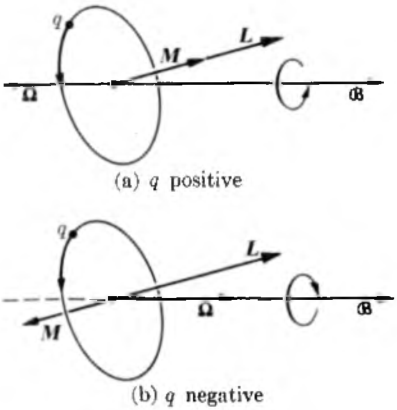


Fig. 7-10. Precessional motion of the angular momentum of a charged particle around the magnetic field.

The energy of an orbiting charged particle in a magnetic field is found by combining Eqs. (5.12) and (7.30) and results in

$$E_p = -\frac{q}{2m}\boldsymbol{L}\cdot\boldsymbol{\mathscr{B}} = \boldsymbol{\Omega}\cdot\boldsymbol{L}. \tag{7.37}$$

If the particle has spin as well, Eq. (7.34) for the magnetic dipole moment should be used, and the expression becomes

$$E_p = -\frac{e}{2m}(\pm\boldsymbol{L}+\gamma\boldsymbol{S})\cdot\boldsymbol{\mathscr{B}}. \tag{7.38}$$

These results are very important to understanding the behavior of an atom or a molecule in an external magnetic field, a subject of interest from both the theoretical and the practical points of view.

For example when an atom is placed in an external magnetic field (taken to be along the Z=axis), the motion of the electrons is disturbed and the energy is changed according to Eq. (7.38). When this theoretical value of E_p is compared with the experimental results, it is found that the Z-components of the orbital and spin angular momenta are quantized. That is, L_z and S_z can attain only certain values, which are expressed in the form

$$L_z = m_l\hbar, \qquad S_z = m_s\hbar. \tag{7.39}$$

According to experiment and quantum mechanics, the possible values of m_l are 0, ± 1, ± 2, ± 3, ..., and m_s can attain only two values, $+\frac{1}{2}$ or $-\frac{1}{2}$. The number m_l is called the *magnetic quantum number* of the electron and m_s is the *spin quantum number*. A similar result is obtained for protons and neutrons. For that reason it is said that *the electron, the proton, and the neutron have spin* $\frac{1}{2}$.

In Section 7.4 it was stated that *the angular momentum L is also quantized in units of* $n\hbar$; however, an analysis more detailed than that of Section 7.4 and using the methods of quantum mechanics shows that the allowed values of the angular momentum of an electron are given by

$$L=\sqrt{l(l+1)}\hbar \tag{7.40}$$

where $l=0, 1, 2, 3, \ldots$ is a positive integer and is called the *angular momentum quantum number*. Since L_z cannot be larger than L, the values of m_l cannot exceed l; that is,

$$m_l=0, \pm 1, \pm 2, \ldots, \pm(l-1), \pm l, \tag{7.41}$$

or a total of $2l+1$ different values or orientations of $\mathbf{L}$. For $l=0$, only $m_l=0$ is possible. For $l=1$, the values of m_l may be 0, ± 1, and so on.

Also it is shown in quantum mechanics that for the energy level E_n, the allowed values of the angular momentum quantum number are

$$l=0, 1, 2, \ldots (n-1), \tag{7.42}$$

which gives a total of n different values of l. States with $l=0, 1, 2, 3, \ldots$ are designated s, p, d, f, ..., respectively. In the same atom all electrons having both the same energy and the same angular momentum (that is, having the same n and l values, such as all 3p electrons in an atom) constitute an *electron shell* of the atom. Each shell of course contain electrons with spin up and spin down.

That for a given value of L only certain values of L_z are possible implies that L can attain only certain directions in space (Fig. 7-11a). This is what is meant by space quantization. In the case of spin since m_s has only two possible values $(\pm\frac{1}{2})$, we conclude that S can attain in space only two directions relative to the Z-axis; these directions are usually called up (↑) and down (↓). The allowed orientation of the spin vector is shown in Fig. 7-11b.

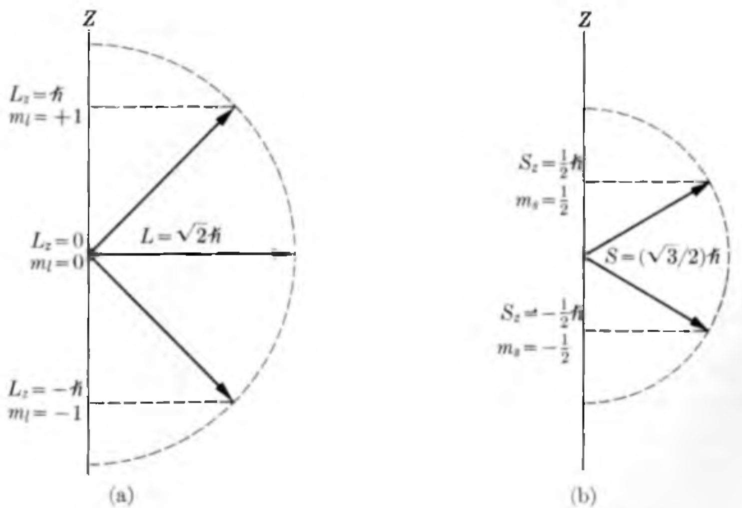


Fig. 7-11. Possible orientations of (a) the angular momentum corresponding to $l=1$, $L=\sqrt{2}\hbar$, and (b) the spin $s=\frac{1}{2}$, $S=(\sqrt{3/2})\hbar$.

For a given value of n the number of different combinations of l and m_l are n^2; if the two possible values of m_s are included, the combinations of the quantum numbers l, m_l, m_s is $2n^2$. Thus for $n=1, 2, 3, 4, \ldots$ the number of different angular momentum states are 2, 8, 18, 32, ... This is of considerable relevance when atomic structures are analyzed since it has been found that no two electrons in the same energy state in an atom can have the same set of quantum numbers l, m_l, and m_s. Thus $2n^2$ gives the maximum number of electrons that can exist in the energy state E_n.

When Eqs. (7.39) are introduced in Eq. (7.38) with the direction of the magnetic field chosen as the Z-axis so that $\boldsymbol{L}\cdot\mathscr{B}=L_z\mathscr{B}$ and $\boldsymbol{S}\cdot\mathscr{B}=S_z\mathscr{B}$, we find that the energy of an electron in a magnetic field has the possible values

$$E_p=-\frac{e\hbar}{2m_e}\mathscr{B}(m_l+2m_s) \tag{7.43}$$

where 2 has been substituted for the gyromagnetic ratio of the electron. This result means that the energy levels for a given energy E_n are split into several levels spaced apart by the amount $(e\hbar/2m_e)$ when the atom is placed in a magnetic field. This result is known as the *Zeeman effect*, after the Dutch physicist Peter Zeeman (1865–1943), who discovered the phenomenon experimentally.

Example 7.6. Magnetic interaction between an orbiting electron and the nucleus in an atom.

▼ Consider an electron having a charge of $-e$ and revolving with velocity $\boldsymbol{v}$ around a nucleus whose charge is Ze. The electron's path relative to the proton is the solid curve of Fig. 7-12, which for simplicity is assumed to be a circle. However if the motion is referred to a frame of reference attached to the electron, the electron will be at rest; and the proton will appear to be describing the broken path, also a circle, with velocity $-\boldsymbol{v}$. If we neglect the electron's acceleration (the student should be able to compute it and judge the reasonableness of this assumption), this new frame may be considered inertial. Thus relative to the electron, the nucleus produces an electric field given nonrelativistically by $\mathscr{E}=(Ze/4\pi\epsilon_0 r^2)\boldsymbol{u}_r$ and a magnetic field related to $\mathscr{E}$ by Eq. (4.17) with $\boldsymbol{v}$ replaced by $-\boldsymbol{v}$. That is,

$$\begin{aligned}\mathscr{B}&=\frac{1}{c^2}(-\boldsymbol{v})\times\mathscr{E}=\frac{1}{c^2}\mathscr{E}\times\boldsymbol{v}\\&=\frac{Ze}{4\pi\epsilon_0 c^2 r^2}\boldsymbol{u}_r\times\boldsymbol{v}.\end{aligned}$$

However the angular momentum of the electron relative to the nucleus is $\boldsymbol{L}=m\boldsymbol{r}\times\boldsymbol{v}=mr\boldsymbol{u}_r\times\boldsymbol{v}$. Thus, $\mathscr{B}$ and $\boldsymbol{L}$ are related by

$$\mathscr{B}=\frac{Ze}{4\pi\epsilon_0 c^2 mr^3}\boldsymbol{L}.$$

Therefore the magnetic field produced by the relative motion of the nucleus is proportional and parallel to the angular momentum of the electron as indicated in the figure.

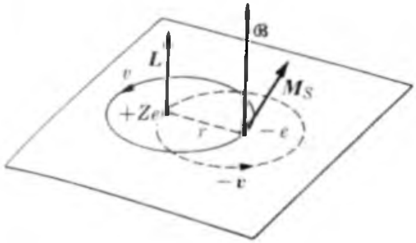


Fig. 7-12. Spin-orbit interaction of an electron orbiting about a positive nucleus.

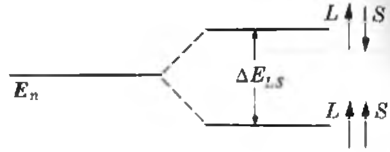


Figure 7-13

Because $\mathscr{B}$ is a magnetic field referred to a frame in which the electron is at rest. $\mathscr{B}$ does not produce any interaction with the orbital motion of the electron; but the electron has a magnetic dipole moment M_S caused by its spin. Thus the magnetic interaction of the electron with the nuclear magnetic field from Eqs. (5.12) and (7.33) is

$$E_p = -M_S \cdot \mathscr{B} = -\left(\gamma \frac{e}{2m} S\right) \cdot \left(\frac{Ze}{4\pi\epsilon_0 c^2 m r^3} L\right) = -\frac{\gamma Z e^2}{8\pi\epsilon_0 c^2 m^2 r^3} S \cdot L.$$

The most important aspect of this result is that the magnetic interaction depends on the relative orientation of the spin S and the orbital angular momentum L of the electron. For that reason this interaction is called *spin-orbit interaction* and is often designated by E_{SL}. A more detailed, relativistic calculation indicates that the value of E_{SL} is one-half the value obtained above.

Next we shall estimate its order of magnitude. Recall from Table 7-1 that, for the electron, γ is approximately -2. Also from Eq. (7.21) the energy of the electron in a circular orbit is to the zeroth order of approximation $E = -Ze^2/4\pi\epsilon_0(2r)$. Thus with E_{SL} corrected by the factor of one-half mentioned above, the spin-orbit interaction energy is

$$E_{SL} \approx \frac{E}{c^2 m^2 r^2} S \cdot L.$$

However L has a magnitude mrv, and S is of the same order of magnitude as L. Thus $S \cdot L$ is approximately $(mrv)^2$. With these substitutions the energy may be approximated by

$$E_{SL} \approx \frac{v^2}{c^2} |E|.$$

Comparing this value with the result of Example 7.5 shows that the spin-orbit interaction of an orbiting electron is of the same order of magnitude as the relativistic correction to the energy. However, the spin-orbit interaction has the peculiarity of showing a distinct directional effect because of the $S \cdot L$ factor, which depends on the relative orientation of L and S.

A careful analysis of experimental evidence of the energy levels of an electron in an atom shows that S can have only two orientations relative to L, either parallel or antiparallel, in agreement with the earlier discussion. Thus the spin-orbit interaction breaks each energy level into pairs (or doublets) of closely spaced energy levels.

The details of atomic structure and the consequences they entail for the structure and behavior of materials are refined with quantum mechanical details; these are discussed in Volume III. ▲

Problems

7.1 (a) Calculate the mass of copper (bivalent) deposited on an electrode by a current of 2.0 A during one hour. (b) How many atoms of copper have been deposited?

7.2 One mole of sodium is deposited on the cathode of an electrolytic cell. (a) What electric current and (b) what charge passed through the cell if the deposit took one day?

7.3 A proton produced in a 1-MeV Van de Graaff accelerator is sent against a gold foil. Calculate the distance of closest approach (a) for a head-on collision, and for collisions with impact parameters of (b) 10^{-15} m and (c) 10^{-14} m. (d) What is the deflection of the proton in each case?

7.4 An alpha particle with a kinetic energy of 4 MeV is directed straight toward the nucleus of a mercury atom. (a) Find the distance of closest approach of the alpha particle to the nucleus. (b) Compare the result with the nuclear radius, $\sim 10^{14}$ m.

7.5 When they investigated low-atomic-number nuclei with 4-MeV alpha particles, Geiger and Marsden observed deviations from the predicted Coulomb scattering. Considering that nuclei have a radius of approximately 10^{-14} m (this is the region in which the nuclear interaction has an effect), calculate the atomic number at which deviations from Rutherford scattering for head-on collisions will be observed with 4-MeV alphas.

7.6 In a hydrogen atom in its state of lowest energy (also called the ground state) the electron moves around the proton in what can be described as a circular orbit of radius 0.53×10^{-10} m. Compute (a) the potential energy, (b) the kinetic energy, (c) the total energy, and (d) the frequency of the motion. (For comparison, the frequency of the radiation emitted by the hydrogen atom is of the order of 10^{15} Hz.)

7.7 (a) Using the virial theorem for one particle, determine the energy of an electron (charge $-e$) revolving around a nucleus of charge $+Ze$ at a distance r. (b) Apply your result to a hydrogen atom ($r \sim 0.53 \times 10^{-10}$ m) and compare with the result obtained in (c) of Problem 7.6.

7.8 If the average lifetime of an excited state of hydrogen is of the order of 10^{-8} s, estimate how many orbits an electron makes (a) when it is in the state $n=2$ and (b) when it is in the state $n=15$, before it suffers a transition to state $n=1$. (c) Compare these numbers with the number of orbits the earth has made around the sun in its approximately 2×10^9 years of existence.

7.9 It is customary to write the energy of the stationary state of atoms with one electron in the form $E_n = -RZ^2hc/n^2$ where R is called the *Rydberg constant*. Using the expression given in Section 7.4 for E_n, show that R is equal to 1.0974×10^7 m^{-1}.

7.10 (a) Compute the energies of the first four stationary states of H and He$^+$. (b) In each case find the energy required to raise the system from the ground state to the first excited state. (c) Represent the energies on a vertical scale by properly spaced horizontal lines. Note that some energies coincide. (d) Can you derive a general rule?

7.11 Using the result of Problem 7.6, (a) estimate the velocity of an electron in a hydrogen atom in its ground state and (b) check the calculations made at the end of Example 7.5.

7.12 (a) Calculate the angular velocity of precession of a spinning electron in a magnetic field of 0.5 T. (b) Calculate the same quantity for a proton in the same field if a proton spins with the same angular momentum as that of an electron. (*Hint:* Use the γ values given in Table 7-1.)

7.13 (a) Compute the magnetic dipole moment of the electron in a hydrogen atom orbiting in a circular path at a distance of 0.53×10^{-10} m from the proton. (b) Compute the angular precessional velocity of the electron if it is in a magnetic field of 10^{-5} T.

7.14 Compute the gyromagnetic ratio γ for a rotating disk of radius R carrying a charge q uniformly distributed over the surface of the disk. 7.15(a) What is the magnitude of the angular momentum of an electron if $l=2$? (b) What angles, relative to a magnetic field parallel to the Z-axis, may the angular momentum vector have when $l=2$?

7.16 (a) Into how many levels does a magnetic field split the $n=3$ level of the hydrogen atom? (b) What is the magnetic energy difference between these levels when the field is 4.0 T?

7.17 Compare the energy of the magnetic splitting of the $n=3$ level in atomic hydrogen with the energy difference between the electronic energies for $n=2$ and $n=3$ when the magnetic splitting is due to a 4-T magnetic field.

CHALLENGING PROBLEMS

7.18 The electron in a hydrogen atom may be assumed to be "spread" over all space with a density $\rho = Ce^{-2r/a_0}$ where $a_0 = 0.53 \times 10^{-10}$ m. (a) Find the constant C such that the total charge is $-e$. (b) Determine the total charge within a sphere of radius a_0, which corresponds to the orbit radius of the electron. (c) Obtain the electric field as a function of r. (d) At what distance does the electric field differ from $-e/4\pi\epsilon_0 r^2$ by 1%? (*Hint:* For part (a), divide the space into spherical shells, each of volume $4\pi r^2\,dr$.)

7.19 Protons accelerated by a voltage of 8×10^5 V fall on a gold foil ($Z=79$). Compute the differential cross section for Coulomb scattering, in intervals of 20°, for ϕ between 20° and 180°. Make a polar graph of $\sigma(\phi)$. (*Note:* Equation (7.14) becomes infinite for $\phi=0$ because we have assumed that the scattering nucleus is a point charge. When the finite size of the nucleus is taken into account, this infinity disappears.)

7.20 The average separation of protons within an atomic nucleus is of the order of 10^{-15} m. Estimate in J and in MeV the order of magnitude of the electric potential energy of two protons in a nucleus.

7.21 If one assumes that all protons in an atomic nucleus of radius R are uniformly distributed, the internal electric potential energy can be computed by

$$\frac{\frac{3}{5}Z(Z-1)e^2}{4\pi\epsilon_0 R}$$

(see Problem 1.72). The nuclear radius can in turn be computed by $R = 1.2 \times 10^{-15} A^{1/3}$ m. Write expressions giving the nuclear electric potential energy in J and in MeV as a function of Z and A.

7.22 Using the results of Problem 7.21, compute the total electric potential energy and the energy per proton for the following nucleus: (a) ^{16}O ($Z=8$), (b) ^{40}Ca ($Z=20$), (c) ^{91}Zr ($Z=40$), (d) ^{144}Nd ($Z=60$), (e) ^{200}Hg ($Z=80$), and (f) ^{238}U ($Z=92$). (g) What do your results tell you about the effect of the electric interaction between protons on the stability of the nucleus? (h) Using your data, plot the potential energy against the mass number.

7.23 (a) Repeat Problem 7.14 for a sphere uniformly charged throughout its volume. (*Hint:* Divide the sphere into disks perpendicular to the axis of rotation.) (b) From the result of this problem, what do you conclude about the electron's structure?

7.24 An electron changes its value of m_s from $+\frac{1}{2}$ to $-\frac{1}{2}$. (a) What is the change in angular momentum of the electron? (b) If this change occurs in a magnetic field of 2 T, what is the change in the electron's energy?

CHAPTER EIGHT

THE TIME-DEPENDENT ELECTROMAGNETIC FIELD

8.1 Introduction

In previous chapters, the electric and magnetic fields were considered to be time independent, or in other words static. In this chapter fields that are time dependent will be studied; that is, at a given point in space the fields may change with time. New relations are found to exist in this case. In Section 4.6 the close relationship between the electric and magnetic parts of an electromagnetic field was investigated, especially with respect to the electromagnetic field's transformation properties, which are required by the principle of relativity. In this chapter it will be seen that a varying magnetic field requires the presence of an electric field, that conversely a varying electric field requires a magnetic field, and that this relationship is required by the principle of relativity. The laws describing these two situations are called the *Faraday-Henry law* and the *Ampère-Maxwell law*.

8.2 The Faraday – Henry Law

One of the many electromagnetic phenomena familiar to the student is *electromagnetic induction*, which was discovered independently and almost simultaneously around 1830 by Michael Faraday (1791–1867) and Joseph Henry (1797–1878). Electromagnetic induction is the working principle of the electric generator, the transformer, and many other devices in daily use. Suppose that an electric conductor that forms a closed path is placed in a region in which a magnetic field exists. If the magnetic flux Φ_m through the closed path *varies with time*, a current may be observed in the circuit while the flux is varying. The presence of an electric current indicates the existence or induction of an emf in the circuit. Measurement of this induced emf shows that it depends on the *time rate* of change of the magnetic flux $d\Phi_m/dt$. For example if a magnet is placed near a closed conductor, an emf appears in the circuit when the magnet (or the circuit) is moved in such a way that the magnetic flux through the circuit changes. The magnitude of the induced emf depends on whether the magnet (or circuit) is moved rapidly or slowly. The greater the rate of change of the flux, the larger the induced emf. The direction in which the induced emf acts depends on whether the magnetic field is increasing or decreasing.

To be more precise, refer to Fig. 8-1, in which the curve L has been oriented in the same sense as the fingers of the right hand when the thumb points in the direction of the magnetic field $\mathscr{B}$. When the magnetic flux increases (that is, $d\Phi_m/dt$ is positive), the induced emf V acts in the negative sense; when the magnetic flux decreases (that is $d\Phi_m/dt$ is negative), V acts in the positive sense. Thus the sign of the induced emf V is always opposite to that of $d\Phi_m/dt$. More detailed measurement reveals that

$\overline{\partial t}$ ∂t

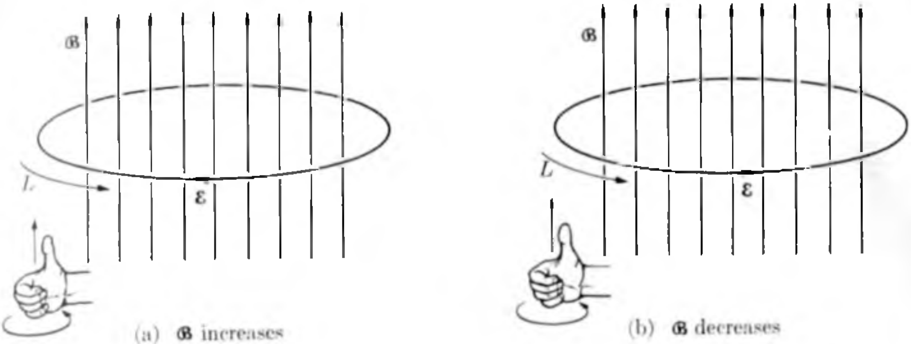


Fig. 8-1. Electric field produced by a time-dependent magnetic field; (a) $d\Phi_m/dt$ positive, V negative, (b) $d\Phi_m/dt$ negative, V positive.

the value of the induced emf when expressed in volts is equal to the time rate of change of the magnetic flux when it is expressed in Wb s^{-1}. That is,

$$V = -\frac{d\Phi_m}{dt}, \tag{8.1}$$

which expresses the *Faraday-Henry law* of electromagnetic induction:

> *in a varying magnetic field an emf is induced in any closed circuit and is equal to the negative of the time rate of change of the magnetic flux through the circuit.*

The negative sign in Eq. (8.1) may be explained in terms of the conservation of energy. In fact if the sign of the induced emf is the same as that of $d\Phi_m/dt$, the magnetic field produced by the current generated by V tends to change Φ_m in the same sense and thereby contributes to an increased value of V, and so on. Thus a small change in Φ_m would initiate a continuous change so that a small amount of energy used to change Φ_m initially would give rise to a large change in the magnetic energy of a system.

In Fig. 8-2, if the area surrounded by L is divided into infinitesimal area elements, each oriented according to the right-hand rule, the magnetic flux through L is $\Phi_m = \int_S \mathscr{B}\cdot u_N\, dS$ according to Section 6.4. Also the emf V implies the existence of an electric field $\mathscr{E}$ such that $V = \oint_L \mathscr{E}\cdot dl$ according to Eq. (3.15). Thus Eq. (8.1) may be written in the alternate form

$$\oint_L \mathscr{E}\cdot dl = -\frac{d}{dt}\int_S \mathscr{B}\cdot u_N\, dS. \tag{8.2}$$

The path L need not coincide with an electric conductor such as a closed wire; instead consider a region of space in which a magnetic field, varying with time, exists. Then Eq. (8.2) is equivalent to saying

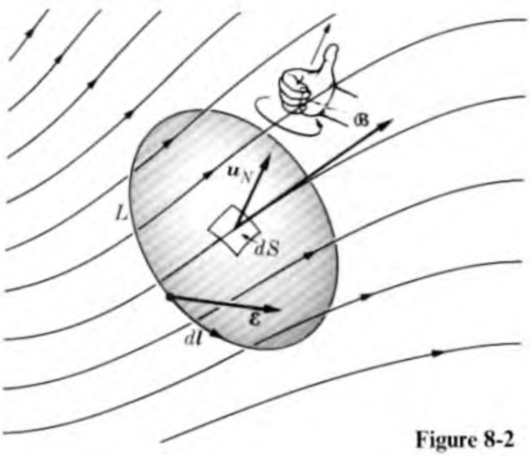


Figure 8-2

> *a time-dependent magnetic field implies the existence of an electric field such that the circulation of the electric field along an arbitrarily closed path is equal to the negative of the time rate of change of the magnetic flux through a surface bounded by the path.*

This way of stating the Faraday-Henry law of electromagnetic induction gives a deeper insight into the physical content of the phenomenon of electromagnetic induction; that is, an electric field *must* exist whenever a magnetic field is changing with time, the two fields being related by Eq. (8.2). The electric field can be determined by measuring the force on a charge at rest in the region in which the magnetic field is varying. This experiment may be carried out and thus confirm an interpretation such as Eq. (8.2).

Example 8.1 Computation of the emf induced in a simple circuit.

▼ Consider a plane circuit composed of N turns, each of area S, placed perpendicular to an alternating uniform magnetic field that varies with time. The equation of the field is $\mathscr{B}=\mathscr{B}_0 \sin \omega t$. The magnetic flux through one turn of the circuit is $\Phi_m=S\mathscr{B}=S\mathscr{B}_0 \sin \omega t$, and the total flux through the N turns is

$$\Phi_m=NS\mathscr{B}_0 \sin \omega t.$$

Therefore applying Eq. (8.1) gives for the induced emf:

$$V=-\frac{d\Phi_m}{dt}=-NS\mathscr{B}_0\omega \cos \omega t, \tag{8.3}$$

which indicates that the induced emf is oscillatory or alternating with the same frequency as the magnetic field. ▲

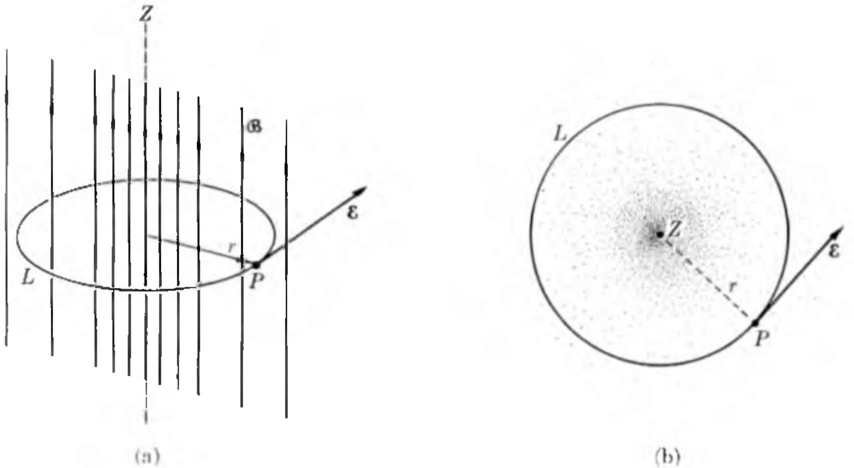


Fig. 8-3. Electric field produced by a time-dependent magnetic field having cylindrical symmetry; (a) side view. (b) top view.

Example 8.2. Determination of the electric field $\mathscr{E}$ produced by an alternating axial magnetic field.

▼ Assume that in a region of space there is a magnetic field parallel to the Z-axis and having axial symmetry; that is, the field's magnitude at each point depends on the distance r to the Z-axis only. The magnitude also varies with time. Assume also that the magnetic field decreases with the distance from the Z-axis. Figure 8-3(a) shows a side view of the field, and Fig. 8-3(b) shows a cross section.

The symmetry of the problem suggests that the electric field $\mathscr{E}$ must depend on the distance r alone, and at each point be perpendicular to the magnetic field $\mathscr{B}$ and to the radius r. In other words the lines of force of the electric field $\mathscr{E}$ are circles concentric with the Z-axis. Choosing the path L in Eq. (8.2) as one of these circles gives the induced emf:

$$V=\oint_L \mathscr{E}\cdot d\boldsymbol{l}=\mathscr{E}(2\pi r).$$

Therefore Eq. (8.1) yields

$$\mathscr{E}(2\pi r)=-\frac{d\Phi_m}{dt}. \tag{8.4}$$

The *average* magnetic field $\mathscr{B}_{\text{ave}}$ in a region covering an area S is defined as $\mathscr{B}_{\text{ave}}=\Phi_m/S$ or $\Phi_m=\mathscr{B}_{\text{ave}}S$. Here $S=\pi r^2$ so that $\Phi_m=\mathscr{B}_{\text{ave}}(\pi r^2)$. Then Eq. (8.4) gives the electric field at a distance r from the axis as

$$\mathscr{E}=-\tfrac{1}{2}r\left(\frac{d\mathscr{B}_{\text{ave}}}{dt}\right). \tag{8.5}$$

If the magnetic field were uniform, $\mathscr{B}_{\text{ave}}$ would simply equal $\mathscr{B}$. ▲

8.3 The Betatron

The results of Example 8.2 were used to design an electron accelerator called a *betatron*, invented in 1941 by the American physicist D. Kerst (1911–). The idea is very simple in principle. If an electron (or any kind of charged particle) is injected into the region in which a varying magnetic field exists, the electron will be accelerated by the associated electric field $\mathscr{E}$ as given by Eq. (8.2) or Eq. (8.5). As the electron gains velocity, the electron's path will be bent by the magnetic field $\mathscr{B}$. If the magnetic field has axial symmetry and the electric and magnetic fields are adjusted properly, the orbit of the electron is a circle such that in each revolution the electron gains energy.

To see the problem in more detail, consider the electron at point P (Fig. 8-3). If things are arranged so that the electron describes a circular path of radius r, the electric field will produce a tangential component to the electron's motion; this component is computed by using $dp/dt = F_T$ where the tangential force $F_T = -e\mathscr{E}$ so that

$$\frac{dp}{dt} = -e\mathscr{E} = \tfrac{1}{2}er\left(\frac{d\mathscr{B}_{\text{ave}}}{dt}\right). \tag{8.6}$$

To generate circular motion, the magnetic field must produce the necessary centripetal acceleration. According to Eq. (4.1) the magnitude of the centripetal force is $F_N = ev\mathscr{B}$. Recall that in circular motion the centripetal acceleration is $mv^2/r = (mv)(v/r) = p(v/r)$ so that $pv/r = F_N$; then the two equations may be combined to give

$$pv/r = ev\mathscr{B} \qquad \text{or} \qquad p = er\mathscr{B}. \tag{8.7}$$

Taking the time derivative of Eq. (8.7) and observing that r is constant because the path is a circle give

$$\frac{dp}{dt} = er\frac{d\mathscr{B}}{dt}.$$

When this equation is compared with Eq. (8.6), the necessary condition for the electron to describe a circular orbit of radius r under the combined action of the electric and magnetic fields is that at the distance r the magnetic field must be

$$\mathscr{B} = \tfrac{1}{2}\mathscr{B}_{\text{ave}} \tag{8.8}$$

where $\mathscr{B}_{\text{ave}}$ is the average value of $\mathscr{B}$ in the region surrounded by L. This condition imposes certain requirements on the manner in which the magnetic field $\mathscr{B}$ may vary as a function of the radial distance r from the axis. The exact variation of $\mathscr{B}$ with r is determined by the requirement of a certain stability of the orbital motion. That is, given the radius of the desired orbit, the forces on the electron must be such that if the motion of the electron is slightly disturbed (i.e., if it is pushed to one side or the other

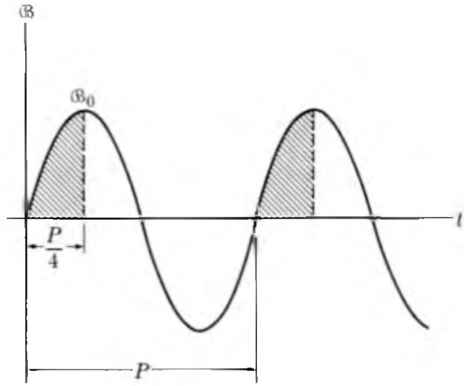


Fig. 8-4. Accelerating time in a betatron.

of the orbit), the electric and magnetic forces acting on the electron tend to pull it back into the correct orbit.

In general the magnetic field is oscillatory with some angular frequency ω. Now because of Eq. (8.6) the electron is accelerated only while the magnetic field is increasing. On the other hand since electrons are in practice injected with very small momentum, they must be injected when the magnetic field is zero. Therefore only one quarter of the period of variation of the magnetic field is good for accelerating the electrons. The accelerating times have been indicated by the shaded areas in Fig. 8-4.

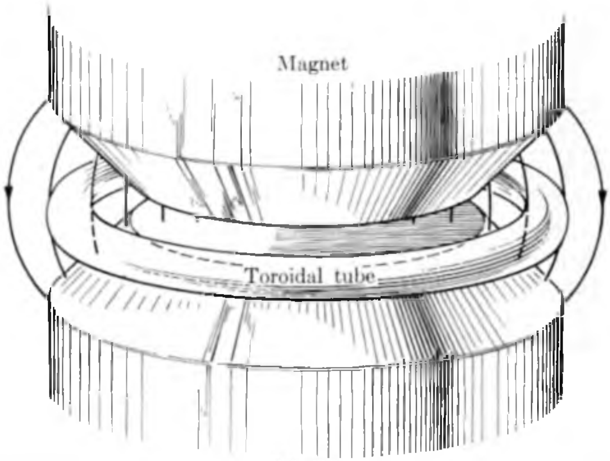
According to Eq. (8.7), the maximum momentum gained by the electron is $p_{\max} = er\mathscr{B}_0$, and therefore the maximum kinetic energy of the accelerated electrons is

$$E_{k,\max} = \frac{1}{2m_e} p_{\max}^2 = \frac{e^2r^2\mathscr{B}_0^2}{2m_e}$$

if they are not accelerated to very high energy compared with the electron rest energy m_ec^2. However when the energy is rather large, comparable to or larger than the rest energy m_ec^2 of the electron, relativistic equations must be used and result in

$$E_{k,\max} = c\sqrt{m_e^2c^2 + e^2r^2\mathscr{B}_0^2} - m_ec^2.$$

Actual betatrons consist of a toroidal tube (Fig. 8-5) placed in the magnetic field produced by a magnet whose pole faces have been so designed or shaped that the correct variation of the magnetic field $\mathscr{B}$ with r according to Eq. (8.8) is produced and stability conditions are fulfilled. The electrons are injected at the beginning of the accelerating period and slightly deflected at its end so that they may hit a target properly located. The kinetic energy of the electrons is given off as electromagnetic radiation and/or as internal energy of the target that is heated up. Betatrons have been built with energies up to 350 Mev. Betatrons are used for studies of certain types of nuclear reactions and for radiation treatment of cancer.



(a)

(b)

Fig. 8-5. (a) View of the accelerating tube and pole faces of a betatron. (b) Assembling the accelerating tube in a betatron.

8.4 Electromagnetic Induction Caused by Relative Motion of Conductor and Magnetic Field

The law of electromagnetic induction as expressed in Eq. (8.2) implies the existence of a local electric field whenever the magnetic field at that point is changing with time. As expressed in Eq. (8.1), the law implies the existence of an emf when the magnetic flux through the circuit changes with time. It is important to discover whether the same results occur when the change in flux is due to a motion or deformation of the path L without $\mathscr{B}$ necessarily changing with time. Consider two simple cases.

First consider the arrangement of conductors illustrated in Fig. 8-6, in which the conductor PQ can move parallel to itself with velocity v while maintaining contact with conductors RT and SU. The system $PQRS$ forms a closed circuit. Suppose also that there is a uniform magnetic field $\mathscr{B}$ perpendicular to the plane of the system.

When the bar PQ is moving, each charge q in the conductor PQ is subject to a force $q\mathbf{v}\times\mathscr{B}$ acting along QP, according to Eq. (2.1). Now the same force on the charge could be assumed to be due to an "equivalent" electric field $\mathscr{E}_{eq}$ given by

$$q\mathscr{E}_{eq}=q\mathbf{v}\times\mathscr{B}$$

or

$$\mathscr{E}_{eq}=\mathbf{v}\times\mathscr{B}.$$

Since $\mathbf{v}$ and $\mathscr{B}$ are perpendicular, the relation among the magnitudes is

$$\mathscr{E}_{eq}=v\mathscr{B}. \tag{8.9}$$

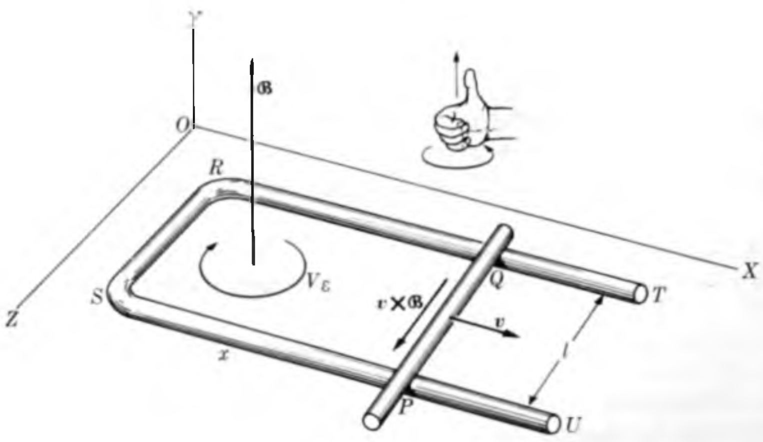


Fig. 8-6. Emf induced in a conductor moving in a magnetic field.

If $PQ=l$, the potential difference existing between P and Q is given by $\Delta V=\mathscr{E}_{eq}l=\mathscr{B}vl$. No forces are exerted on the sections QR, RS, and SP since they are stationary relative to the magnetic field. Therefore the circulation of $\mathscr{E}_{eq}$ (or the emf) along circuit $PQRS$ is just $V=\Delta V$ in the direction of $\boldsymbol{v}\times\mathscr{B}$; that is,

$$V=\mathscr{B}vl.$$

On the other hand if the length SP is called x, the area of $PQRS$ is lx; and the magnetic flux through $PQRS$ is

$$\Phi_m=\int_{PQRS}\mathscr{B}\cdot\boldsymbol{u}_N\,dS=\mathscr{B}lx.$$

The change of flux per unit time is then

$$\frac{d\Phi_m}{dt}=\frac{d}{dt}(\mathscr{B}lx)=\mathscr{B}l\frac{dx}{dt}.$$

Because $dx/dt=v$,

$$\frac{d\Phi_m}{dt}=\mathscr{B}lv=V.$$

In other words Eq. (8.1) is the result. The minus sign is not included because only the relation between the magnitudes has been considered. However, relation (8.1) still holds in sign since the flux Φ_m is increasing; and the sign of V is that of $\boldsymbol{v}\times\mathscr{B}$ so that it agrees with Fig. 8-1.

As a second example, consider a rectangular circuit rotating in a uniform magnetic field $\mathscr{B}$ with angular frequency ω (Fig. 8-7). When the normal $\boldsymbol{u}_N$ to the plane of the circuit makes an angle $\theta=\omega t$ with the magnetic field $\mathscr{B}$, all points of PQ are moving with a velocity $\boldsymbol{v}$ such that the "equivalent" electric field $\mathscr{E}_{eq}=\boldsymbol{v}\times\mathscr{B}$ points from Q to P and has a magnitude $\mathscr{E}_{eq}=v\mathscr{B}\sin\theta$. Similarly for points on RS the direction of $\boldsymbol{v}\times\mathscr{B}$ is from S to R and has the same magnitude. On the sides RQ and PS, we see that $\boldsymbol{v}\times\mathscr{B}$ is perpendicular to the conductors; and thus no potential difference exists between S and P and between R and Q. If $PQ=RS=l$, the circulation of the equivalent electric field $\mathscr{E}_{eq}$ around $PQRS$, is

$$V=\oint_L\mathscr{E}\cdot d\boldsymbol{l}=\mathscr{E}_{eq}(PQ+SR)=2lv\mathscr{B}\sin\theta.$$

If $x=SP$, the radius of the circle described by the charges in PQ and SR is $\frac{1}{2}x$; and therefore $v=\omega(\frac{1}{2}x)=\frac{1}{2}\omega x$. Then since $S=lx$ is the area of the circuit and $\theta=\omega t$,

$$V=2l(\tfrac{1}{2}\omega x)\mathscr{B}\sin\omega t=\omega\mathscr{B}(lx)\sin\omega t=\omega\mathscr{B}S\sin\omega t$$

for the emf induced in the circuit as a result of its rotation in the magnetic field. On the other hand the magnetic flux through the circuit is

$$\Phi_m=\mathscr{B}\cdot\boldsymbol{u}_NS=\mathscr{B}S\cos\theta=\mathscr{B}S\cos\omega t.$$

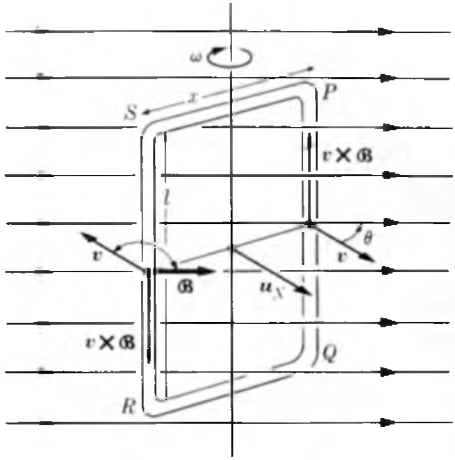


Fig. 8-7. Emf induced in a rotating coil placed in a magnetic field.

Then

$$-\frac{d\Phi_m}{dt}=\omega\mathcal{B}S\sin\omega t=V.$$

Therefore the induced emf resulting from the motion of the conductor can also be calculated by applying Eqs. (8.1) or (8.2) instead of Eqs. (4.1) and (3.15).

Although the discussion has dealt only with circuits of special shapes, a more detailed mathematical calculation indicates that for any circuit

> *the law of electromagnetic induction $V=-d\Phi_m/dt$ can be applied when the change in magnetic flux Φ_m is due either to a change in the magnetic field $\mathcal{B}$ or to a motion or a deformation of the circuit along which the emf is calculated, or both.*

The induced emf in the second case is sometimes called *motional emf.*

8.5 Electromagnetic Induction and the Principle of Relativity

Although the law of electromagnetic induction as expressed by Eqs. (8.1) and (8.2) is valid no matter what the origin of the change of magnetic flux, there is a profound difference in the physical situations in the two possibilities. When an observer recognizes that the change of magnetic flux through a circuit stationary in the observer's own frame of reference is due to a change in the magnetic field $\mathcal{B}$, an electric field $\mathcal{E}$, related to $\mathcal{B}$ as indicated by Eq. (8.2), is measured at the same time; and the presence of the electric field is recognized by measuring the force on a charge *at rest* in the observer's frame of reference. However when the observer recognizes that the

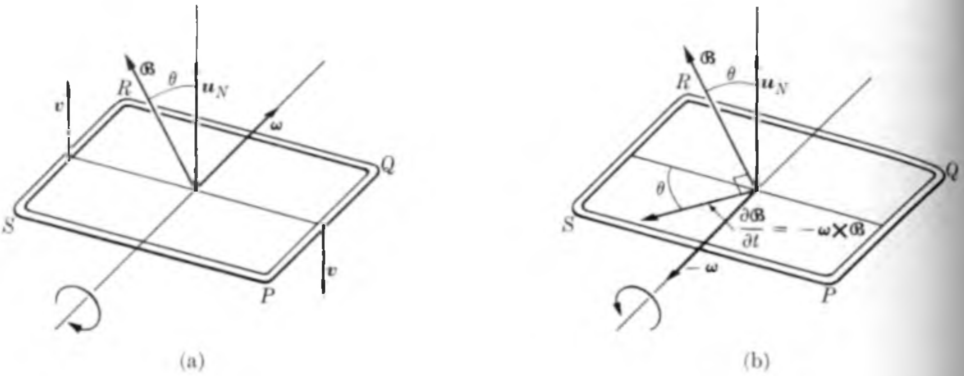


Figure 8-8

change of magnetic flux is due to the conductor's motion relative to its frame of reference, no electric field is observed but the emf measured is assigned to the force $q \times \mathscr{B}$ exerted by the magnetic field on the charges of the moving conductor.

How does it happen that two different and apparently unrelated situations have a common description? This is not a matter of coincidence, but strictly a consequence of the principle of relativity. A full mathematical analysis will not be attempted here: instead, the situation will be examined from an intuitive point of view. Consider the case of the rotating circuit discussed in connection with Fig. 8-7. In a frame of reference in which the magnetic field $\mathscr{B}$ is constant (Fig. 8-8a) and the circuit is rotating, no electric field is observed and the forces on the electrons in the circuit are due to Eq. (3.1); but an observer attached to a frame moving with the circuit sees a stationary conductor and a magnetic field $\mathscr{B}$ whose direction rotates in space (Fig. 8-8b). The observer then relates the forces on the electrons in the circuit to the electric field $\mathscr{E}$ associated with a changing magnetic field, according to the law of electromagnetic induction as expressed by Eq. (8.2).

The experimental verification of the law of electromagnetic induction for changing magnetic fields is simply a reaffirmation of the general validity of the principle of relativity.

8.6 Electric Potential and Electromagnetic Induction

In Chapters 1 and 2 it was shown that a static electric field $\mathscr{E}$ is associated with an electric potential V in such a way that the components of $\mathscr{E}$ along the X-, Y-, and Z-axes are the negatives of the derivatives of V relative to x, y, and z. That is, $\mathscr{E}_x = -\partial V/\partial x$, etc.: or to state the matter more simply, the electric field is the negative of the gradient of the electric potential. A consequence of this fact is that the circulation of the static electric field around any closed path is zero, a property that is expressed mathematically by the statement (3.17) or

$$\oint_L \mathscr{E} \cdot dl = 0.$$

However when the electromagnetic field is time dependent. we have seen that the equation above is no longer valid; instead Eq. (8.2) applies:

$$\oint_L \mathscr{E} \cdot dl = -\frac{d}{dt} \int_S \mathscr{B} \cdot u_N \, dS.$$

Thus in a time-dependent electromagnetic field the circulation of the electric field is not zero. and therefore the electric field cannot be expressed as the negative of the gradient of the electric potential. This statement does not mean that the concept of potential is completely inapplicable in this case. but only that the concept must be used in a different form. In fact, *two* potentials are required. One is called the *scalar potential*. similar to the one used in the static case. and the other is a *vector potential*. There will be no occasion to use these potentials in this text; they are mentioned here only to point out the need for great care about which static-field concepts may be kept and which must be modified in passing from static to time-dependent fields.

8.7 The Faraday – Henry Law in Differential Form

The law of electromagnetic induction as expressed by Eq. (8.2) can be applied to a path of any shape. This law will now be applied to a very small or infinitesimal rectangular path *PQRS* placed in the *XY*-plane and having sides dx and dy (Fig. 8-9). First the circulation of the electric field $\mathscr{E}$ must be evaluated. The procedure is exactly similar to the one followed in the discussion of Ampère's law in differential form;

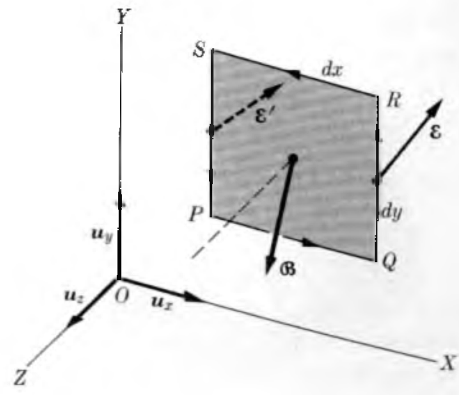


Fig. 8-9. Elementary circuit for deriving the Faraday-Henry law in differential form.

the student is referred to Section 6.3 for the details. For the infinitesimal surface $PQRS$ in the XY-plane.

$$\oint_{PQRS} \mathscr{E}\cdot dl = \int_{PQ} + \int_{QR} + \int_{RS} + \int_{SP} \mathscr{E}\cdot dl.$$

Now $\int_{QR} \mathscr{E}\cdot dl = \mathscr{E}_y\, dy$ and $\int_{SP} \mathscr{E}\cdot dl = -\mathscr{E}'_y\, dy$ so that

$$\int_{QR} + \int_{SP} \mathscr{E}\cdot dl = (\mathscr{E}_y - \mathscr{E}'_y)\, dy = d\mathscr{E}_y\, dy = \frac{\partial \mathscr{E}_y}{\partial x}\, dx\, dy.$$

We may replace $d\mathscr{E}_y$ by $(\partial \mathscr{E}_y/\partial x)\, dx$ since $d\mathscr{E}_y$ corresponds to the difference in $\mathscr{E}_y$ for two points separated by a very small distance dx, but having the same y and z coordinates. Similarly for the other two sides

$$\int_{PQ} + \int_{RS} \mathscr{E}\cdot dl = -\frac{\partial \mathscr{E}_x}{\partial y}\, dx\, dy.$$

Adding the two results gives

$$\oint_{PQRS} \mathscr{E}\cdot dl = \left(\frac{\partial \mathscr{E}_y}{\partial x} - \frac{\partial \mathscr{E}_x}{\partial y}\right) dx\, dy. \tag{8.10}$$

Next compute the magnetic flux through the surface. Because the surface $PQRS$ is in the XY-plane, the normal unit vector $\boldsymbol{u}_N$ is just $\boldsymbol{u}_z$ and $\mathscr{B}\cdot \boldsymbol{u}_N = \mathscr{B}\cdot \boldsymbol{u}_z = \mathscr{B}_z$. Therefore the magnetic flux is

$$\int_{PQRS} \mathscr{B}\cdot \boldsymbol{u}_N\, dS = \mathscr{B}_z\, dx\, dy \tag{8.11}$$

since $dx\, dy$ is the area of the rectangle. Substituting Eqs. (8.10) and (8.11) into Eq. (8.2), and canceling the common factor $dx\, dy$ on both sides give

$$\frac{\partial \mathscr{E}_y}{\partial x} - \frac{\partial \mathscr{E}_x}{\partial y} = -\frac{\partial \mathscr{B}_z}{\partial t}. \tag{8.12}$$

By placing the rectangle in the YZ- and ZX-planes, two other expressions may be written:

$$\frac{\partial \mathscr{E}_z}{\partial y} - \frac{\partial \mathscr{E}_y}{\partial z} = -\frac{\partial \mathscr{B}_x}{\partial t} \tag{8.13}$$

and

$$\frac{\partial \mathscr{E}_x}{\partial z} - \frac{\partial \mathscr{E}_z}{\partial x} = -\frac{\partial \mathscr{B}_y}{\partial t}. \tag{8.14}$$

Expressions (8.12), (8.13), and (8.14) together constitute the Faraday-Henry law expressed in differential form. They can be combined into a single vector equation as was done in Section 6.3 for Ampère's law:

$$\operatorname{curl} \mathcal{E} = -\frac{\partial \mathcal{B}}{\partial t}. \tag{8.15}$$

Equation (8.15), as well as its equivalents (8.12), (8.13), and (8.14), expresses the relations that must exist between the time rate of change of the magnetic field at a point and the electric field existing at the *same* point of space. Equation (8.15) illustrates in a very obvious way the close interrelationship between the electric and magnetic components of an electromagnetic field.

8.8 The Principle of Conservation of Charge

In Section 1.2 the fact that electric charge is conserved was discussed. In other words in all processes that occur in the universe, the net amount of charge must always remain the same. This statement may be expressed in a quantitative way that is very useful. Consider a closed surface S (Fig. 8-10), and designate by q the net charge inside S at a given time. Since the problem is dynamic and not static, free charges (such as electrons in metals or ions in a plasma) are moving through the medium and crossing the surface S. At some times there may be more outgoing charges than entering ones; this difference results in a decrease in the net charge q within the surface S. At other times the situation may be reversed, and the incoming charges may exceed those leaving; the result is an increase in the net charge q. Of course if the outgoing and the incoming charge fluxes through S are the same, the net charge q remains the same. The principle of conservation of charge obviously requires that

$$\begin{aligned} \text{Loss of charge} &= \text{outgoing charge flux} - \text{incoming charge flux} \\ &= \text{net outgoing charge flux.} \end{aligned} \tag{8.16}$$

Now the net charge flux per unit time, or current, through a surface S was found in Example 4.1 to be $I = \int_S \boldsymbol{j} \cdot \boldsymbol{u}_N \, dS$ where $\boldsymbol{j}$ is the current density. In the present case the surface S is closed so that

$$I = \oint_S \boldsymbol{j} \cdot \boldsymbol{u}_N \, dS \tag{8.17}$$

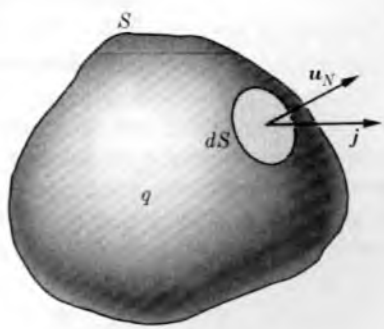


Fig. 8-10. Current across a closed surface enclosing a charge q.

gives the net charge passing out through the surface per unit time. On the other hand the loss of charge per unit time within S is $-dq/dt$. Therefore in mathematical terms Eq. (8.16) becomes $-dq/dt = I$ or

$$-\frac{dq}{dt} = \oint_S \boldsymbol{j} \cdot \boldsymbol{u}_N \, dS. \tag{8.18}$$

an equation that expresses the principle of conservation of charge under the assumption that charge is neither created nor annihilated. Now according to Gauss's law for the electric field as given by Eq. (5.3), the total charge within a closed surface is expressed in terms of the electric field at the surface by

$$q = \epsilon_0 \oint_S \boldsymbol{\mathscr{E}} \cdot \boldsymbol{u}_N \, dS$$

so that

$$\frac{dq}{dt} = \epsilon_0 \frac{d}{dt} \oint_S \boldsymbol{\mathscr{E}} \cdot \boldsymbol{u}_N \, dS.$$

Substituting this result into Eq. (8.18) allows us to write

$$\oint_S \boldsymbol{j} \cdot \boldsymbol{u}_N \, dS + \epsilon_0 \frac{d}{dt} \oint_S \boldsymbol{\mathscr{E}} \cdot \boldsymbol{u}_N \, dS = 0 \tag{8.19}$$

for the expression of the principle of conservation of charge in a way that incorporates Gauss's law. When the fields are static, the integral $\oint_S \boldsymbol{\mathscr{E}} \cdot \boldsymbol{u}_N \, dS$ does not depend on time. The integral's time derivative is therefore zero and results then in

$$\oint_S \boldsymbol{j} \cdot \boldsymbol{u}_N \, dS = 0 \qquad \text{for static fields.} \tag{8.20}$$

This expression means that for static fields there is no accumulation or loss of charge in any region of space, and the net current across a closed surface is zero. (This is essentially the content of Kirchhoff's first law for network analysis, introduced in Example 3.4.)

8.9 The Ampère – Maxwell Law

The Faraday-Henry law as expressed in Eqs. (8.2) or (8.15) establishes a relation between the magnetic field and the electric field in the same region of space. The close relationship that exists between the electric and magnetic fields suggests that an analogous relation should exist between the time rate of change of an electric field and a magnetic field at the same place. That is, since

$$\oint_L \boldsymbol{\mathscr{E}} \cdot d\boldsymbol{l} = -\frac{d}{dt} \int_S \boldsymbol{\mathscr{B}} \cdot \boldsymbol{u}_N \, dS$$

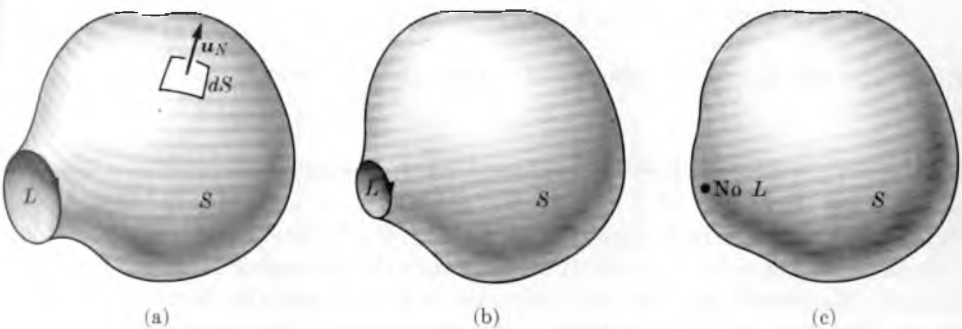


Fig. 8-11. Surface bounded by line L. When line L shrinks to a point, the surface becomes closed.

relates the circulation of the electric field to the time rate of change of the flux of the magnetic field, it might be expected that a similar expression must relate the circulation of the magnetic field to the time rate of change of the flux of the electric field. The circulation of the magnetic field was expressed in Ampère's law as

$$\oint_L \mathscr{B}\cdot dl = \mu_0 \int_S j\cdot u_N\, dS; \tag{8.21}$$

but this expression does not contain any time rate of change of the flux of the electric field. This lack is not surprising since this expression was derived under static conditions. However, Ampère's law needs a revision if it is to be applied to time-dependent fields.

Ampère's law in the form (8.21) applies to a surface S bounded by a contour L. The surface S is arbitrary so long as it is bounded by L. If the line L shrinks, the value of $\oint_L \mathscr{B}\cdot dl$ decreases (Fig. 8-11) until eventually the value becomes zero when L shrinks to a point, and the surface S becomes a *closed* surface. Ampère's law as expressed by Eq. (8.21) then requires that

$$\oint_S j\cdot u_N\, dS = 0.$$

This equation agrees with Eq. (8.20) for the conservation of charge so long as the field is static. However, when the field is not static but time dependent, Eq. (8.20) is no longer correct. Instead it is Eq. (8.19), which incorporates Gauss's law in the conservation of charge, that is always correct. Thus the suspicion that Ampère's law must be modified when dealing with time-dependent fields is confirmed. The modification seems obvious. The integral $\int_S j\cdot u_N\, dS$ in Eq. (8.21) must be replaced by

$$\int_S j\cdot u_N\, dS + \epsilon_0 \frac{d}{dt}\int_S \mathscr{E}\cdot u_N\, dS$$

in accordance with Eq. (8.19). The result is

$$\oint_L \mathscr{B}\cdot dl = \mu_0 \int_S j\cdot u_N\, dS + \epsilon_0\mu_0 \frac{d}{dt}\int_S \mathscr{E}\cdot u_N\, dS. \tag{8.22}$$

Remember that $\int_S j\cdot u_N\, dS$ is the current I through the surface S; then Eq. (8.22) may be rewritten as

$$\oint_L \mathscr{B}\cdot dl = \mu_0 I + \epsilon_0\mu_0 \frac{d}{dt}\int_S \mathscr{E}\cdot u_N\, dS. \tag{8.23}$$

This equation should be compared with Eq. (6.3) for Ampère's law. Equation (8.23) reduces to Ampère's law for static fields since then the last term is zero; and Eq. (8.23) becomes Eq. (8.19) when the line L shrinks to a point and the surface S becomes closed. Therefore Eq. (8.23) satisfies all the physical principles previously discussed.

So far we have merely played with mathematics in an attempt to make Ampère's law compatible with the law of conservation of charge. One necessary further step is to verify experimentally that Eq. (8.22) is correct. and that it describes the actual situation found in nature. The best proof is the existence of electromagnetic waves, a subject that will be discussed in a later chapter.

The person who first suggested the modification of Ampère's law in the way indicated here was the British physicist James Clerk Maxwell (1831–1879), and therefore Eq. (8.22) is called the *Ampère-Maxwell law*. Maxwell's modification came about more because of the urge for mathematical consistency than because of experimentation. In fact the experiments substantiating Maxwell's ideas came only some years later.

Ampère's law (Eq. 8.21) relates a steady current to the magnetic field the current produces. The Ampère-Maxwell law (Eq. 8.22) goes a step further and indicates that a time-dependent electric field $\mathscr{E}$ also contributes to the magnetic field. For example in the absence of currents Eq. (8.22) becomes

$$\oint_L \mathscr{B}\cdot dl = \epsilon_0\mu_0 \frac{d}{dt}\int_S \mathscr{E}\cdot u_N\, dS. \tag{8.24}$$

which shows more clearly the relation between a time-dependent electric field and its associated magnetic field. In other words

> *a time-dependent electric field at some point in space implies the existence of a magnetic field at the same place.*

The circulation of the magnetic field is called the *magnetomotive force* applied to the closed line L and is designated by $\Lambda_{\mathscr{B}}$. The electric flux across the surface S bounded by the contour L is designated by Φ_e. Then Eq. (8.24) may be written in the form

$$\Lambda_{\mathscr{B}} = \epsilon_0\mu_0 \frac{d\Phi_e}{dt},$$

which the student should compare with Eq. (8.1) for the law of electromagnetic induction. The student should also verify that the factor $\epsilon_0\mu_0$ is consistent with SI

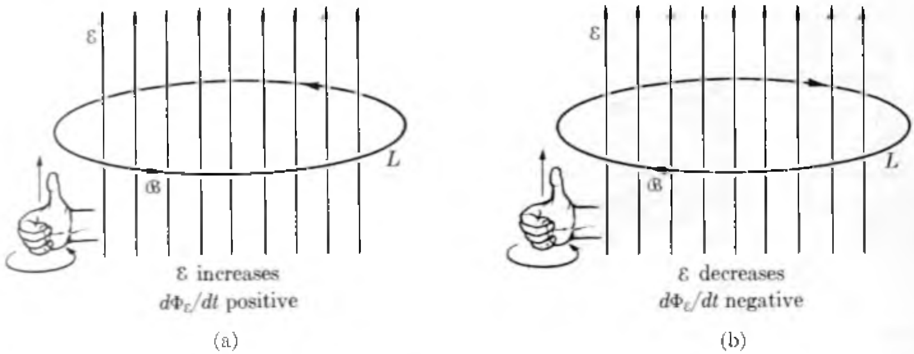


Fig. 8-12. Magnetic field produced by a time-dependent electric field.

units. The relative orientation of the electric and the magnetic fields is shown in Fig. 8-12. corresponding to a time-dependent uniform electric field. If the electric field increases (decreases). the orientation of the magnetic lines of force is the same as (opposite to) the sense of rotation of a right-handed screw that advances in the direction of the electric field. The student should compare this result with Fig. 8-1.

The Ampère-Maxwell law as expressed by Eq. (8.23) differs from the Faraday-Henry law as expressed by Eq. (8.2). in several respects. In the first place Eq. (8.23) has a term corresponding to an *electric* current: in Eq. (8.2) no term corresponds to a *magnetic* current simply because there are apparently no free magnetic poles in nature. In the second place the time rate of change of the electric flux appears with a positive sign in Eq. (8.23); the magnetic flux appears with a negative sign in Eq. (8.2).

Although Ampère's law has been amended using the principle of conservation of charge as guide. the principle of relativity could have equally been used to develop this modification. With relativity. it has been found that when the electric and magnetic fields are related in two inertial frames of reference as in Eqs. (2.21) and (2.23), and the Faraday-Henry law is correct. then Eq. (8.23) must also be satisfied. This procedure is a little more difficult. but in a sense it is more fundamental.

8.10 The Ampère – Maxwell Law in Differential Form

Since Eq. (8.22) for the Ampère-Maxwell law is very similar to Eq. (8.2) for the Faraday-Henry law. the technique used in Section 8.7 for obtaining the Ampère-Maxwell law in differential form may again be applied. Figure 8-9 is now replaced by Fig. 8-13. In analogy with Eq. (8.10) the circulation of the magnetic field along the rectangular path $PQRS$ whose sides are dx and dy is

$$\oint_{PQRS} \mathscr{B}\cdot d\mathbf{l} = \left(\frac{\partial \mathscr{B}_y}{\partial x} - \frac{\partial \mathscr{B}_x}{\partial y}\right) dx\, dy. \tag{8.25}$$

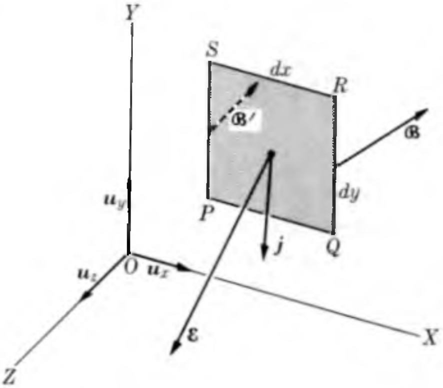


Fig. 8-13. Elementary circuit for deriving the Ampère-Maxwell law in differential form.

The flux of the electric current through the surface bounded by $PQRS$ was obtained in Eq. (6.10) when Ampère's law in differential form was derived. It was

$$\int_{PQRS} \boldsymbol{j} \cdot \boldsymbol{u}_N \, dS = j_z \, dx \, dy. \tag{8.26}$$

Finally the flux of the electric field through the surface bounded by $PQRS$ is analogous to Eq. (8.26) and is

$$\int_{PQRS} \boldsymbol{\mathscr{E}} \cdot \boldsymbol{u}_N \, dS = \mathscr{E}_z \, dx \, dy;$$

and therefore

$$\frac{d}{dt} \int_{PQRS} \boldsymbol{\mathscr{E}} \cdot \boldsymbol{u}_N \, dS = \frac{\partial \mathscr{E}_z}{\partial t} \, dx \, dy. \tag{8.27}$$

Substituting Eqs. (8.25), (8.26), and (8.27) in Eq. (8.22), and canceling the common factor $dx\,dy$ on both sides yield

$$\frac{\partial \mathscr{B}_y}{\partial x} - \frac{\partial \mathscr{B}_x}{\partial y} = \mu_0 j_z + \epsilon_0 \mu_0 \frac{\partial \mathscr{E}_z}{\partial t}. \tag{8.28}$$

With the rectangle placed in the YZ- and ZX-planes, two other expressions are obtained:

$$\frac{\partial \mathscr{B}_z}{\partial y} - \frac{\partial \mathscr{B}_y}{\partial z} = \mu_0 j_x + \epsilon_0 \mu_0 \frac{\partial \mathscr{E}_x}{\partial t} \tag{8.29}$$

and

$$\frac{\partial \mathscr{B}_x}{\partial z} - \frac{\partial \mathscr{B}_z}{\partial x} = \mu_0 j_y + \epsilon_0 \mu_0 \frac{\partial \mathscr{E}_y}{\partial t}. \tag{8.30}$$

Expressions (8.28), (8.29), and (8.30) together constitute the Ampère-Maxwell law in differential form. As done previously for Ampère's law and Faraday-Henry's law, these equations may be combined into a single vector equation by writing

$$\operatorname{curl} \mathscr{B} = \mu_0 \left(\boldsymbol{j} + \epsilon_0 \frac{\partial \mathscr{E}}{\partial t} \right), \tag{8.31}$$

which expresses a relation between the electric current at a point in space and the electric and the magnetic fields at the same point. In empty space with no currents, $j=0$, and Eq. (8.31) becomes

$$\operatorname{curl} \mathscr{B} = \mu_0 \epsilon_0 \frac{\partial \mathscr{E}}{\partial t}, \tag{8.32}$$

which is the equivalent to Eq. (8.24) in differential form. Equation (8.32) is similar to Eq. (8.15) for the Faraday-Henry law, and clearly shows the relationship between the magnetic field and the time rate of change of the electric field at the same point.

8.11 Maxwell's Equations

At this point let us summarize our discussion of the electromagnetic field. An important kind of interaction among the fundamental particles composing matter is the *electromagnetic interaction*. It is associated with *electric charge*, a characteristic property of each particle. To describe the electromagnetic interaction, the notion of *electromagnetic field* (characterized by two vectors: the *electric field* $\mathscr{E}$ and the *magnetic field* $\mathscr{B}$) has been introduced, such that the force on an electric charge is given by

$$\boldsymbol{F} = q(\mathscr{E} + \boldsymbol{v} \times \mathscr{B}). \tag{8.33}$$

The electric and the magnetic fields $\mathscr{E}$ and $\mathscr{B}$ are in turn determined by the positions of the charges themselves and by their motions. The separation of the electromagnetic field into its electric and magnetic components depends on the relative motion of the observer and the charges producing the field. Also the fields $\mathscr{E}$ and $\mathscr{B}$ are directly correlated with each other by the Ampère-Maxwell and Faraday-Henry laws. All these relations are expressed by four laws, which have been analyzed in the previous chapters, and which may be written both in their integral and differential forms as in Table 8-1.

The entire theory of the electromagnetic field is condensed into these four laws. They are called *Maxwell's equations* since it was Maxwell who, in addition to formulating the fourth law, recognized that they, together with Eq. (8.33), constitute the basic framework of the theory of electromagnetic interactions. The electric charge q and the current I are called the *sources* of the electromagnetic field since given q and I, Maxwell's equations allow the calculation of $\mathscr{E}$ and $\mathscr{B}$.

Table 8-1. Maxwell's Equations for the Electromagnetic Field

Law	Integral form	Differential form
I. Gauss's law for the electric field [(2.3) and (2.5)]	$\oint \mathscr{E}\cdot \boldsymbol{u}_N\, dS = \dfrac{q}{\epsilon_0}$	$\operatorname{div} \mathscr{E} = \dfrac{\rho}{\epsilon_0}$ IV
II. Gauss's law for the magnetic field [(6.17) and (6.18)]	$\oint \mathscr{B}\cdot \boldsymbol{u}_N\, dS = 0$	$\operatorname{div} \mathscr{B} = 0$ III
III. Faraday-Henry law [(8.2) and (8.15)]	$\oint \mathscr{E}\cdot d\boldsymbol{l} = -\dfrac{d}{dt}\displaystyle\int_S \mathscr{B}\cdot \boldsymbol{u}_N\, dS$	$\operatorname{curl} \mathscr{E} = -\dfrac{\partial \mathscr{B}}{\partial t}$ II
IV. Ampère-Maxwell law [(8.23) and (8.31)]	$\oint \mathscr{B}\cdot d\boldsymbol{l} = \mu_0 I + \epsilon_0\mu_0 \dfrac{d}{dt}\displaystyle\int_S \mathscr{E}\cdot \boldsymbol{u}_N dS$	$\operatorname{curl} \mathscr{B} = \mu_0 \boldsymbol{j} + \epsilon_0\mu_0 \dfrac{\partial \mathscr{E}}{\partial t}$ I

Note that Gauss's laws for the electric and magnetic fields, Eqs. (2.3) and (6.17), were derived for static fields. However, we are now incorporating these laws into a theory involving time-dependent fields. The student may wonder if perhaps these laws may have to be revised in the same way that Ampère's law was modified to make it applicable to a time-dependent situation. The answer is no. It has been found that this set of laws is in agreement with experiment, and the consequences derived from them have so far been found to agree with experimental results. Therefore the two Gauss laws stay the same when applied to time-dependent electric and magnetic fields.

Maxwell's equations also form a consistent set of equations. On one side, Eqs. (2.3) and (8.23), which involve a surface integral of the electric field, are consistent: this consistency was our basic requirement in revising Ampere's law. Also Eqs. (6.17) and (8.2), which involve a surface integral of the magnetic field, are consistent. For example, applying Eq. (8.2) to the surface of Fig. (8-11) when the curve L shrinks until the surface is closed, the circulation of $\mathscr{E}$ becomes zero: and therefore

$$\frac{d}{dt}\oint_S \mathscr{B}\cdot \boldsymbol{u}_N\, dS = 0 \qquad \text{or} \qquad \oint_S \mathscr{B}\cdot \boldsymbol{u}_N\, dS = \text{const},$$

which coincides with Eq. (6.17) if the constant of integration is zero.

In free or empty space where there are no charges ($\rho = 0$) nor currents ($\boldsymbol{j} = 0$), Maxwell's equations are slightly simpler and become in the differential form

$$\operatorname{div} \mathscr{E} = 0, \qquad \operatorname{div} \mathscr{B} = 0,$$
$$\operatorname{curl} \mathscr{E} = -\frac{\partial \mathscr{B}}{\partial t}, \quad \operatorname{curl} \mathscr{B} = \epsilon_0\mu_0 \frac{\partial \mathscr{E}}{\partial t}. \tag{8.34}$$

which exhibit a certain symmetry. The student should compare Maxwell's equations, in either the integral or the differential form, with the equations listed in Table 6-1 for the static field, and note the main differences introduced. In particular observe that the Faraday-Henry and Ampère-Maxwell laws provide the connection between the electric and the magnetic fields that was absent in the equations for the static fields.

Maxwell's equations are used in integral or in differential form, depending on the problem to be solved. In Chapter 11, for example, they are used for discussing electromagnetic waves.

Remembering all these equations may seem a formidable task but is not so. In the first place they have a certain symmetry that, once recognized, helps to organize them in one's mind; and continuous application produces familiarity. In the second place, more important than remembering them in detail is understanding the physical message they convey.

Maxwell's equations are compatible with the principle of relativity in that they remain invariant under a Lorentz transformation. That is, their form does not change when the coordinates x, y, z and the time t are transformed according to the Lorentz transformation (6.28) of Volume I, and the fields $\mathscr{E}$ and $\mathscr{B}$ are transformed according to Eqs. (4.22) and (4.24). The mathematical proof of this belongs in a more advanced course, and so will be omitted here.

The synthesis of electromagnetic interactions as expressed by Maxwell's equations is one of the greatest achievements in physics, and that synthesis is what places these interactions in a unique position. They are the best understood of all interactions and the only ones, so far, that can be expressed in a closed, consistent, mathematical form. This fact has been rather fortunate since much of modern civilization has been made possible because of our understanding of electromagnetic interactions, which are responsible for most of the processes, natural and resulting from technological development, that affect our daily life.

However, Maxwell's equations as they have been presented have their limitations. They work very well when dealing with electromagnetic interactions between large aggregates of charges, such as radiating antennas, electric circuits, and even beams of ionized atoms or molecules. However although the equations themselves are still correct, the electromagnetic interactions between fundamental particles (especially at high energies) must be treated somewhat differently according to the laws of quantum mechanics by a technique called *quantum electrodynamics*. Even granted these limitations, the results derived from the form of Maxwell's equations given in this chapter are an excellent approximation for describing electromagnetic interactions. This method is called *classical electrodynamics*. It is this approximate technique that is used in this book when electromagnetic waves and the structure of matter are discussed.

Problems

8.1 A coil consisting of 200 turns and having a radius of 0.10 m is placed perpendicular to a uniform magnetic field of 0.2 T. Find the emf induced in the coil if in 0.1 s (a) the field is doubled, (b) the field is reduced to zero, (c) the field is reversed in direction, (d) the coil is rotated 90°, and (e) the coil is rotated 180°. In each case make a diagram showing the direction of the emf.

8.2 Refer to Problem 6.10; if the current varies according to $I = I_0 \sin \omega t$, determine the emf induced in the circuit.

8.3 Show that if V_1 is an oscillating emf applied to terminals AB, the emf V_2 at terminals $A'B'$ as a result of the mutual induction between the two coils is $V_2 = (N_2/N_1)V_1$ (see Fig. 8-14). This is the principle of the transformer; the formula is correct as long as the magnetic flux is the same through both coils, and as long as the resistance is negligible.

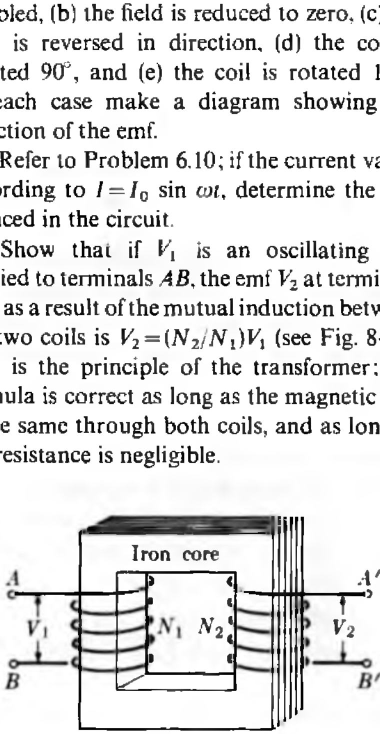


Figure 8-14

8.4 The magnetic field $\mathscr{B}$ at all points within the dashed circle of Fig. 8-15 equals 0.5 T. The field is directed into the plane of the paper and is decreasing at the rate of 0.1 T s^{-1}. (a) What is the shape of the lines of force of the induced electric field in Fig. 8-15, within the dashed circle? (b) What are the magnitude and the direction of this field at any point of the circular conducting ring, and what is the emf in the ring? (c) What is the current in the ring if its resistance is 2 ohms? (d) What is the potential difference between any two points of the ring? (e) How do you reconcile your answers to (c) and (d)? (f) If the ring is cut at some point and the ends are separated slightly, what will be the potential difference between the ends?

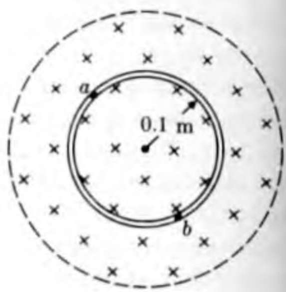


Figure 8-15

8.5 A square loop of wire is moved at constant velocity v across a uniform magnetic field confined to a square region whose sides are twice the length of those of the square loop (see Fig. 8-16). Sketch a graph of the induced emf in the loop as a function of x, from $x = -2l$ to $x = +2l$; plot clockwise emf's upward and counterclockwise emf's downward.

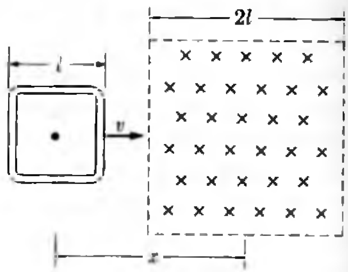


Figure 8-16

8.6 A rectangular loop is moved through a region in which the magnetic field is given by $\mathscr{B}_y = \mathscr{B}_z = 0$, $\mathscr{B}_x = (6 - y)$ T (see Fig. 8-17). Find the emf in the loop as a function of time, with $t = 0$ when the loop is in the position shown in the figure, if (a) $v = 6.5$ m s^{-1}, and (b) the loop starts at rest and has an acceleration of 2 m s^{-2}.

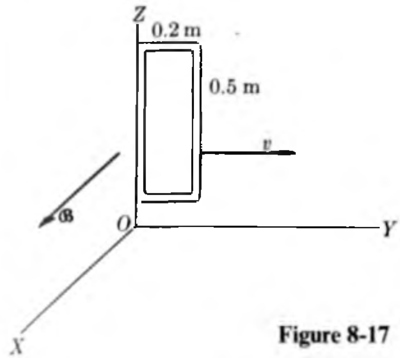


Figure 8-17

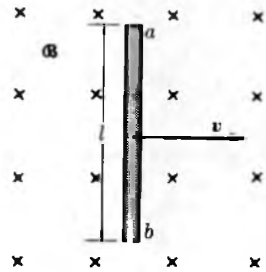


Figure 8-18

(c) Repeat for uniform motion parallel to OZ in place of OY. (d) Repeat for uniform acceleration of 2 m s^{-2} parallel to OZ.

8.7 Suppose that the loop in Problem 8.6 is pivoted about the OZ-axis and rotates at a constant rate ω. (a) Calculate the instantaneous emf as a function of time. (b) What is the average emf during the first 90° of rotation if the period of rotation is 0.2 s?

8.8 In Fig. 8-18, let $l=1.5$ m, $\mathscr{B}=0.5$ T, and $v=6\times10^4$ m s^{-1}. (a) What is the potential difference between the ends of the conductor? (b) Which end is at the higher potential?

8.9 In Fig. 8-19 the cube, one meter on a side, is in a uniform magnetic field of 0.2 T directed along the Y-axis. Wires A, C, and D move in the directions indicated, each with a velocity of 0.5 m s^{-1}. What is the potential difference between the ends of each wire?

8.10 If the rectangular circuit of Fig. 8-20 is moving away from the rectilinear current with velocity v, find the induced emf. Use two methods. (*Hint:* Remember Problem 6.10 and note that $v=dr/dt$.)

8.11 Referring to the situation discussed in Sections 8.2 and 8.4, compute the electric field in the frame of reference attached to the moving conductor, and determine the potential difference between its end points.

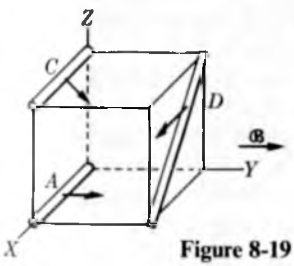


Figure 8-19

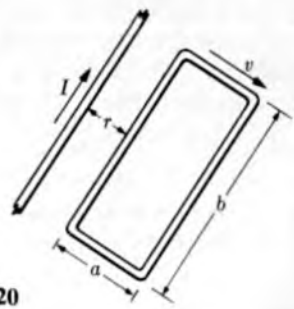


Figure 8-20

CHALLENGING PROBLEMS

8.12 A closed square wire loop with sides of length l is allowed to fall with the top portion of the loop in a uniform magnetic field $\mathscr{B}$. The magnetic field $\mathscr{B}$ is perpendicular to and directed into the plane of the paper as shown in Fig. 8-21. The loop has a resistance R and a weight mg. (a) Find the magnitude of the current in the loop when the speed of the loop is v.

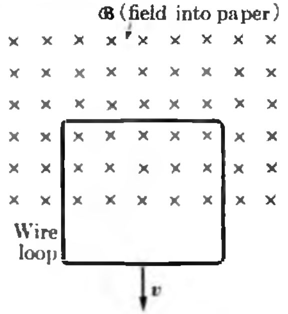


Figure 8-21

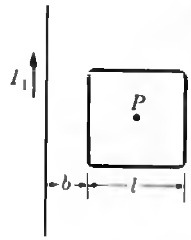


Figure 8-22

Indicate on a diagram the direction of the current using the positive current convention. (b) What is the total magnetic force on the loop? (c) At what speed v would the resultant force on the loop be zero? (AP-B: 1970)

8.13 A long straight wire carrying constant current I_1 is situated in the same plane as a square loop of wire as shown in Fig. 8-22. The long wire is parallel to one side of the loop and a distance b from it. The length of each side of the loop is l. (a) Show that the expression for the magnetic induction (field) $\mathcal{B}$ is proportional to I_1/r, where r is the perpendicular distance from the long straight wire. (b) Determine the total magnetic flux through the loop. The current I_1 now varies with time t according to the relationship $I_1 = kt$ where k is a constant. The resistance of the square loop of wire is R. (c) Determine the induced current in the loop. (d) What is the direction of the resultant magnetic force on the loop? Explain your reasoning in arriving at your answer. (AP-C; 1971)

8.14 A uniform magnetic field $\mathcal{B}$ is confined to a square region with sides of length $2L$. A square loop of wire with sides of length L is moved through the region at a constant speed v. Let $t=0$ be the time at which the loop first encounters the magnetic field. The resistance of the loop is R (Fig. 8-23). (a) Plot a graph of the induced emf $\mathcal{E}$ in the wire vs. time t. Carefully label significant points on the graph with the appropriate values of $\mathcal{E}$ and t in terms of L, v, and B. Express the answers to parts (b), (c), and (d) in terms of B, v, L, and R. (b) At time $t = L/2v$, determine the induced current I in the loop. (Neglect the self-inductance of the loop.) (c) At time $t = L/2v$, determine the magnitude of the magnetic force on each side of the loop. Indicate the directions of these forces. (d) Determine the net work performed by the external agent that moves the loop completely through the field at constant speed. (AP-B and C: 1972)

8.15 A single loop of wire fits closely around the center of a bar magnet as shown in Fig. 8-24. The cross-sectional area of the magnet

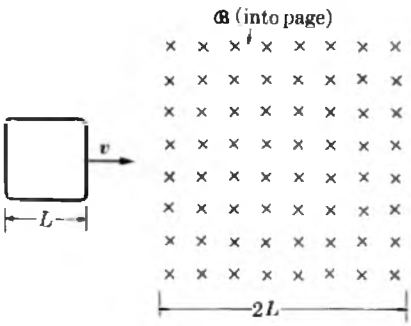


Figure 8-23

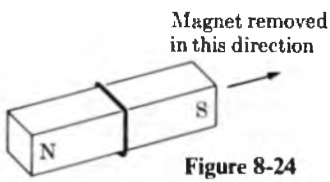


Figure 8-24

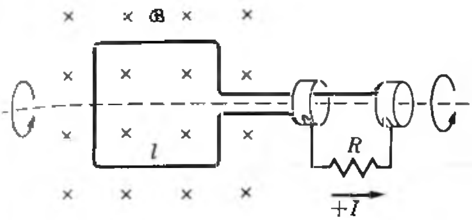


Figure 8-25

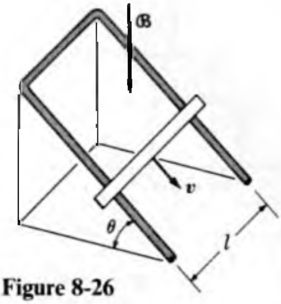


Figure 8-26

is 2.0×10^{-4} meter^2, the resistance of the loop is 0.010 ohm, and the magnetic field $\mathscr{B}$ in the magnet (assumed uniform) is 1.0×10^{-4} T. (a) The bar magnet is moved far away from the loop in the direction indicated. Determine the total charge that flows past any point in the loop. (b) Indicate on a sketch the direction of the conventional current in the loop as the magnet is removed. (c) If, instead, the magnet were removed in the opposite direction, would the current in the loop have the same direction as in part (b)? (AP-C; 1972)

8.16 In a uniform magnetic field, a square loop of side l is rotated with a frequency f in the sense shown in Fig. 8-25. The resistor R is connected to the loop by means of slip rings. At time $t=0$, the plane of the loop is in the plane of the figure. (a) Sketch a graph of the current I through R as a function of time t; take the positive direction of current as shown in the figure. (b) Calculate the maximum value of the current through R in terms of the given quantities. (AP-B; 1973)

8.17 A surveyor attempts to use a compass below a power line carrying a steady current of 10^3 ampères. The compass is 6.0 m directly below the wire. (a) If the horizontal component of the earth's field is 1×10^{-4} T, could the power line disturb the compass reading? Give a quantitative argument. (b) Suppose, instead, that the current were 10^3 ampères of 60 Hz alternating current. Would the compass reading be disturbed? Explain your answer qualitatively in terms of the properties of the compass. (AP-C; 1973)

8.18 In a uniform magnetic field $\mathscr{B}$ directed vertically downward, a metal bar of mass m is released from rest and slides without friction down a track inclined at an angle θ as shown in Fig. 8-26. The electrical resistance of the bar between its two points of contact with the track is R; the track has negligible resistance. The width of the track is l. (a) Show on a diagram the direction of the current in the sliding bar. (b) Denoting by v the instantaneous speed with which the bar is sliding down the incline, determine an expression for the magnitude of the current in the bar. (c) Determine an expression for the force exerted on the bar by the magnetic field. (d) Determine an expression for the terminal velocity of the sliding bar. (AP-C; 1973)

8.19 A small circular loop of wire with radius r is placed at the center of a large circular loop of wire with radius R. The two loops lie in the same plane, and $r \ll R$ (Fig. 8-27). In the outer loop there is a sinusoidal current $I = I_0 \sin \omega t$ where t is time and I_0 and ω are constants. Find an expression for the induced emf in the inner loop. (AP-C; 1974)

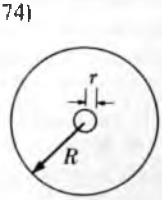


Figure 8-27

8.20 A long straight conductor lies in the plane of a rectangular loop of wire as shown in Fig. 8-28. The total resistance of the loop

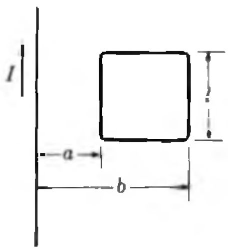


Figure 8-28

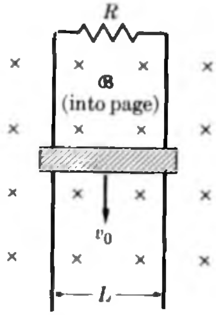


Figure 8-29

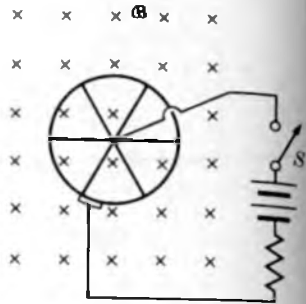


Figure 8-30

is R. The current in the long straight conductor increases at a constant rate dI/dt. (a) Indicate on a diagram the direction of the induced current in the loop and explain your reasoning. (b) Determine the magnitude of the current on the assumption that the self-inductance of the loop may be neglected. (AP-C; 1975)

8.21 A conducting bar of mass M slides without friction down two vertical conducting rails which are separated by a distance L and are joined at the top through an unknown resistance R. The bar maintains electrical contact with the rails at all times. There is a uniform magnetic field $\mathscr{B}$, directed into the page as shown in Fig. 8-29. The bar is observed to fall with a constant terminal speed v_0. (a) On a diagram, draw and label all the forces acting on the bar. (b) Determine the magnitude of the induced current I in the bar as it falls with constant speed v_0 in terms of B, L, g, v_0, and M. (c) Determine the voltage induced in the bar in terms of B, L, g, v_0, and M. (d) Determine the resistance R in terms of B, L, g, v_0, and M. (AP-C; 1976)

8.22 A wheel with six spokes is positioned perpendicular to a uniform magnetic field $\mathscr{B}$ of magnitude 0.5 tesla (weber per square meter). The field is directed into the plane of the paper and is present over the entire region of the wheel as shown in Fig. 8-30. When the switch S is closed, there is an initial current of 6 ampères between the axle and the rim; and the wheel begins to rotate. The resistance of the spokes and the rim may be neglected. (a) What is the direction of rotation of the wheel? Explain. (b) The radius of the wheel is 0.2 m. Calculate the initial torque on the wheel. (c) Describe qualitatively the angular velocity of the wheel as a function of time. (AP-C; 1977)

8.23 Two parallel conducting rails, separated by a distance L of 2 m, are connected through a resistance R of 3 ohms as shown in Fig. 8-31. A uniform magnetic field with a magnitude B of 2 tesla points into the page. A conducting

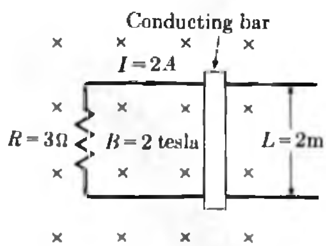


Figure 8-31

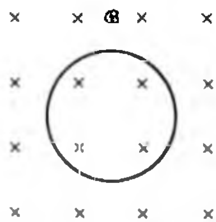


Figure 8-32

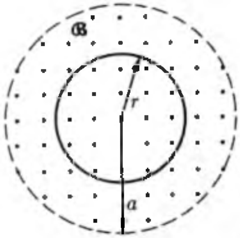


Figure 8-33

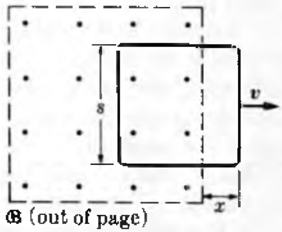


Figure 8-34

bar with mass m of 4 kilograms can slide without friction across the rails. (a) Determine at what speed the bar must be moved and in what direction to induce a counterclockwise current I of 2 ampères as shown. (b) Determine the magnitude and direction of the external force that must be applied to the bar to keep it moving at this velocity. (c) Determine the rate at which heat is being produced in the resistor and determine the mechanical power being supplied to the bar. (d) Suppose the external force is suddenly removed from the bar. Determine the energy in joules dissipated in the resistor before the bar comes to rest. (AP-B; 1978).

8.24 A circular loop of wire of area A and electrical resistance R is placed in a spatially uniform magnetic field $\mathcal{B}$ directed into the page and perpendicular to the plane of the loop as shown in Fig. 8-32. The magnetic field is gradually reduced from an initial value of B_0, in such a way that the magnetic-field strength as a function of time is $B(t)=B_0e^{-\alpha t}$. (a) Indicate on a diagram the direction of the induced current. Applying the fundamental relation for electromagnetic induction, explain your choice. (b) Do the electromagnetic forces on this current tend to make the loop expand or contract? Explain. (c) Determine an expression, in terms of B_0, A, and R, that describes the total quantity of charge that flows past a point in the loop during the time the magnetic field is reduced from B_0 to zero. (d) Determine an expression for the amount of energy dissipated as heat in the loop, in terms of B_0, A, R, and α, during the time the magnetic field is reduced from B_0 to zero. (AP-C; 1978)

8.25 A spatially uniform magnetic field directed out of the page is confined to a cylindrical region of space of radius a as shown in Fig. 8-33. The strength of the magnetic field increases at a constant rate such that $B=B_0+Ct$ where B_0 and C are constants and t is time. A circular conducting loop of radius r and resistance R is placed perpendicular to the magnetic field. (a) Indicate on a diagram the direction of the induced current in the loop. Explain your choice. (b) Derive an expression for the induced current in the loop. (c) Derive an expression for the magnitude of the induced electric field at any radius $r<a$. (d) Derive an expression for the magnitude of the induced electric field at any radius $r>a$. (AP-C; 1980)

8.26 A square loop of wire of side s and resistance R is pulled at constant velocity $\boldsymbol{v}$ out of a uniform magnetic field of intensity B. The plane of the loop is always perpendicular to the magnetic field. After the leading edge of the loop has passed the edge of the B field as shown in Fig. 8-34, there is an induced current in the loop. (a) On a figure, indicate the direction of this induced current. (b) Using Faraday's law of induction, develop an expression for the induced emf $\mathcal{E}$ in the loop. (c) Determine the induced current I in the loop. (d) Determine the power required to keep the loop moving at constant velocity. (AP-C; 1981)

8.27 Show that Eq. (8.8) is satisfied if $\mathcal{B}=C/r$. (*Hint:* Compute $\mathcal{B}_{\text{ave}}$ for an arbitrary r, insert the value in Eq. (8.8), and compute the derivative with respect to r.)

8.28 A charge q, of mass m, is moving in a circular orbit of radius ρ under a centripetal force F. In a certain time interval, a uniform magnetic field is set up in a direction perpendicular to the plane of the orbit. Using the law of electromagnetic induction, show that the change in the magnitude of the velocity of the ion is $\Delta v = -q\rho\mathscr{B}/2m$, and that the corresponding change in magnetic moment is $\Delta \boldsymbol{m} = -(q^2\rho^2/4m)\boldsymbol{\mathscr{B}}$. Compare with Example 6.4. (*Hint:* To obtain the tangential acceleration while the magnetic field is changing, use Eq. (8.6), derived in discussing the betatron.)

8.29 A metallic disk of radius a rotates with angular velocity ω in a plane in which there is a uniform magnetic field parallel to the disk axis (see Fig. 8-35). Show that the potential difference between the center and the rim is $\frac{1}{2}\omega a^2 \mathscr{B}$.

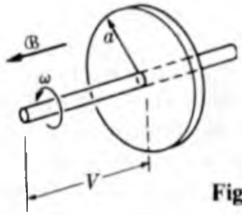


Figure 8-35

8.30 Referring to the situation described in Section 8.5, (a) show that in the frame of reference in which the circuit is at rest and the magnetic field rotates with angular velocity $-\boldsymbol{\omega}$, $\partial\boldsymbol{\mathscr{B}}/\partial t = -\boldsymbol{\omega} \times \boldsymbol{\mathscr{B}}$. (b) Write Eq. (8.15) with this value of $\partial\boldsymbol{\mathscr{B}}/\partial t$; and using the result of Problem 16.20, show that the electric field observed in this frame of reference is $\boldsymbol{\mathscr{E}} = \frac{1}{2}(\boldsymbol{\omega} \times \boldsymbol{\mathscr{B}}) \times \boldsymbol{r}$. (c) Show that the emf produced by this electric field is the same as the emf measured by the observer attached to the magnetic field. (*Hint:* Note that $\frac{1}{2}\boldsymbol{r} \times d\boldsymbol{l}$ is the area of the triangle determined by both vectors, and that $\boldsymbol{A} \times \boldsymbol{B} \cdot \boldsymbol{C} = \boldsymbol{A} \cdot \boldsymbol{B} \times \boldsymbol{C}$.)

8.31 In a region in which a uniform magnetic field $\boldsymbol{\mathscr{B}}$ exists, the magnitude of the field is increasing at a constant rate; that is, $\partial\boldsymbol{\mathscr{B}}/\partial t = \boldsymbol{b}$ where $\boldsymbol{b}$ is a constant vector parallel to $\boldsymbol{\mathscr{B}}$. (a) Show that according to Eq. (8.15), the electric field at each point is $\boldsymbol{\mathscr{E}} = -\frac{1}{2}\boldsymbol{b} \times \boldsymbol{r}$. (b) Placing the Z-axis parallel to the magnetic field, obtain the rectangular components of $\boldsymbol{\mathscr{E}}$. (c) Plot the lines of force of the magnetic and the electric fields.

8.32 Find the electric flux through a sphere concentric with a charge moving with a high velocity. (*Hint:* Use Eq. (4.27) in Gauss's law.)

8.33 By using the operator ∇, write the Maxwell equations in differential form (Table 8-1). (See Problem 6.17.)

8.34 Show that the equation of continuity (8.18) expressed in differential form is $\partial\rho/\partial t = -\operatorname{div} \boldsymbol{j}$.

8.35 Show that in order for the equation of continuity as written in Problem 8.34 to remain invariant for all inertial observers under a Lorentz transformation, it is necessary that the current and charge density transform according to the law

$$j'_x = \frac{j_x - \rho v}{\sqrt{1 - v^2/c^2}}, \qquad j'_y = j_y,$$

$$j'_z = j_z, \qquad \rho' = \frac{\rho - j_x v/c^2}{\sqrt{1 - v^2/c^2}}.$$

Write the nonrelativistic limit of these expressions and discuss their plausibility. (*Hint:* Recall that $\boldsymbol{j} = \rho \boldsymbol{v}$ is the current density for charges moving with velocity $\boldsymbol{v}$.)

CHAPTER NINE

TIME-DEPENDENT ELECTRIC CIRCUITS

9.1 Introduction

The circuits considered in Chapter 3 were such that neither the applied emf nor the current varied with time. In these circuits the current was constant in magnitude and direction; for this reason these circuits are sometimes called direct-current circuits. The only parameter that needs to be considered in such circuits is the resistance of the different elements that compose the circuit. In many instances however, the applied emf and the electric current vary with time, and then new effects must be taken into account. These effects are in general related to the phenomenon of electromagnetic induction. In such cases the simple relation $V=RI$ (where V is the applied emf, R is the resistance, and I is the current) corresponding to Ohm's law must be modified to take the induction effects into account.

9.2 Self-Induction

Consider a circuit carrying a current I (Fig. 9-1). According to Ampère's law the current produces a magnetic field that is proportional to I at each point. The magnetic flux through the circuit produced by its own magnetic field is called the *self-flux*. This magnetic flux is then proportional to the current I and may be written

$$\Phi_m = LI. \tag{9.1}$$

The coefficient L depends on the geometric shape of the conductor and is called the *self-inductance* of the circuit. Self-inductance is expressed in Wb A^{-1}, a unit called the *henry* (H) in honor of the American scientist Joseph Henry (1797–1878). That is, $H = Wb\,A^{-1} = m^2\,kg\,C^{-2}$.

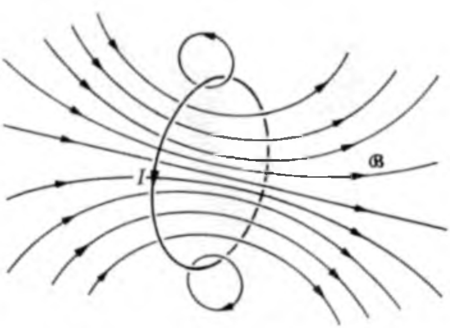


Fig. 9-1. Self flux in a circuit.

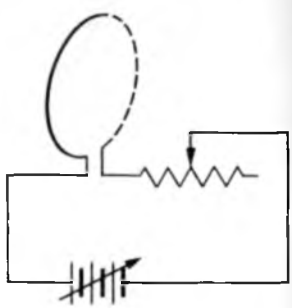


Figure 9-2

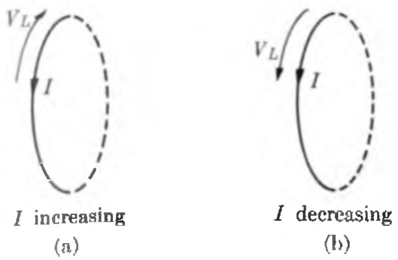


Fig. 9-3. Direction of the self-induced emf in a circuit.

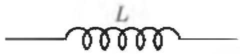


Fig. 9-4. Representation of a self-inductance.

Suppose that the electric current in a circuit is not constant. For example the current may be varied by varying either the emf applied to the circuit or the electric resistance of the circuit, or by varying both (Fig. 9-2). When the current I changes with time, the magnetic flux through the circuit also changes; and according to the law of electromagnetic induction, an emf is induced in the circuit. This special case of electromagnetic induction is called *self-induction*. Combining Eqs. (8.1) and (9.1) gives the self-induced emf

$$V_L = -\frac{d\Phi_m}{dt} = -L\frac{dI}{dt}. \tag{9.2}$$

The minus sign indicates that V_L is opposed to the change in the current. Thus if the current increases, dI/dt is positive and V_L is opposed to the current (Fig. 9-3a). If the current decreases, dI/dt is negative and V_L acts in the same direction as the current (Fig. 9-3b). Therefore V_L always acts in a direction that opposes the *change* in the current. When Eq. (9.2) was written, the circuit was assumed rigid and therefore L was considered constant when the time derivative was computed. If the shape of the circuit is variable, L is not constant; and instead of Eq. (9.2), the self-induced emf is

$$V_L = -\frac{d}{dt}(LI). \tag{9.3}$$

In diagrams drawn to indicate that a conductor has an appreciable inductance, the symbol of Fig. 9-4 is used. However, note that the self-inductance of a circuit is not concentrated at a particular point, but is a property of the whole circuit.

Example 9.1. Establishment of a current in a circuit.

▼ When an emf V is applied to a circuit by closing a switch (Fig. 9-5), the current does not instantaneously attain the value V/R corresponding to Ohm's law, but increases gradually and steadily approaches the value given by Ohm's law. This process is due to the self-induced emf V_L, which opposes the change in the current and is present while the current increases from zero up to the

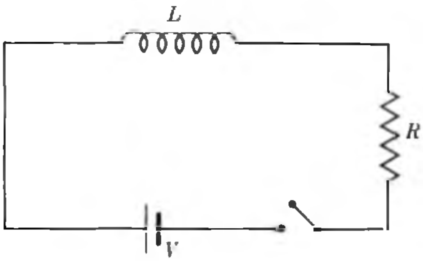


Fig. 9-5. Electric circuit containing a resistance and a self-inductance.

final constant value. The total emf in the circuit is then $V+V_L=V-L(dI/dt)$. Ohm's law is now

$$RI=V+V_L \qquad \text{or} \qquad RI=V-L(dI/dt). \tag{9.4}$$

The new feature of this equation is that it relates the current in the circuit to the time rate of change of the current. By writing Eq. (9.4) in the form

$$L\left(\frac{dI}{dt}\right)=V-RI, \tag{9.5}$$

the student may see that as the current increases, the time rate of change of the current decreases until it becomes zero when the current reaches the value V/R associated with Ohm's law. To show how the current varies with time, Eq. (9.5) may be written as

$$L\left(\frac{dI}{dt}\right)=-R\left(I-\frac{V}{R}\right),$$

or with the variables I and t separated,

$$\frac{dI}{I-V/R}=-\frac{R}{L}\,dt.$$

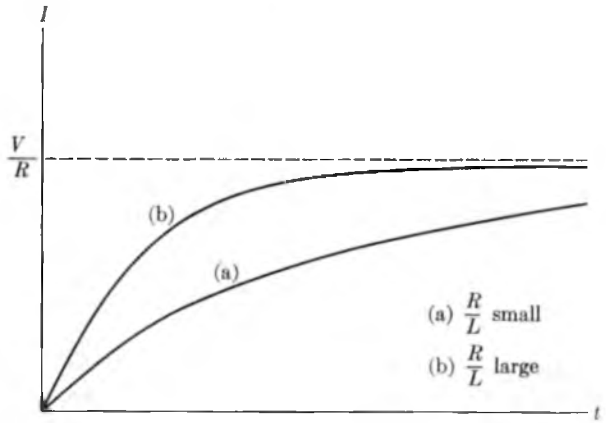


Fig. 9-6. Growth of the current in an inductive circuit.

Note that at $t=0$ the current is also zero ($I=0$); then

$$\int_0^I \frac{dI}{I-V/R} = -\frac{R}{L}\int_0^t dt,$$

and integration gives

$$\ln(I-V/R)-\ln(-V/R) = -(R/L)t.$$

Using the identity $\ln e^x = x$, the current in the circuit is

$$I=\frac{V}{R}(1-e^{-Rt/L}). \tag{9.6}$$

The second term in the parentheses decreases with time, and the current asymptotically approaches the value V/R that is given by Ohm's law (Fig. 9-6). If R/L is large, the current reaches this value very fast; but if R/L is small, it may take a long time before the current stabilizes. The student may recognize the mathematical similarity between Eq. (9.5) and the expression for the motion of a body through a viscous fluid; the latter was given in Section 7.9 of Volume I as $m(dv/dt) = F - K\eta v$. The following correspondences may be seen: $V \leftrightarrow F$, $L \leftrightarrow m$, and $R \leftrightarrow K\eta$. ▲

Example 9.2. Decay of the current in the circuit of Fig. 9-7 when the switch is moved from position 1 to position 2.

▼ Assume that the switch has been in position 1 for a very long time so that the current in the circuit has achieved its limiting (or steady) value V/R. Moving the switch over to position 2 removes the applied emf without actually opening the circuit. The only emf that remains is $V_L = -L\,dI/dt$, and Ohm's law for the circuit becomes

$$RI = -L\frac{dI}{dt} \quad \text{or} \quad \frac{dI}{I} = -\frac{R}{L}dt.$$

If time is set at 0 from the instant that V is removed from the circuit, the initial current is V/R. Integrating gives

$$\int_{V/R}^I \frac{dI}{I} = -\frac{R}{L}\int_0^t dt$$

or

$$\ln I - \ln(V/R) = -(R/L)t.$$

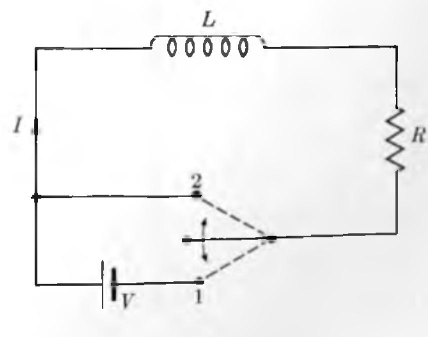


Fig. 9-7. Device for removing the emf applied to a circuit without changing the resistance.

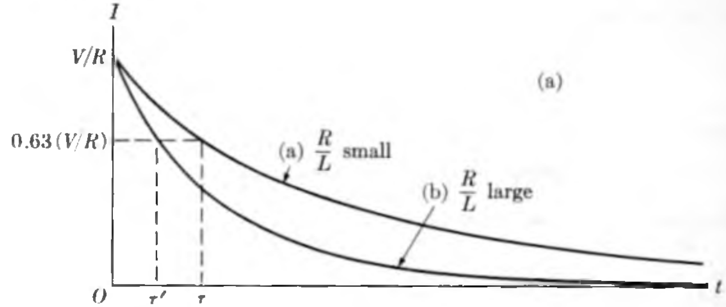


Fig. 9-8. Decay of the current in a circuit after the emf has been removed.

Removing logarithms results in

$$I = (V/R)e^{-Rt/l}. \tag{9.7}$$

The current decreases exponentially as shown in Fig. 9-8. The larger the resistance R or the smaller the inductance L, the faster is the drop in the current. The time required for the current to drop to $1/e$, approximately 63% of its initial value, is $\tau = L/R$. This time is called *relaxation time*. ▲

Example 9.3. Self-inductance of coaxial cylinders.

▼ Suppose a circuit is composed of two coaxial, cylindrical, metallic sheets of radii a and b, each sheet carrying a current I, but in the opposite direction (Fig. 9-9). The space between the cylinders is filled with a substance whose permeability is μ.

In Example 6.1 the magnetic field for this current arrangement was computed as $\mathscr{B} = \mu I/2\pi r$ in the region within the two cylinders, and zero elsewhere. The vacuum permeability μ_0 used in Example 6.1 has been replaced by μ, the permeability of the medium filling the space within the two cylinders. To calculate the self-inductance, the magnetic flux through any section of the conductor must be computed. Consider the section $PQRS$, having a length l. If we divide this

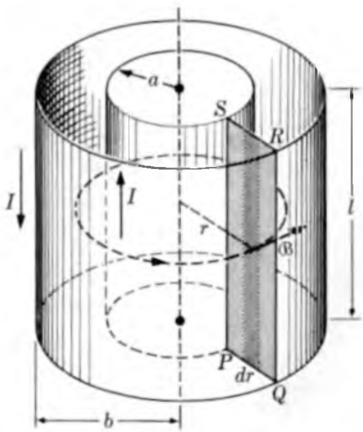


Figure 9-9

section into strips of width dr, the area of each strip is $l\,dr$. The magnetic field $\mathscr{B}$ is perpendicular to $PQRS$. Therefore

$$\Phi_m = \int_{PQRS} \mathscr{B}\,dS = \int_a^b \left(\frac{\mu I}{2\pi r}\right)(l\,dr)$$

$$= \frac{\mu I l}{2\pi}\int_a^b \frac{dr}{r} = \frac{\mu I l}{2\pi}\ln\frac{b}{a}.$$

Therefore the self-inductance of a portion of length l is

$$L = \frac{\Phi_m}{I} = \frac{\mu l}{2\pi}\ln\frac{b}{a}, \tag{9.8}$$

and the self-inductance per unit length will be $(\mu/2\pi)\ln b/a$. ▲

9.3 Energy of the Magnetic Field

In Section 3.2 it was seen that to maintain a current in a circuit, energy must be supplied. The energy required per unit time (in other words, the power) is VI. Now Eq. (9.4) may be written in the form

$$V = RI + L\frac{dI}{dt}.$$

Multiplying this equation by I, we have

$$VI = RI^2 + LI\frac{dI}{dt}. \tag{9.9}$$

According to Eq. (3.10), the term RI^2 is the energy spent in moving the electrons through the crystal lattice of the conductor and is transferred to the ions that make up the lattice. The last term in Eq. (9.9) is then interpreted as the energy required per unit time to build up the current or to establish its associated magnetic field in space. Therefore the rate of increase of the magnetic energy is

$$\frac{dE_m}{dt} = LI\frac{dI}{dt}.$$

The magnetic energy required to increase a current from zero to the value I is thus

$$E_m = \int_0^E dE_m = \int_0^I LI\,dI = \tfrac{1}{2}LI^2. \tag{9.10}$$

For example in the circuit of Example 9.3 the magnetic energy of a section of length l is from Eq. (9.8)

$$E_m = \frac{1}{2}\left(\frac{\mu l}{2\pi}\ln\frac{b}{a}\right)I^2 = \frac{\mu l I^2}{4\pi}\ln\frac{b}{a}. \tag{9.11}$$

The magnetic energy E_m can also be calculated by using the expression

$$E_m = \frac{1}{2\mu} \int \mathscr{B}^2 \, dv \tag{9.12}$$

where the integral extends throughout all the volume in which the magnetic field exists, and dv is a volume element. For example in the case of the circuit of Fig. 9-9, which has been redrawn in Fig. 9-10, the magnetic field is given by $\mathscr{B} = \mu I/2\pi r$. When the volume element is a cylindrical shell of radius r and thickness dr, the volume is $dv = (2\pi r)l\, dr$. Substituting in Eq. (9.12) and remembering that the magnetic field extends only from $r = a$ to $r = b$ yield

$$E_m = \frac{1}{2\mu} \int_a^b \left(\frac{\mu I}{2\pi r}\right)^2 (2\pi l r\, dr) = \frac{\mu l I^2}{4\pi} \int_a^b \frac{dr}{r} = \frac{\mu l I^2}{4\pi} \ln \frac{b}{a}.$$

Expression (9.12) may be interpreted by saying that the energy spent in establishing the current has been *stored* in the surrounding space so that an energy $(\mathscr{B}^2/2\mu)\, dv$ corresponds to a volume dv; and the energy per unit volume E_m stored in the magnetic field is

$$\mathrm{E}_m = \frac{1}{2\mu} \mathscr{B}^2. \tag{9.13}$$

Although expression (9.12) has been justified for the magnetic energy density by using a circuit of very special symmetry, a more detailed analysis, not given here, would indicate that the result is completely general. When both electric and magnetic fields are present, the electric energy density given by Eq. (2.40) must also be considered; and thus the total energy per unit volume in the electromagnetic field is

$$\mathrm{E} = \tfrac{1}{2}\epsilon \mathscr{E}^2 + \frac{1}{2\mu} \mathscr{B}^2. \tag{9.14}$$

Example 9.4. Energy of the magnetic field of a slowly moving electron.

▼ From Section 4.5 a slowly moving charge was shown to produce a magnetic field whose lines of force are circles perpendicular to the direction of motion and whose magnitude is obtained from Eq. (4.16) as

$$\mathscr{B} = \frac{\mu_0}{4\pi} q \frac{v \sin\theta}{r^2}$$

with $q = -e$ for an electron. Suppose that the crude model of the electron introduced in Example 2.15, in which R is the "radius" of the electron, is used. The energy of the magnetic field *exterior* to the charge is obtained by using Eq. (9.12) with the integral extended over all space *outside* the charge. An appropriate volume element, illustrated by the ring in Fig. 9-11, has a perimeter equal to $2\pi r \sin\theta$, and a cross section with sides dr and $r\, d\theta$, and therefore an area $r\, dr\, d\theta$. The volume of the ring is

$$dv = \text{perimeter} \times \text{cross section} = 2\pi r^2 \sin\theta\, dr\, d\theta.$$

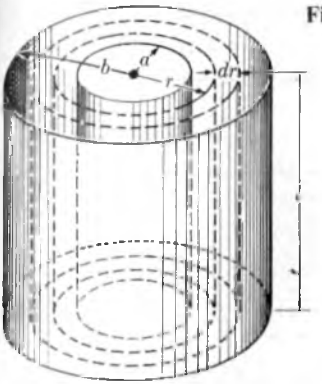


Figure 9-10

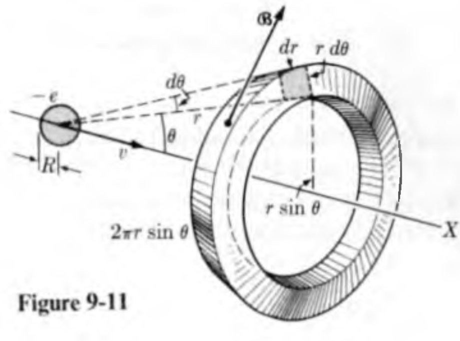


Figure 9-11

Therefore Eq. (9.12) gives

$$E_m = \frac{1}{2\mu_0} \int_R^\infty \int_0^\pi \left(\frac{\mu_0}{4\pi} \frac{qv \sin\theta}{r^2} \right)^2 2\pi r^2 \sin\theta \, dr \, d\theta$$

$$= \frac{\mu_0}{16\pi} q^2 v^2 \int_R^\infty \frac{dr}{r^2} \int_0^\pi \sin^3\theta \, d\theta = \frac{1}{2} \left(\frac{\mu_0}{4\pi} \frac{2q^2}{3R} \right) v^2.$$

This result gives only an order of magnitude for the total magnetic energy because the contribution from the magnetic field *inside* the charged particle must be added, and therefore the charge distribution inside the particle must be known. The most interesting feature of E_m is that it depends on v^2 and therefore resembles the kinetic energy of a particle whose mass is

$$m = \frac{\mu_0}{4\pi} \frac{2q^2}{3R}.$$

In the case of the electron, $q = -e$ and $m = m_e$ so that

$$m_e = \frac{\mu_0}{4\pi} \frac{2e^2}{3R} = \frac{1}{4\pi\epsilon_0} \frac{2e^2}{3Rc^2}$$

where Eq. (4.18) has been used to eliminate μ_0. Solving for R yields

$$R = \frac{2}{3} \left(\frac{e^2}{4\pi\epsilon_0 m_e c^2} \right) = \tfrac{2}{3} r_e$$

where r_e is the radius of the electron as defined in Eq. (2.45). That our rough calculation gives a result of the same order of magnitude as in Example 2.15, in which $R = \frac{3}{5} r_e$, is a proof of the consistency of the theory since only the order of magnitude can be estimated. When the present result is combined with that of Example 2.15, it seems plausible to think that the rest energy of a charged particle is associated with the energy of its electric field, and that the kinetic energy corresponds to the energy of the magnetic field. However, it is logical to think that the fields associated with the other interactions existing in nature also contribute to the rest and kinetic energies of a particle. However, our incomplete knowledge of those interactions makes it impossible to state definitely that this is the case. In fact, the calculations considered both in Example 2.15 and here are what are known as the determinations of the *self energy* of the electron. ▲

9.4 Free Electrical Oscillations in a Circuit

As shown previously three parameters characterize the flow of electricity through an electric circuit: the capacitance C, the resistance R, and the self-inductance L. We shall now analyze the way in which the three together determine the current produced by a given emf in a closed circuit. If the current I in the circuit of Fig. 9-12(a) is in the direction indicated, charges q and $-q$ appear on the plates of the capacitor C such that

$$I=\frac{dq}{dt}. \tag{9.15}$$

These charges produce an emf $V_C=-q/C$. The minus sign appears because the emf opposes the current I as a result of the tendency of the capacitor to discharge through the circuit. At the inductance L there is another emf equal to $V_L=-L(dI/dt)$ according to Eq. (9.2). [In addition there may be applied to the circuit some other emf, such as V, shown in Fig. 9-12(b).]

Consider the situation in which only the two emf's V_L and V_C are present. The current in this case has its origin in the charging of the capacitor or in varying the magnetic flux through the inductance or by inserting (and later on removing) an external emf before closing the circuit loop. Therefore, applying Ohm's law, Eq. (3.1), gives

$$RI=V_L+V_C \qquad \text{or} \qquad RI=-L\frac{dI}{dt}-\frac{q}{C}. \tag{9.16}$$

Taking the derivative of the whole equation with respect to t gives

$$R\frac{dI}{dt}=-L\frac{d^2I}{dt^2}-\frac{1}{C}\frac{dq}{dt}.$$

Using Eq. (9.15) and writing all terms on the left-hand side of the equation yield

$$L\frac{d^2I}{dt^2}+R\frac{dI}{dt}+\frac{1}{C}I=0. \tag{9.17}$$

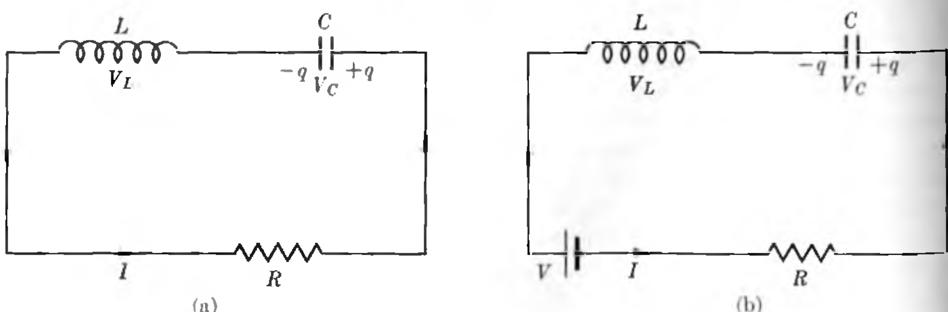


Figure 9-12

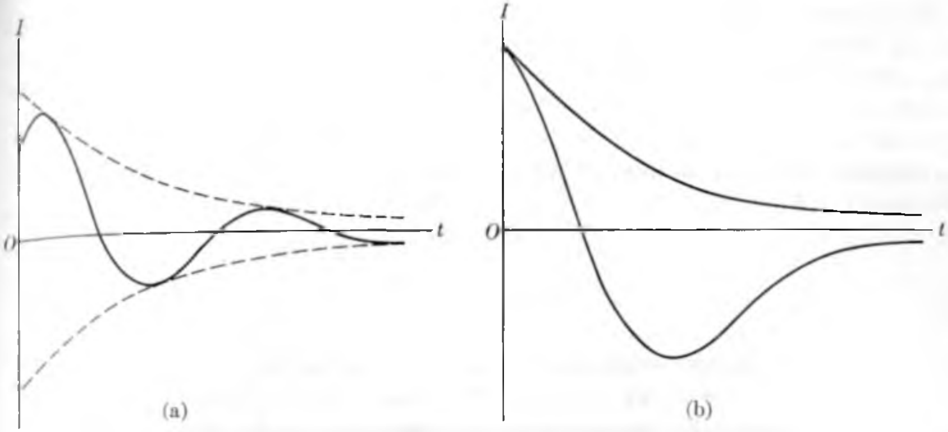


Fig. 9-13. Variation of the current of a discharging capacitor as a function of time: (a) when $R^2 < 4L/C$; (b) when $R^2 > 4L/C$.

This expression is a differential equation whose solution gives the current I as a function of t. The parameters L, R, and C characterize the circuit.

Now this equation is formally identical to that corresponding to the damped oscillations of a particle; the latter equation [see Eq. (12.45) of Volume I] is

$$m\frac{d^2x}{dt^2}+\lambda\frac{dx}{dt}+kx=0$$

if the following correspondences are established: $L \leftrightarrow m$, $R \leftrightarrow \lambda$, $1/C \leftrightarrow k$. Therefore the description of the motion of a particle can formally be applied in this case. Assume that $R^2 < 4L/C$ and introduce the quantities

$$\gamma=\frac{R}{2L}, \qquad \omega=\sqrt{\frac{1}{LC}-\frac{R^2}{4L^2}}. \tag{9.18}$$

The current as a function of time is then given by the expression

$$I=I_0e^{-\gamma t}\sin(\omega t+\alpha) \tag{9.19}$$

as the student may verify by direct substitution of I and its derivatives into Eq. (9.17). The graph of the current versus time has been given in Fig. 9-13(a). An oscillatory or alternating current is established whose amplitude decreases with time when $R < \sqrt{4L/C}$. When the resistance R is very small compared with the inductance L, both γ and the last term in the expression for ω may be neglected with the result that $I = I_0 \sin(\omega t+\alpha)$ so that the electric oscillations are undamped and have a frequency

$$\omega_0=\sqrt{\frac{1}{LC}}. \tag{9.20}$$

This expression is called the *characteristic frequency* of an LC circuit, and is equivalent to the frequency $\omega_0 = \sqrt{k/m}$ for an undamped oscillator. Note that the damping in an electric circuit results from the dissipation of energy in the resistance R. These oscillations in which no external emf is applied are the *free* oscillations of the circuit.

If the resistance is large enough so that $R^2/4L^2 > 1/LC$ or $R^2 > 4L/C$, the frequency ω becomes imaginary. In this case the current decreases gradually without oscillating as shown in Fig. 9-13b. The lower curve in Fig. 9-13b is the special case in which $R^2 = 4L/C$ and is called the *critically damped* circuit.

9.5 Forced Electrical Oscillations in a Circuit

Forced electric oscillations are produced when to the circuit depicted in Fig. 9-12 is added an alternating emf of the form $V = V_0 \sin \omega t$ as shown in Fig. 9-14. In this case Eq. (9.16) now has the form

$$RI = V_L + V_C + V_0 \sin \omega t.$$

Repeating the procedure used to obtain Eq. (9.17), differentiate with respect to time and arrange the terms as

$$L\frac{d^2I}{dt^2} + R\frac{dI}{dt} + \frac{I}{C} = \omega V_0 \cos \omega t. \tag{9.21}$$

This equation is very similar to that for forced oscillations of a particle [see Eq. (12.50) in Volume I] with an important difference: the frequency ω appears as a factor in the right-hand side of Eq. (9.21). The reason is that because of the relation $I = dq/dt$, the current in an electric circuit corresponds to the velocity $v = dx/dt$ in the motion of a particle. It is reasonable to assume that the current will oscillate with the same frequency as that of the applied emf. Therefore the current that satisfies Eq. (9.21) is given by

$$I = I_0 \sin(\omega t - \alpha) \tag{9.22}$$

where α is the *phase lag* of the current with respect to the applied emf. Substituting Eq. (9.22) into Eq. (9.21) gives the current amplitude as

$$I_0 = \frac{V_0}{\sqrt{R^2 + (\omega L - 1/\omega C)^2}}; \tag{9.23}$$

the phase difference α between the current and the applied emf is obtained from

$$\tan \alpha = \frac{\omega L - 1/\omega C}{R}, \tag{9.24}$$

which the student should compare with Eq. (12.53) of Volume I.

Table 9-1. Correspondence between a Damped Oscillator and an Electric Circuit

Oscillator	Electric circuit
Mass, m	Inductance, L
Damping, λ	Resistance, R
Elastic constant, k	Inverse capacitance, $1/C$
Displacement, x	Charge, q
Velocity, $v=dx/dt$	Current, $I=dq/dt$
Applied force, F_0	Applied emf, V_0

The *impedance* of the electric circuit is defined as V_0/I_0 or

$$Z=\sqrt{R^2+(\omega L-1/\omega C)^2}. \tag{9.25}$$

The *reactance* of the circuit is defined

$$X=\omega L-1/\omega C, \tag{9.26}$$

so that

$$Z=\sqrt{R^2+X^2} \tag{9.27}$$

and

$$\tan\alpha=\frac{X}{R}. \tag{9.28}$$

The quantities Z, R, X, and α are related as shown in Fig. 9-15. Note that both the reactance and the impedance are expressed in ohms. For example the term ωL expressed in terms of the fundamental units gives $\mathrm{s^{-1}\,H=m^2\,kg\,s^{-1}\,C^{-2}}$, the same expression obtained in Section 2.2 for the ohm. [The student can make the same verification for the term $1/\omega C$.] If R and X are expressed in ohms, then in view of its definition (9.27) Z must also be expressed in ohms.

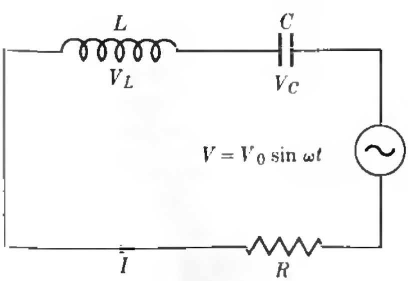


Figure 9-14

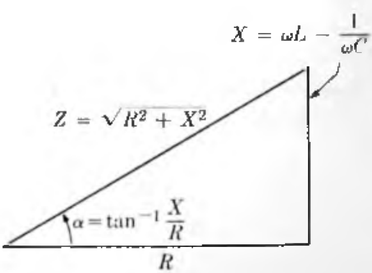


Fig. 9-15. Relation between the magnitudes of the resistance, the reactance, and the impedance.

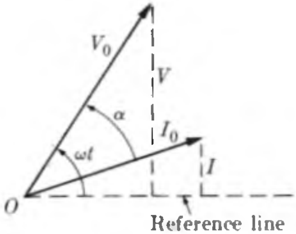


Fig. 9-16. Rotating vectors of the current and the emf in an ac circuit.

The emf V and the current I can be represented by rotating vectors as illustrated in Fig. 9-16. The components of the vectors normal to the reference line are the instantaneous values of V and I. The current I lags or leads the emf according to whether α is positive or negative, or ωL is larger or smaller than $1/\omega C$. Figure 9-17 gives the plot of V and I versus time for the case where I lags V by α.

The power required to maintain the current is

$$\begin{aligned} P = VI &= V_0 I_0 \sin \omega t \sin (\omega t - \alpha) \\ &= V_0 I_0 (\sin^2 \omega t \cos \alpha - \sin \omega t \cos \omega t \sin \alpha). \end{aligned} \tag{9.29}$$

To obtain the average power required to maintain the current, note that over a single cycle or over a long length of time $(\sin^2 \omega t)_{ave} = \frac{1}{2}$ and $(\sin \omega t \cos \omega t)_{ave} = 0$. Therefore

$$P_{ave} = \tfrac{1}{2} V_0 I_0 \cos \alpha = \tfrac{1}{2} R I_0^2. \tag{9.30}$$

Resonance is obtained when P_{ave} is maximum, which occurs when $\alpha = 0$; that is, when $\omega L = 1/\omega C$, corresponding to a frequency $\omega = \sqrt{1/LC}$, equal to Eq. (9.20). At resonance the current has maximum amplitude and is in phase with the emf; the result is maximum average power. The rotating vectors V and I are in phase or superposed (Fig. 9-18a): and the current and emf vary with time as shown in Fig. 9-18b.

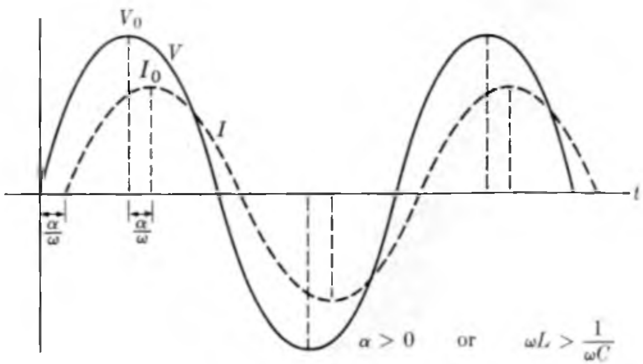


Fig. 9-17. Variation of current and emf as a function of time in an ac circuit.

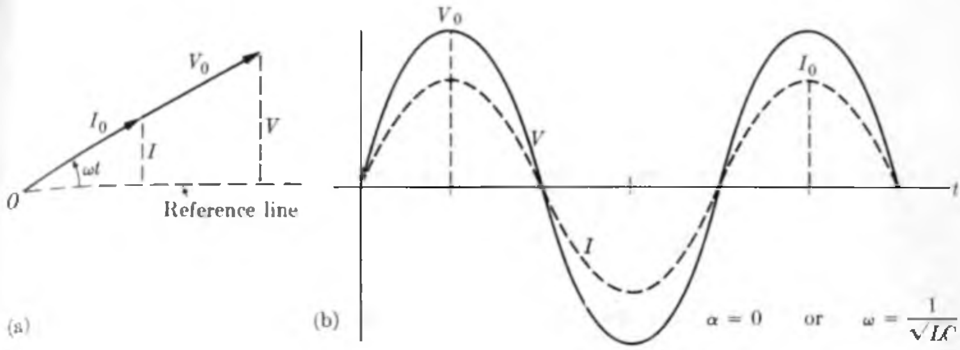


Fig. 9-18. Relation between emf and current when the phase difference is zero (resonance).

As in the case of the forced oscillations of a particle, the general solution of Eq. (9.21) is the sum of Eq. (9.22) and a transient current given by Eq. (9.19). However because of the resistor in the circuit, the term corresponding to Eq. (9.19) quickly becomes negligible and only Eq. (9.22) need be taken into account. Nevertheless when some modification, such as a variation in L, C, or R, occurs in the circuit, the transient term does appear for a short time until the circuit adjusts to the new conditions.

Example 9.5. Discussion of an alternating-current circuit in terms of the rotating-vector technique.

▼ The results stated in Section 9.5 can be derived very easily by means of the technique of rotating vectors. Note that the equation of the circuit can be written in the form

$$V_0 \sin \omega t = RI - V_L - V_C = RI + L\frac{dI}{dt} + \frac{q}{C}.$$

In the same way that RI is the potential difference across the resistance R, $L(dI/dt)$ and q/C are the respective potential differences (or voltage drops) across the inductance and the capacitance.

If it is assumed that $I = I_0 \sin(\omega t - \alpha)$, the rotating vector of the current lags that of the emf by the angle α (Fig. 9-19). Now consider that the rotating vector of the emf is the sum of the rotating vectors corresponding to the three terms on the right in the equation above. Note that $dI/dt = \omega I_0 \cos(\omega t - \alpha)$ and $q = \int I\, dt = -(1/\omega) I_0 \cos(\omega t - \alpha)$. Therefore

Potential drop across the resistance:

$$RI = RI_0 \sin(\omega t - \alpha), \text{ in phase with } I.$$

Potential drop across the inductance:

$$L\left(\frac{dI}{dt}\right) = \omega I_0 \cos(\omega t - \alpha) = \omega L I_0 \sin(\omega t - \alpha + \tfrac{1}{2}\pi), \text{ leading } I \text{ by } \tfrac{1}{2}\pi.$$

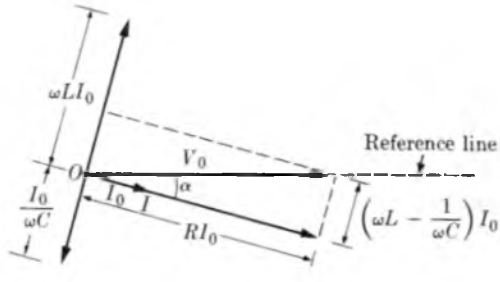


Fig. 9-19. Rotation vector diagram for the circuit shown in Fig. 9-14.

Potential drop across the capacitor:

$$\frac{q}{C}=-\left(\frac{1}{\omega C}\right)I_0\sin(\omega t-\alpha)=\left(\frac{1}{\omega C}\right)I_0\sin(\omega t-\alpha-\tfrac{1}{2}\pi), \text{ lagging } I \text{ by } \tfrac{1}{2}\pi.$$

The three rotating vectors are shown in Fig. 9-19, in which the reference line is given by the rotating vector corresponding to V. Their amplitudes are RI_0, ωLI_0, and $I_0/\omega C$. Their resultant must be V_0 since the three potential drops must add up to the applied emf. Therefore

$$V_0^2=R^2I_0^2+\left(\omega L-\frac{1}{\omega C}\right)^2 I_0^2$$

or

$$V_0=\sqrt{R^2+(\omega L-1/\omega C)^2}\,I_0.$$

Solving this equation for I_0 gives a result identical to Eq. (9.23). Also, from the figure the phase angle α may be computed. Its value agrees with Eq. (9.24). The rotating-vector technique is widely used in the engineering analysis of alternating-current circuits. ▲

9.6 Coupled Circuits

Consider two circuits such as (1) and (2) in Fig. 9-20. When a current I_1 circulates in circuit (1), a magnetic field proportional to I_1 is established throughout space; through circuit (2) there is then a magnetic flux Φ_2, which is also proportional to I_1. It is customary to write

$$\Phi_2=MI_1 \tag{9.31}$$

where M is a coefficient of proportionality and represents the magnetic flux through circuit (2) per unit current in circuit (1). Similarly if a current I_2 circulates in circuit (2), a magnetic field is produced; and it in turn produces through circuit (1) a magnetic flux Φ_1, which is proportional to I_2. Hence

$$\Phi_1=MI_2. \tag{9.32}$$

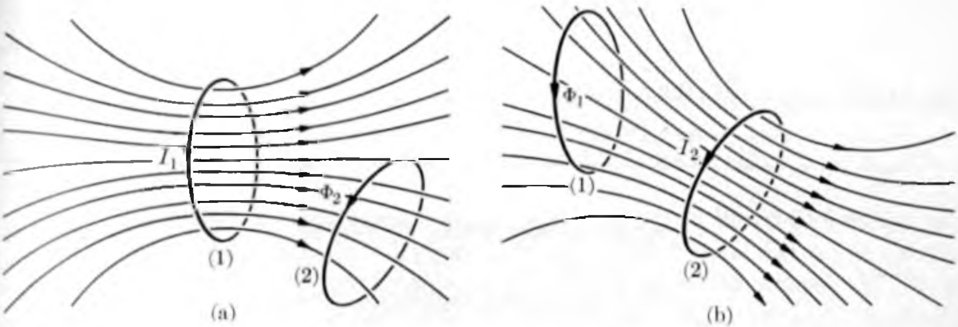


Fig. 9-20. Mutual induction.

Note that Eq. (9.32) uses the same coefficient M that Eq. (9.31) uses. This common element means that the magnetic flux through circuit (1) caused by the unit current in circuit (2) is the same as the magnetic flux through circuit (2) caused by the unit current in circuit (1). This common coefficient is called the *mutual inductance* of the two circuits; and one can prove that it must be the same in both cases as indicated. In other words, mutual induction is symmetrical. The coefficient M depends on the shapes of the circuits and their relative orientation. Mutual inductance, like self-inductance, is measured in henrys since it corresponds to Wb A^{-1}.

If the current I_1 is variable, the flux Φ_2 through circuit (2) changes; and an emf V_{M2} is induced in this circuit. This emf is given by

$$V_{M2} = -M\frac{dI_1}{dt}.$$

The assumption in writing this equation is that the circuits are rigid and fixed in space so that M is constant. Similarly if the current I_2 is variable, an emf V_{M1} is induced in circuit (1) and is given by

$$V_{M1} = -M\frac{dI_2}{dt}. \tag{9.33}$$

Therefore M is called "mutual inductance" since it describes the mutual effect or influence between the two circuits. In addition if the circuits are moved relative to each other, the result is a change in M; and emf's are again induced in both current loops.

The equation relating the current in circuit (1) to the parameters of the system is found by using Ohm's law. All that is needed is to add to Eq. (9.6) the emf V_{M1} given by Eq. (9.33). That is,

$$RI_1 = V_{L1} + V_{C1} + V_{M1}$$

where $V_{L1} = -L_1\,dI_1/dt$ and $V_{C1} = -q_1/C$. Therefore if the time derivative of the preceding equation is taken (note that $I_1 = dq_1/dt$), instead of Eq. (9.17) the result is

$$L_1 \frac{d^2 I_1}{dt^2} + R_1 \frac{dI_1}{dt} + \frac{1}{C_1} I_1 = -M \frac{d^2 I_2}{dt^2}. \qquad (9.34)$$

Similarly for circuit (2) the equation is

$$L_2 \frac{d^2 I_2}{dt^2} + R_2 \frac{dI_2}{dt} + \frac{1}{C_2} I_2 = -M \frac{d^2 I_1}{dt^2}. \qquad (9.35)$$

Equations (9.34) and (9.35) form a set of two simultaneous differential equations similar to Eq. (12.32) of Volume I for two coupled oscillators. The coupling constant is M. The general solutions will not be considered, but from the discussion of mechanically coupled oscillators we conclude that there will be an exchange of energy between the circuits. Common and practical applications of this process are the *transformer* and the *induction generator*. Another application of mutual induction in a broader sense is the transmission of a signal from one place to another by producing a variable current in one circuit, called the *transmitter*. This circuit in turn acts on another circuit, the *receiver*, coupled to it. This procedure is the case for telegraph, radio, television, radar, etc.

Example 9.6. The mutual inductance of the system shown in Fig. 9-21.

▼ The system of Fig. 9-21 consists of a coil that contains N turns and is wrapped around the central portion of a toroidal solenoid having n turns per unit length and a cross section of area S. It is first necessary to find the magnetic flux through the solenoid when there is a current in the coil, or conversely to find the magnetic flux through the coil when there is a current in the solenoid. We shall follow the second procedure, which is the easier of the two. Recall from Example 6-2 that in a toroidal solenoid the magnetic field is confined to the interior of the solenoid and has a value $\mathscr{B} = \mu_0 nI$, given by Eq. (6.7). The magnetic flux through any cross section of the solenoid is

$$\Phi_m = \mathscr{B}S = \mu_0 nSI$$

where S is the cross-sectional area of the solenoid. This flux is the same as the flux through any turn of the coil even if its cross section is larger than that of the solenoid. Therefore the magnetic flux through the coil is

$$\Phi = N\Phi_m = \mu_0 nNSI.$$

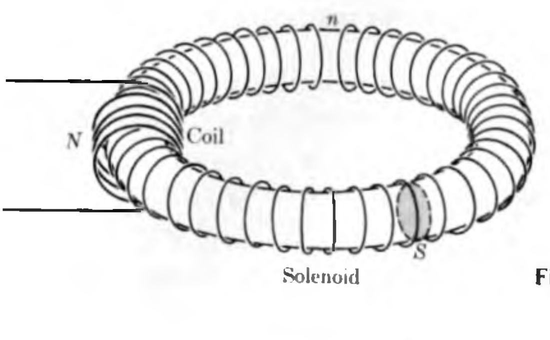


Figure 9-21

Comparison with Eq. (9.31) gives for the mutual inductance of the system

$$M=\mu_0 nNS,$$

which is seen to be a function only of the geometry of the two circuits. This arrangement is widely used in the laboratory when a standard mutual inductance is required. ▲

9.7 Concluding Remarks

The most important and fundamental aspect of mutual induction is that *energy can be exchanged between two circuits via the electromagnetic field*. That is, the electromagnetic field produced by the currents in the circuits acts as a carrier of energy and transports the energy through space from one circuit to the other. Since mutual induction between two circuits is a macroscopic phenomenon, resulting from elementary interactions between the moving charges that constitute their respective currents, we may conclude from this phenomenon that the electromagnetic interaction between any two charged particles can also be described as an exchange of energy via their mutual electromagnetic field.

When two charged particles participate in an electromagnetic interaction, the principle of conservation of energy must be restated to include the energy of the field. [Recall that the principle of conservation of momentum in Eq. (2.31) also had to be restated to take the momentum of the field into account.] Thus the total energy of a system of two interacting charged particles is

$$E=E_1+E_2+E_{\text{field}} \tag{9.36}$$

where E_1 and E_2 are the total energies of each particle, each energy being the sum of the kinetic and potential energies resulting from any force acting on the particle, and E_{field} is the energy associated with their mutual electromagnetic interactions. It can be proved that under static conditions (or conditions that vary very slowly with time), E_{field} corresponds exactly to the potential energy

$$E_p=\frac{q_1q_2}{4\pi\epsilon_0 r_{12}}$$

produced by the Coulomb interaction between the two charges.

It is the sum of the three terms in Eq. (9.36) that remains constant during the motion of two particles if they are subject to no other forces.

Problems

9.1 The magnetic flux through a circuit carrying a current of 2 A is 0.8 Wb. Find its self-inductance. Compute the emf induced in the circuit if in 0.2 s the current is (a) doubled, (b) reduced to zero, and (c) reversed.

9.2 Find the self-inductance of a toroidal solenoid of N turns. Assume that the radius of the coils is very small compared with the radius of the torus.

9.3 A capacitor C that has an initial charge q_0 is connected to a resistor R. If the switch S of Fig. 9-22 is closed, the capacitor discharges through the resistor. Show that (a) the current in the circuit is

$$I=-\frac{dq}{dt},$$

(b) the equation of the circuit is

$$\frac{q}{C}=RI,$$

(c) the charge of the capacitor at time t is

$$q=q_0e^{-t/RC},$$

and (d) the energy dissipated in the resistance by the Joule effect is equal to the initial energy of the capacitor. (*Hint:* For (c), combine (a) and (b); for (d), compute the integral $\int_0^\infty RI^2\,dt$.)

9.4 A capacitor C_1 has an initial charge q_0. When the switch S is closed (Fig. 9-23), the capacitor is connected in series with a resistor R and an uncharged capacitor C_2. (a) Show that the equation of the circuit is

$$\frac{q}{C_1}+\frac{(q_0-q)}{C_2}=RI.$$

(b) Find q and I as functions of time.

9.5 A capacitor C having an initial charge q_0 is connected to a self-inductance L of negligible resistance (Fig. 9-24). If the switch S is closed, the capacitor discharges through the inductance. Show that (a) the current in the circuit is

$$I=-\frac{dq}{dt},$$

(b) the equation of the circuit is

$$\frac{q}{C}-L\frac{dI}{dt}=0,$$

(c) the charge on the capacitor at time t is $q=q_0\cos\omega t$ where $\omega=1/\sqrt{LC}$ so that electric oscillations are set up. This arrangement is used to obtain high-frequency oscillations.

9.6 A battery of emf V and negligible internal resistance is connected in series with a resistance R and an uncharged capacitor C (Fig. 9-25). After the switch S is closed, show that (a) the current in the circuit is

$$I=+\frac{dq}{dt}$$

where q is the charge accumulated in the capacitor, (b) the equation of the circuit is

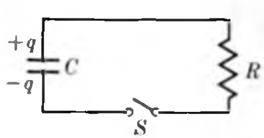


Figure 9-22

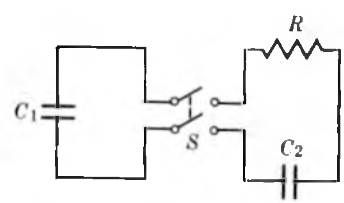


Figure 9-23

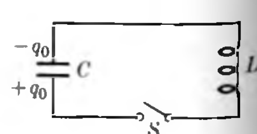


Figure 9-24

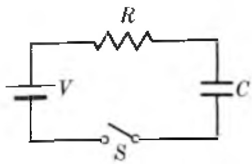


Figure 9-25

$V-q/C=RI$. (c) the charge as a function of time is

$$q=VC(1-e^{-t/RC})$$

and (d) the current as a function of time is

$$I=(V/R)e^{-t/RC}.$$

Plot q and I as functions of time.

9.7 A circuit is composed of a resistance to which an alternating emf $V=V_0 \sin \omega t$ is applied. Show that the current is given by

$$I=\left(\frac{V_0}{R}\right)\sin \omega t,$$

plot the rotating vector of the emf and the current, and show that they are in phase. What is the impedance of the circuit?

9.8 A circuit is composed of an alternating emf of amplitude V_0 and angular frequency ω connected to a capacitor C. (a) Find the current. (b) Draw the rotating vectors corresponding to the applied emf and to the current. (c) Plot the current as a function of ω and of C.

9.9 A 1-μF capacitor is connected across an a-c source whose voltage amplitude is kept constant at 50 V, but whose frequency can be varied. Find the current amplitude when the angular frequency is (a) 100 s^{-1}, (b) 1000 s^{-1}, and (c) 10,000 s^{-1}. (d) Construct a log-log plot of current amplitude versus frequency.

9.10 An inductor of self-inductance 10 H and of negligible resistance is connected across the source of Problem 9.9. Find the current amplitude when the angular frequency is (a) 100 s^{-1}, (b) 1000 s^{-1}, (c) 10,000 s^{-1}. (d) Construct a log-log plot of current amplitude versus frequency.

9.11 The voltage amplitude of an a-c source is 50 V and its angular frequency is 1000 s^{-1}. Find the current amplitude if the capacitance of a capacitor connected across the source is (a) 0.01 μF, (b) 1.0 μF, and (c) 100 μF. (d) Construct a log-log plot of current amplitude versus capacitance.

9.12 Find the current amplitude if the self-inductance of a resistanceless inductor connected across the source of Problem 9.11 is (a) 0.01 H, (b) 1.0 H, and (c) 100 H. (d) Construct a log-log plot of current amplitude versus self-inductance.

9.13 A circuit is composed of an alternating emf of amplitude V_0 and angular frequency ω connected to a self-inductance L. (a) Find the current. (b) Draw the rotating vectors corresponding to the applied emf, the potential drop across the self-inductance and the current. (c) Plot the current as a function of ω and of L.

9.14 A circuit is composed of a resistance and an inductance in series to which an alternating emf $V=V_0 \sin \omega_f t$ is applied. Show that the impedance of the circuit is $\sqrt{R^2+(\omega L)^2}$, and that the current lags the emf by an angle $\tan^{-1}(\omega L/R)$. (*Hint*: Plot the rotating vector of the current. Then, using the results of Problem 9.13, draw the rotating vectors corresponding to the potential difference or emf across the resistance and the inductance. Find their magnitude and compare with V_0 to obtain the impedance. The angle between the resultant rotating vector of the emf and the rotating vector of the current gives the phase difference.)

9.15 Repeat the preceding problem for a circuit composed of (a) a resistance and a capacitor, and (b) an inductance and a capacitor.

9.16 A solenoid has 10^3 turns/m and a cross section of 1.2×10^{-3} m^2. Around its central section a coil of 300 turns is wound. Determine (a) their mutual inductance, and (b) the emf in the coil if the initial current of 2 A in the solenoid is reversed in 0.2 s.

9.17 Coils A and B have 200 and 800 turns, respectively. A current of 2 A in A produces a

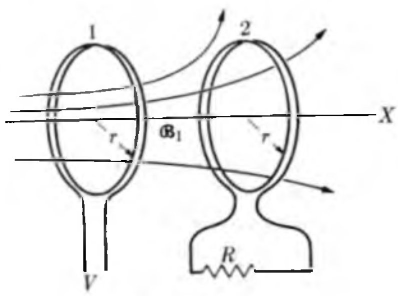


Figure 9-26

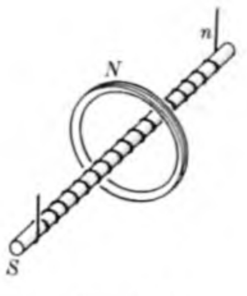


Figure 9-27

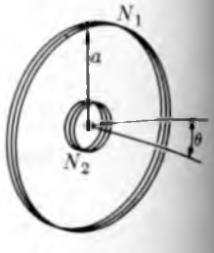


Figure 9-28

magnetic flux of 1.8×10^{-4} Wb in each turn of B. Compute (a) the coefficient of mutual inductance, (b) the magnetic flux through A when there is a current of 4 A in B, and (c) the emf induced in B when the current in A changes from 3 A to 1 A in 0.3 s.

9.18 Two coils are placed coaxially as shown in Fig. 9-26. Coil 1 is connected to an external source of emf labeled V. Assume that the geometry is such that one-fifth of the magnetic flux produced by coil 1 pass through coil 2, and vice versa. The resistances of the coils are R_1 and R_2; and coil 2 is connected to an external resistance R as shown. The numbers of turns in the coils are N_1 and N_2. The total flux produced by coil 1 is given by

$$\Phi_{\mathcal{B}} = (L_1/N_1)I_1,$$

where L_1 is the self-inductance of coil 1. (a) Find the emf induced in coil 2 when I_1 increases uniformly from 0 to I_0 in t s. (b) Find the induced emf in coil 2 when $I_1 = I_0 \sin \omega t$.

9.19 A coil having N turns is placed around a very long solenoid with cross section S and n turns per unit length (see Fig. 9-27). Show that the mutual inductance of the system is $\mu_0 nNS$.

9.20 In the center of a circular coil with radius a and N_1 turns, there is a very small coil with area S and N_2 turns as shown in Fig. 9-28. Show that the mutual inductance is $\frac{1}{2}\mu_0 N_1 N_2 S \cos \theta/a$ where θ is the angle between the normals to the two coils.

CHALLENGING PROBLEMS

9.21 In the arrangement shown in Fig. 9-29, the capacitor on the left has a capacitance C and has been charged to voltage V_0; the capacitor on the right has a capacitance $3C$ and is initially uncharged. The switch is then closed. A long time after closing the switch, what is the voltage across each capacitor in terms of V_0? (AP-C; 1971)

9.22 In Fig. 9-30, $\mathcal{E} = 100$ volts; $C_1 = 12$ microfarads; $C_2 = 24$ microfarads; $R = 10$ ohms. Initially, C_1 and C_2 are uncharged, and all switches are open. (a) First, switch S_1 is closed. Determine the charge on C_1 when equilibrium is reached. (b) Next S_1 is opened and afterward S_2 is closed. Determine the charge on C_1 when equilibrium is again reached. (c) For the equilibrium condition of part (b), determine the voltage across C_1. (d) S_2 remains closed, and now S_1 is also closed. How much *additional* charge flows from the battery? (AP-C; 1975)

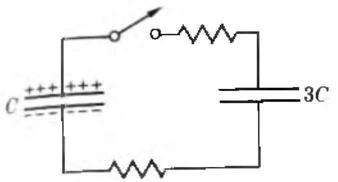


Figure 9-29

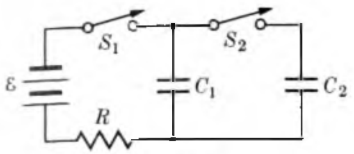


Figure 9-30

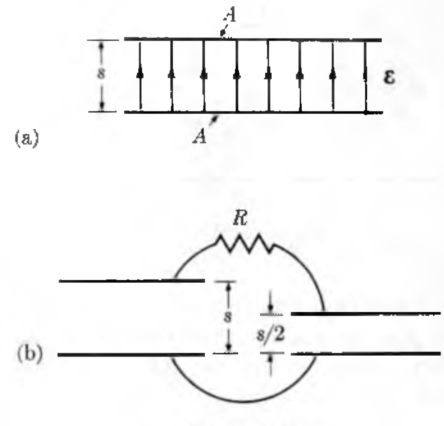


Figure 9-31

9.23 A uniform electric field $\mathcal{E}$ is established between two capacitor plates, each of area A, which are separated by a distance s as shown in Fig. 9-31a. (a) What is the electric potential difference V between the plates? (b) Specify the sign of the charge on each plate. The capacitor is then connected electrically through a resistor to a second parallel-plate capacitor, initially uncharged, whose plates have the same area A but a separation of only $s/2$. (c) Indicate on a diagram like Fig. 9-31b the direction of the current in each wire, and explain why the current will eventually cease. (d) After the current has ceased, which capacitor has the greater charge? Explain your reasoning. (e) The total energy stored in the two capacitors after the current has ceased is less than the initial stored energy. Explain qualitatively what has become of this "lost" energy. (AP-B; 1978)

9.24 The bridge illustrated in Fig. 9-32 can be used to compare the two inductances L_1 and L_2. The bridge is balanced so that the current from B to D is zero at all times when the alternating emf V is applied. Show that $L_1/L_2 = R_3/R_4$.

9.25 In Problem 9.24 the resistance of the inductances was neglected. If their resistances are R_1 and R_2, the procedure is as follows. First the bridge is balanced until there is no current between B and D when a *constant* emf is applied. Next the current is balanced as in Problem 9.24, without changing the resistance. Show that the same relation still holds.

9.26 In the circuit of Fig. 9.33 $V = V_0 \sin \omega t$ is an alternating emf. Find the amplitude and

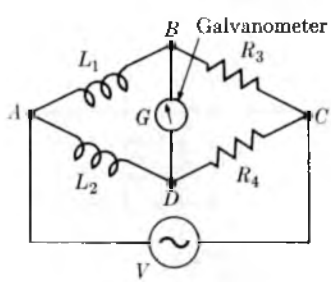


Figure 9-32

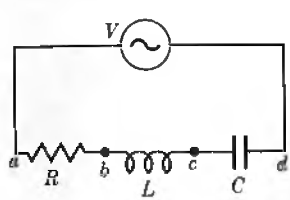


Figure 9-33

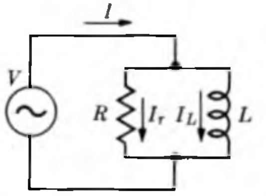


Figure 9-34

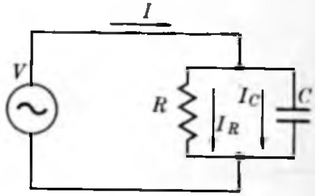


Figure 9-35

phase relative to the emf of the potential difference V_{ab}, V_{bc}, V_{cd}, V_{ac}, V_{bd}. (*Hint*: Draw the corresponding rotating vectors, as indicated before in Fig. 9-19.)

9.27 An alternating emf having a maximum value of 100 V and an angular frequency of 120π s^{-1} is connected in series with a resistance of $1\,\Omega$, a self-inductance of 3×10^{-3} H, and a capacitor of 2×10^{-3} F. Determine (a) the amplitude and phase of the current, and (b) the potential difference across the resistance, the inductance, and the capacitor. (c) Make a diagram showing the rotating vectors corresponding to the applied emf, the current, and the three potential differences. (d) Verify that the three potential difference vectors add to the emf vector.

9.28 If I_{rms} and V_{rms} in an a-c circuit are the root-mean-square values of the current and the emf over one cycle, show that

$$I_{rms}=\frac{I_0}{\sqrt{2}},$$

$$V_{rms}=\frac{\mathscr{E}_0}{\sqrt{2}},$$

and

$$P_{ave}=I_{rms}V_{rms}\cos\alpha$$

where α is the phase angle between the current and the emf.

9.29 A circuit consists of an alternating emf having a maximum value of 100 V, a resistance of $2\,\Omega$, a self-inductance of 10^{-3} H, and a capacitance of 10^{-3} F, all connected in series. Find the maximum value of the current for the following values of the angular frequency of the emf: (a) 0, (b) 10 s^{-1}, (c) 10^2 s^{-1}, (d) resonance, (e) 10^4 s^{-1}, and (f) 10^5 s^{-1}. Plot the current against the logarithm of the frequency.

9.30 A circuit is composed of a resistance and an inductance in parallel as shown in Fig. 9-34. Show that the resultant impedance of the circuit is given by

$$\frac{1}{Z}=\sqrt{\frac{1}{R^2}+\frac{1}{\omega^2L^2}}$$

and the phase by $\tan^{-1}(R/\omega L)$. (*Hint*: Plot the rotating vector of the applied emf. Then, using the results of Problems 9.13 and 9.14, draw the rotating vectors corresponding to the current in the resistance and the inductance.

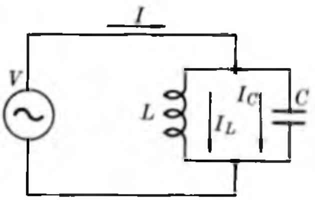


Figure 9-36

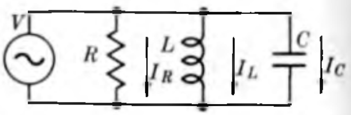


Figure 9-37

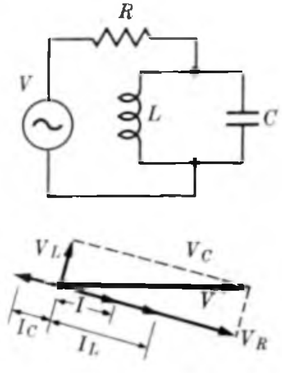


Figure 9-38

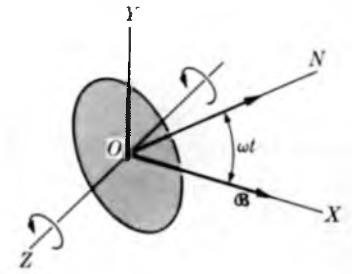


Figure 9-39

Their resultant gives the total current from which the impedance and the phase difference are obtained.)

9.31 Repeat the preceding problem for the circuit illustrated in Fig. 9-35.

9.32 Repeat Problem 9.30 for the circuit illustrated in Fig. 9-36.

9.33 Repeat Problem 9.30 for the circuit illustrated in Fig. 9-37.

9.34 A coil having a resistance of $1\,\Omega$ and a self-inductance of 10^{-3} H is connected in parallel with a second coil having a resistance of $1\,\Omega$ and a self-inductance of 3×10^{-3} H. An alternating emf having an amplitude of 10 V and an angular frequency of 120π s^{-1} is connected to the system. Calculate (a) the current across each coil, (b) the total current. (c) Make a plot showing the rotating vector for the emf, the current in each conductor, and the total current. (d) Verify that the vector of the total current is equal to the sum of the vectors of each current.

9.35 A circuit is composed of an inductance and a capacitor in parallel, connected in series with a resistance R as shown in Fig. 9-38. (a) Draw the rotating vectors corresponding to V, I_L, I_C, RI, V_L, and V_C. (b) Show that the impedance of the circuit is $\sqrt{Z}=[R^2+\omega^2L^2/(1-\omega^2LC)^2]^{1/2}$. (c) What is the value of the impedance when $\omega=1/\sqrt{LC}$? (In this case there is said to be *antiresonance*.) (d) Make a rough sketch of the current versus the frequency. (*Hint:* Note that the rotating vectors of the currents through L and C must add to the current through R, but those corresponding to the potential difference must be identical. To assist the student, the rotating vector diagram has also been shown.)

9.36 A circular coil of radius a, resistance R, and self-inductance L rotates with a constant angular velocity around a diameter perpendicular to a uniform magnetic field (see Fig. 9-39). Find (a) the emf induced and the current in the coil, (b) the average values of the x- and y-components of the magnetic field produced by the coil at O, and (c) the angle with the X-axis made by a magnetic needle placed at O.

9.37 Verify by direct substitution that Eq. (9.22) is a solution of Eq. (9.21) if I_0 and α are given by Eqs. (9.23) and (9.24), respectively. (*Hint:* First expand $\sin(\omega t-\alpha)$ and replace $\sin\alpha$ and $\cos\alpha$ by their corresponding values as derived from Eq. (9.24).)

PART THREE

WAVES

Of all concepts used in physics, two can be understood intuitively by all people, no matter what their cultural level. These concepts are *particle* and *wave*. To the average person, a particle is a small portion of matter where "small" normally is decided in terms of an anthropomorphic scale and means small relative to the environment of the particle. Similarly waves are ordinarily pictured in terms of the waves observed on the surface of water, on a string, or on a spring.

The physicist uses the concept of particle in a somewhat more abstract and fundamental sense so that a great variety of physical situations may be adequately treated. The concept of the wave undergoes a similar transformation; the physicist has extended the concept and applied it to a large number of phenomena that do not resemble a wave on the surface of water, but that have the same mathematical description. The following chapters present a general discussion of wave phenomena in this broad sense.

Several types of waves will be analyzed with special emphasis on electromagnetic waves. In each case the student must concentrate on understanding the physical situation described and the mathematical framework used, and must avoid the inevitable temptation merely to picture all waves as those on the surface of a liquid. The most important aspects of waves are the velocity of their propagation and the modifications they suffer when the physical properties of the medium change (to produce reflection, refraction, polarization), when different kinds of obstacles are interposed in their paths (to produce diffraction, scattering), or when several waves coincide in the same region of space (to produce interference). These are the specific topics to be covered; the primary purpose of this portion of the text is to enable the student to attain a fundamental understanding of the wave description of physical phenomena, namely the propagation of a physical situation described by a time-dependent field.

CHAPTER TEN

WAVE MOTION: ELASTIC WAVES

10.1 Introduction

When a bell is struck, sound is heard at distant points; sound is transmitted through the surrounding air. If a speeding boat passes at some distance from the shore, the wake that the boat has produced eventually reaches the shore. When a light bulb is turned on, the room is filled with light. Section 8.11 demonstrated that it is possible to transmit an electromagnetic signal from one place to another as a result of the physical relations between the electric and magnetic fields. Although the physical mechanism may be different for each of the processes just mentioned, they all have a common feature: they are physical disturbances that are produced at one point in space, propagate through space, and produce an effect later at another point. The propagation of any of these disturbances through a medium is an example of *wave motion*.

To examine the matter more generally, consider a physical property described by a certain field. This field may be an electromagnetic field, the deformation in a spring, the pressure in a gas, the strain in a solid, the transverse displacement of a string, or even the gravitational field. Suppose that the conditions at one place become time dependent or dynamic so that a perturbation of the physical state of the system occurs at that place. The physical properties of the system, which are described by the time-dependent equations of the field (such as Maxwell's equations for the electromagnetic field), result in the *propagation* of this disturbance through space. The disturbance disrupts the static conditions at other places.

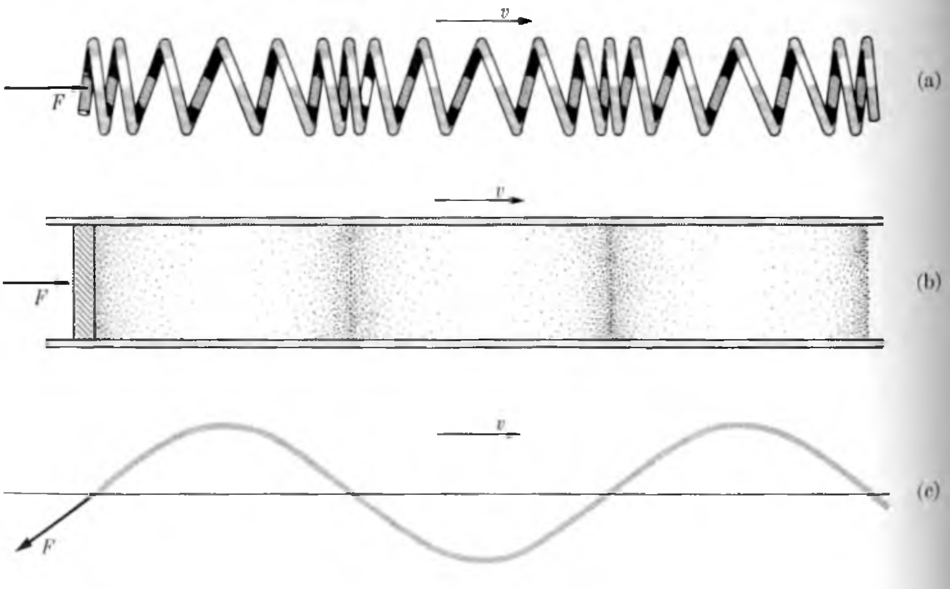


Fig. 10-1. Elastic waves (a) of a spring, (b) in a gas, and (c) on a string.

For example consider the free surface of a liquid. The field in this case is the displacement of each point of the surface relative to the equilibrium position. Under equilibrium or static conditions the free surface of a liquid is plane and horizontal; but if at one point the conditions at the surface are disturbed by dropping a stone into the liquid, it is well known that this disturbance propagates in all directions along the surface of the liquid. To determine the mechanism of propagation and its velocity, one must analyze how the displacement of a point at the surface of the liquid affects the rest of the surface. This analysis produces the dynamical equations for the process. These equations then enable us to obtain quantitative information about the variation in space and time of the disturbance.

In this chapter the general characteristics of wave motion are first discussed and then followed by specific examples. Most of the examples will correspond to elastic waves in a substance. (See Fig. 10-1.) In most of these cases the molecular structure of matter may be ignored and a continuous medium can be assumed. This assumption is valid as long as the space fluctuation of the wave (determined by the wavelength) is large compared with the intermolecular separation of the medium supporting the wave motion.

10.2 Mathematical Description of Wave Motion

Consider a function $\xi = f(x)$ represented graphically by the solid curve in Fig. 10-2. If every point of the curve is translated a distance $\Delta x = a$ to the right (or to the left) without deformation, then the value of the function at each new point, say x', is the same as the function's value at $x' - a$ (or $x' + a$). Thus $f(x-a)$ represents the curve displaced *without deformation* to the right by an amount a, and similarly $f(x+a)$ represents the same curve displaced to the left by an amount a.

Now consider a continuous displacement of the curve $f(x)$. When the curve is displaced a distance Δx from the curve's position at time $t=0$ in a time Δt with a velocity v, such that $a = \Delta x = v\Delta t = vt$ (where v is called the *phase velocity*), then a

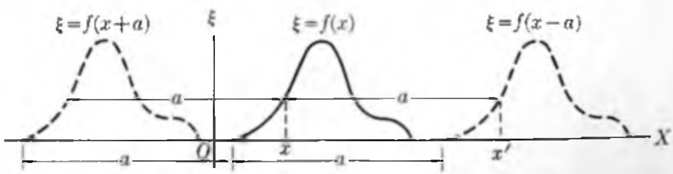


Fig. 10-2. Undistorted translation of a function $\xi(x)$.

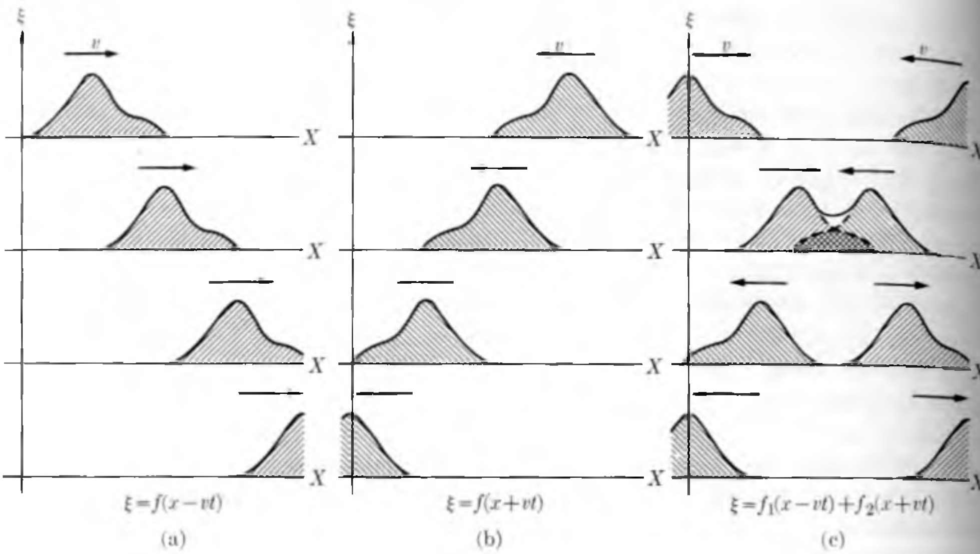


Fig. 10-3. Undistorted propagation of a wave (a) to the right and (b) to the left. (c) Waves propagating in opposite directions produce additive results where the waves interfere.

"pulse" is "traveling" along the x-direction (Fig. 10-3). Therefore a mathematical expression of the form

$$\xi(x, t)=f(x \mp vt) \tag{10.1}$$

is adequate for describing a physical disturbance that travels or "propagates" without deformation along the positive (or negative) X-axis; this propagation is the characteristic feature of *wave motion*. The quantity $\xi(x, t)$ may represent a great diversity of physical quantities, such as the deformation in a solid, the pressure in a gas, an electric or magnetic field, etc.

An especially interesting case is that in which $\xi(x, t)$ is a sinusoidal or harmonic function such as

$$\xi(x, t)=\xi_0 \sin k(x-vt)=\xi_0 \sin [k(x-vt)+2\pi]. \tag{10.2}$$

The quantity k has a special meaning. When the value of x is replaced by $x+2\pi/k$, the function $\xi(x, t)$ has the same value; that is,

$$\begin{aligned}\xi\left(x+\frac{2\pi}{k}, t\right)&=\xi_0 \sin k\left(x+\frac{2\pi}{k}-vt\right)\\ &=\xi_0 \sin [k(x-vt)+2\pi]=\xi_0 \sin k(x-vt)\\ &=\xi(x, t).\end{aligned}$$

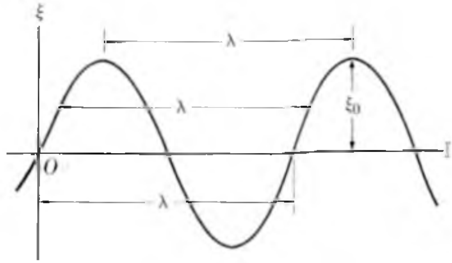


Fig. 10-4. Harmonic wave.

Then

$$\lambda = \frac{2\pi}{k} \tag{10.3}$$

is the "space period" of the curve in Fig. 10-4; that is, the curve repeats itself every length λ. The quantity λ is called the *wavelength*, and the quantity $k = 2\pi/\lambda$ represents the number of wavelengths in the distance 2π and is called the *wave number*.* Therefore

$$\xi(x, t) = \xi_0 \sin k(x - vt) = \xi_0 \sin \frac{2\pi}{\lambda}(x - vt) \tag{10.4}$$

represents a sinusoidal or harmonic wave of wavelength λ propagating to the right along the X-axis with a phase velocity v. Equation (10.4) can also be written in the form

$$\xi(x, t) = \xi_0 \sin (kx - \omega t) \tag{10.5}$$

where

$$\omega = kv = \frac{2\pi v}{\lambda} \tag{10.6}$$

is called the *angular frequency* of the wave. According to Eq. (12.2) of Volume 1, $\omega = 2\pi\nu$ where ν is the *frequency* with which the physical disturbance varies at every point x; therefore

$$\lambda\nu = v, \tag{10.7}$$

which relates the wavelength, the frequency, and the phase velocity of a wave. Thus if P, the period of oscillation at each point, is given by $P = 1/\nu = 2\pi/\omega$, Eq. (10.4) may be written in the form

$$\xi(x, t) = \xi_0 \sin 2\pi\left(\frac{x}{\lambda} - \frac{t}{P}\right). \tag{10.8}$$

---

*Sometimes the term wave number is reserved for $1/\lambda$ or $k/2\pi$, corresponding to the number of wavelengths in one unit of length.

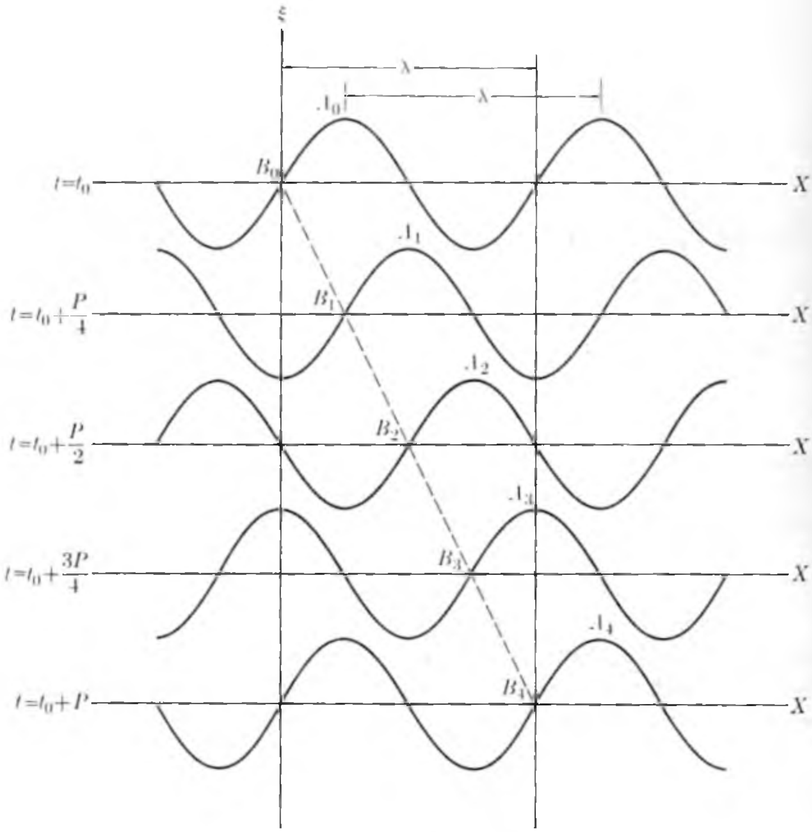


Fig. 10-5. Harmonic wave propagating to the right. The wave advances the distance λ in the time P.

Similarly

$$\xi(x, t)=\xi_0 \sin k(x+vt)=\xi_0 \sin (kx+\omega t)$$
$$=\xi_0 \sin 2\pi\left(\frac{x}{\lambda}+\frac{t}{P}\right) \tag{10.9}$$

represents a sinusoidal or harmonic wave moving in the $-X$-direction.

It is instructive to look at the space distribution of $\xi(x, t)$ at different, successive time intervals. The function $\xi(x, t)$ has been represented in Fig. 10-5 at times t_0, $t_0+P/4$, $t_0+P/2$, $t_0+3P/4$, and t_0+P. Note that while the wave itself propagates to the right, it repeats itself in *space* after one period. The reason for this repetition is that from Eq. (10.7)

$$\lambda=v/\nu=vP.$$

which shows that the wavelength may also be defined as the distance advanced by the wave in one period. Therefore in sinusoidal wave motion there are two periodicities: one in time, given by the period P, and one in space, given by the wavelength λ, with the two related by $\lambda = vP$.

The student may easily verify that the general expressions (10.1) for a traveling harmonic wave may be written in the alternative form

$$\xi(x, t) = F(t \pm x/v)$$

where the positive sign corresponds as before to propagation in the $-X$-direction and the negative sign to propagation in the $+X$-direction. Thus if we choose this functional form for $\xi(x, t)$ we may write

$$\xi(x, t) = \xi_0 \sin \omega(t \pm x/v) = \xi_0 \sin (\omega t \pm kx) \tag{10.10}$$

instead of Eq. (10.5) and (10.9).

Example 10.1. The wavelength of sound from a tuning fork that oscillates with a frequency of 440 Hz. (The velocity of sound in air is ≈ 340 m s^{-1}.)

▼ Equation (10.7) may be rewritten as $\lambda = v/\nu$ so that

$$\lambda = \frac{v}{\nu} = \frac{340 \text{ m s}^{-1}}{440 \text{ Hz}} = 0.772 \text{ m.}$$ ▲

Example 10.2. The wavelength of light in the red region of the visible spectrum. The frequency corresponding to red light is $\approx 5 \times 10^{14}$ Hz. (Light in vacuum propagates with a velocity of 3×10^8 m s^{-1}.)

▼ Applying Eq. (10.7) as in Example 10.1 gives

$$\lambda = \frac{v}{\nu} = \frac{3 \times 10^8 \text{ m s}^{-1}}{5 \times 10^{14} \text{ Hz}} = 6 \times 10^{-7} \text{ m.}$$

From these two examples the student should get a feeling for the vast difference (six orders of magnitude) between the wavelength of audible sound and that of visible light. ▲

10.3 Fourier Analysis of Wave Motion

According to Fourier's theorem in Section 12.14 of Volume 1, any periodic motion may be expressed as a superposition of simple harmonic motions of frequencies $\omega, 2\omega, \ldots, n\omega, \ldots$ or periods $P, P/2, \ldots, P/n, \ldots$. It is usually convenient to analyze

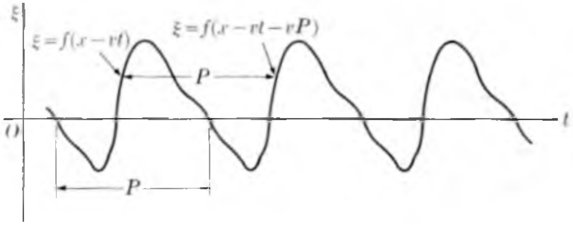


Fig. 10-6. Arbitrary waveform periodic in time as viewed from a fixed spatial position x.

an arbitrary periodic motion by Fourier's theorem because simple harmonic motions are well understood and obey simple equations.

In Section 10.2 simple harmonic waves were shown to have a dual periodicity in space and time; we will now show that an arbitrary waveform that is periodic in either space or time must be periodic also in the alternate variable so that the entire waveform may be expressed in terms of superposition of simple harmonic waves.

Let $\xi = f(x - vt)$ be a wave motion periodic in time; that is, a motion that at a given point in space has the same value at times $P, 2P, \ldots, nP, \ldots$ (Fig. 10-6). In other words

$$\xi = f(x - vt) = f[x - v(t \pm nP)] = f(x - vt \mp nvP).$$

However since $\lambda = vP$, at a given time t the displacement ξ has the same value at points separated by the distances vP, $2vP, \ldots, nvP, \ldots$. Therefore if instead of changing t, we change x by the amount $\lambda = vP$, the wave repeats itself in space (Fig. 10-7). Thus a wave motion periodic in time is also periodic in space.

As discussed in Section 12.14 of Volume 1, Fourier's theorem can be written

$$\begin{aligned} \xi = f(u) = a_0 + a_1 \cos u + a_2 \cos 2u + \cdots + a_n \cos nu \\ + b_1 \sin u + b_2 \sin 2u + \cdots + b_n \sin nu \end{aligned} \tag{10.11}$$

where $f(u)$ is a function periodic in the variable u with a periodicity of 2π. For our purposes u can be either x or t; and the periodicity is either λ or P, respectively. Thus for example

$$\xi = f(x - vt) = a_0 + a_1 \cos [kx - \omega t] + a_2 \cos [2(kx - \omega t)] + \cdots$$

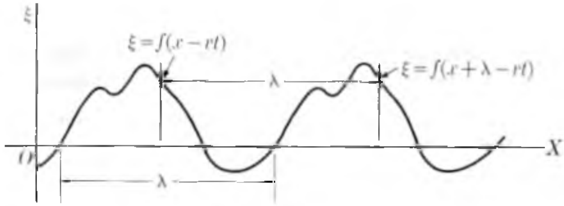


Fig. 10-7. Arbitrary waveform periodic in space as viewed at a fixed time t.

where $k=2\pi/\lambda$ and $\omega=2\pi/P$. This expression is just another way of writing the more general principle of the linear superposition of waves:

> *any periodic wave motion can be expressed as a linear superposition of harmonic waves of frequencies ω, $2\omega, \ldots, n\omega$ and wavelengths λ, $\lambda/2, \ldots, \lambda/n$.*

The coefficients a_n and b_n for each of the harmonic terms in Eq. (10.11) can be obtained by using the orthogonality properties of sines and cosines; that is, for the dimensionless variable u, with period 2π

$$a_0=\frac{1}{2\pi}\int_0^{2\pi} f(u)\,du,$$

$$a_n=\frac{1}{\pi}\int_0^{2\pi} f(u)\cos nu\,du, \tag{10.12}$$

and

$$b_n=\frac{1}{\pi}\int_0^{2\pi} f(u)\sin nu\,du.$$

The Fourier method is useful not only for analyzing periodic curves but even for analyzing nonperiodic curves, such as the pulse of Fig. 10-8. The main difference between this case and the discussion in the text is that instead of the analysis giving a set of *discrete* wavenumbers $k, 2k, 3k, \ldots nk, \ldots$ or a discrete frequency *spectrum* ω, $2\omega, 3\omega, \ldots, n\omega, \ldots$, the curve must be analyzed in terms of a continuum of wavelengths or continuous spectrum of frequencies. The amplitude corresponding to each wavelength or frequency is given by a continuous function of k or ω and is called the *Fourier transform* of the pulse. For a particular pulse form there is always a unique Fourier transform.

Consider the pulse given in Fig. 10-8. Such a pulse could be produced by allowing a loudspeaker to oscillate for only a short interval. If the curve of Fig. 10-8 had extended from $-\infty$ to $+\infty$ in space, no Fourier analysis would be necessary because the curve would have only one wavenumber, k_0. However in order that the curve may be annihilated for $x<x_1$ and $x>x_2$, other wavenumbers must be added so that the resultant Fourier series in those regions is zero. Therefore a finite pulse in space is a composite of many wavenumbers even if the source has a single well-defined frequency.

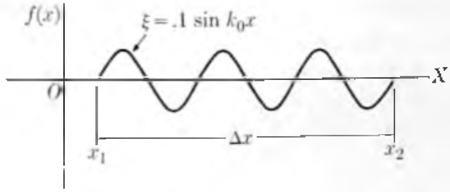


Fig. 10-8. A harmonic pulse in space.

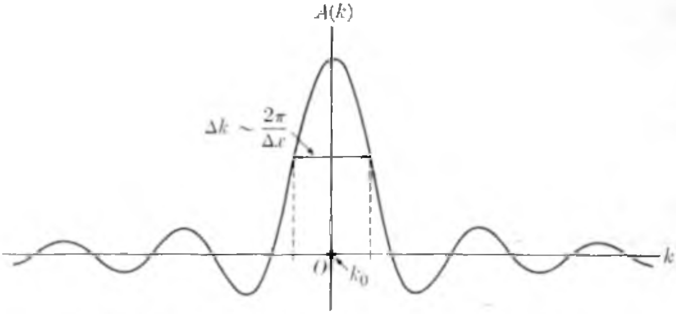


Fig. 10-9. Fourier transform of the pulse in Fig. 10-8.

Example 10.3. Fourier analysis of a wave described at time $t=0$ by the function $f(x)$ shown in Fig. 10-8. The expression for this wave is

$$\xi = A \sin k_0 x$$

in the interval $\Delta x = x_2 - x_1$, and zero outside that interval. This type of wave is called a *pulse* or *wave packet*.

▼ It may be proved that the amplitude function for k (i.e., the Fourier transform of the pulse) is given by

$$F(k) = \tfrac{1}{2} A \Delta x \left[\frac{\sin(\frac{1}{2}\Delta x \Delta k)}{(\frac{1}{2}\Delta x \Delta k)} \right]$$

where $\Delta x = x_2 - x_1$, the length of the pulse, and $\Delta k = k - k_0$. This amplitude function is illustrated in Fig. 10-9. When $\Delta k = 0$ or $k = k_0$, the function has a maximum of $\frac{1}{2}A\Delta x$ because the value of the term within the square brackets in the equation above tends toward unity as Δk tends toward zero. Furthermore because the numerator of the term within the brackets never has a magnitude greater than one, as the value of Δk increases in absolute value, the value of $F(k)$ decreases in an oscillatory manner. The range of values of k for which $F(k)$ is greater than 50 percent of the central maximum corresponds roughly to the condition

$$\left|\tfrac{1}{2}\Delta x \Delta k\right| < \frac{\pi}{2} \qquad \text{or} \qquad -\frac{\pi}{\Delta x} < \Delta k < \frac{\pi}{\Delta x}.$$

Thus the only wavenumbers whose amplitudes are appreciable are those in the range Δk around k_0 given by

$$\Delta x \Delta k \sim 2\pi. \tag{10.13}$$

This equation indicates that the shorter the space length of the pulse. the larger the range of wavenumbers required to represent the pulse accurately. Similarly one may consider a pulse in time and obtain a Fourier transform of frequency amplitude centered around the single frequency of oscillation where the relationship between frequency and time is

$$\Delta \omega \Delta t \sim 2\pi. \tag{10.14}$$

Equation (10.14) indicates that the shorter the time interval of a single frequency pulse, the larger the range of frequencies required to represent the pulse accurately. ▲

10.4 Differential Equation of Wave Motion

A second step is an investigation of how to determine when a given time-dependent field propagates as a wave without distortion. The fields associated with each physical process are governed by dynamical laws (characteristic of each process). These laws can be expressed in the form of differential equations as in the case of the electromagnetic field. We may look therefore into the possibility of finding a differential equation applicable to *all* kinds of wave motion. Then every time its physical properties indicate that a particular field satisfies such an equation, we may be sure that the field propagates through space with a definite velocity and without distortion.* Conversely, if experiment shows that a field propagates in space with a definite velocity and without distortion, the field can be described by a set of equations compatible with the wave equation.

The equation we shall encounter over and over that describes a wave motion propagating with a definite velocity v and without distortion along either the $+X$- or the $-X$-direction is

$$\frac{\partial^2 \xi}{\partial t^2}=v^2\frac{\partial^2 \xi}{\partial x^2}. \tag{10.15}$$

This expression is called the *differential equation of wave motion*. The general solution of Eq. (10.15) is of the form of Eq. (10.1):

$$\xi(x, t)=f_1(x-vt)+f_2(x+vt). \tag{10.16}$$

Thus the general solution of Eq. (10.15) can be expressed as the superposition of two wave motions propagating in opposite directions. Of course for a wave propagating in one direction, only one of the two functions appearing in Eq. (10.15) is required. However when (for example) there are an incoming wave along the $+X$-direction and a reflected wave along the $-X$-direction, the general form of Eq. (10.16) must be used.

To prove that an expression of the form of Eq. (10.16) is a solution of the wave equation (10.15) we must first introduce some mathematical results. When there is a

*This technique was used in Chapter 12 of Volume I, in which an oscillatory motion was shown to follow an equation of the type $d^2x/dt^2+\omega^2x=0$, and this equation was used to identify several types of simple harmonic motion, once the physical laws of motion associated with this equation were understood.

function $y=f(u)$ where u is, in turn, a function of x [that is, $u(x)$], then we can write

$$\frac{dy}{dx}=\frac{dy}{du}\frac{du}{dx}.$$

This expression is called the *chain rule* for derivatives. For example if $y=\sin(3x^2)$, then $y=\sin u$, $u=3x^2$, $dy/du=\cos u$, and $du/dx=6x$ so that

$$dy/dx=(\cos u)(6x)=6x\cos 3x^2.$$

Now apply the chain rule to $\xi=f(x\pm vt)$. In this case make $u=x\pm vt$ so that $\xi=f(u)$. Because u is a function of two variables, x and t, we have to use partial derivatives where $\partial u/\partial x=1$ and $\partial u/\partial t=\pm v$. Then since ξ is also a function of two variables, its partial derivatives are

$$\frac{\partial\xi}{\partial x}=\frac{d\xi}{du}\frac{\partial u}{\partial x}=\frac{d\xi}{du}, \qquad \frac{\partial\xi}{\partial t}=\frac{d\xi}{du}\frac{\partial u}{\partial t}=\pm v\frac{d\xi}{du}.$$

Taking second derivatives gives

$$\frac{\partial^2\xi}{\partial x^2}=\frac{d}{du}\left(\frac{\partial\xi}{\partial x}\right)\frac{\partial u}{\partial x}=\frac{d^2\xi}{du^2},$$

$$\frac{\partial^2\xi}{\partial t^2}=\frac{d}{du}\left(\frac{\partial\xi}{\partial t}\right)\frac{\partial u}{\partial t}=\left(\pm v\frac{d^2\xi}{du^2}\right)(\pm v)=v^2\frac{d^2\xi}{du^2}.$$

Combining these two equations to eliminate $d^2\xi/du^2$ gives Eq. (10.15), proving that $\xi=f(x\pm vt)$ is a solution of the wave equation, independent of the form of the function f. Since the wave equation is linear, the general solution is of the type indicated in Eq. (10.16).

As a concrete example, we may verify that the wave equation (10.15) is satisfied by the sinusoidal wave, $\xi=\xi_0\sin k(x-vt)$. The partial space and time derivatives of ξ are

$$\frac{\partial\xi}{\partial x}=k\xi_0\cos k(x-vt), \qquad \frac{\partial^2\xi}{\partial x^2}=-k^2\xi_0\sin k(x-vt);$$

$$\frac{\partial\xi}{\partial t}=-kv\xi_0\cos k(x-vt), \qquad \frac{\partial^2\xi}{\partial t^2}=-k^2v^2\xi_0\sin k(x-vt).$$

Therefore $\partial^2\xi/\partial t^2=v^2\partial^2\xi/\partial x^2$, in agreement with Eq. (10.15).

To provide a better understanding of the fundamental ideas of wave motion, certain kinds of waves that are more or less familiar to the student will now be considered. The student must note that in the waves to be discussed in the succeeding sections, Eq. (10.15) results from the dynamical laws of the process, considered together with certain approximations such as small amplitude or long wavelength, etc. Therefore the theory related to Eq. (10.15) is applicable only under the stated approximations.

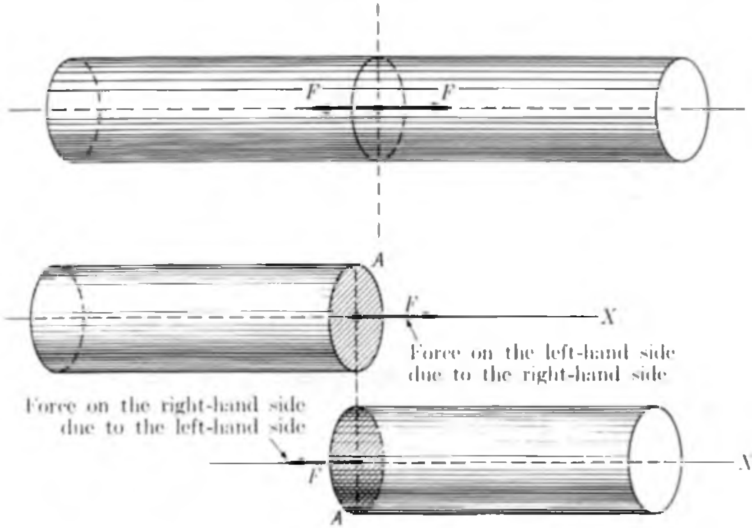


Fig. 10-10. The forces on any section of a rod under stress are equal and opposite.

10.5 Elastic Waves in a Solid Rod

If a disturbance is produced at one end of a solid rod, say by hitting it with a hammer, the disturbance propagates along the rod and eventually is felt at the other end. An *elastic wave* has propagated along the rod. In this section we will examine this elastic wave and show how its velocity of propagation is related to the physical properties of the rod. Consider a rod that has a uniform cross section A and is subject to a stress along the axis as indicated by the force F. The force F is not necessarily the same at all sections, and may vary along the axis of the rod. At each cross section (as shown in Fig. 10-10) there is a *normal stress* $\mathscr{S}$, defined as the force per unit area acting perpendicular to the cross section in either direction. Then

$$\mathscr{S} = \frac{F}{A}. \tag{10.17}$$

The stress is expressed in N m^{-2}, a unit called the *pascal* (Pa) to honor the French scientist Blaise Pascal (1623–1662).

Under the action of such forces each section of the rod suffers a displacement ξ parallel to the axis. If the displacement is the same at all points of the rod, there is no deformation but simply a rigid displacement of the rod along its axis. We are interested in the case in which there is a deformation so that ξ varies along the rod: that is, ξ is a function of x. Consider two sections A and A' separated by a distance dx in the un-

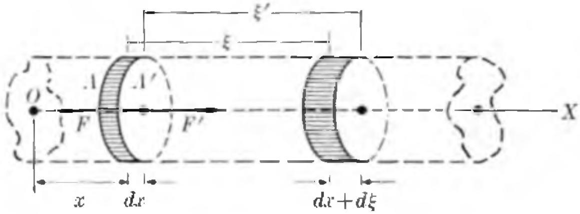


Fig. 10-11. Exaggerated diagram of a longitudinal wave in a rod.

disturbed rod (Fig. 10-11). When forces are applied to the rod, the section A is displaced a distance ξ; and the section A', a distance ξ'. The separation between A and A' in the deformed state is then

$$dx+(\xi'-\xi)=dx+d\xi$$

where $d\xi=\xi'-\xi$. The deformation of the rod in that region has therefore been $d\xi$. The *normal strain* ϵ in the rod is defined as the deformation along the axis per unit length. Since in this case the deformation $d\xi$ corresponds to a length dx, the strain in the rod is $d\xi/dx$. More generally

$$\epsilon=\frac{\partial\xi}{\partial x}. \tag{10.18}$$

When there is no deformation, ξ is constant and $\epsilon=0$: that is, there is no normal strain when there is no deformation. The strain, being the quotient of two lengths, is a pure number or dimensionless quantity.

The relation that exists between the normal stress $\mathscr{S}$ and the normal strain ϵ of the rod is called *Hooke's law*:

> *within the elastic limit of the material, the normal stress is proportional to the normal strain*;

or

$$\mathscr{S}=Y\epsilon \tag{10.19}$$

where the proportionality constant Y is called *Young's modulus of elasticity* after the Englishman Thomas Young (1773–1829). This law was first stated by the Englishman Robert Hooke (1635–1703). Hooke's law is a good approximation for the elastic behavior of a substance as long as the deformations are small. For large stresses and deformations, Eq. (10.19) no longer holds, and the description of the physical situation becomes much more complicated.

Table 10-1 gives the elastic constants for certain materials. These constants are Young's modulus Y; the bulk modulus κ, defined later in Eq. (10.27); and the modulus of rigidity or shear modulus G, later defined in Eq. (10.38).

Table 10-1. Elastic Constants (10^{11} Pa)

Material	Y	κ	G
Aluminum	0.70	0.61	0.24
Copper	1.25	1.31	0.46
Iron	2.06	1.13	0.82
Lead	0.16	0.33	0.054
Nickel	2.1	1.64	0.72
Steel	2.0	1.13	0.80

Introducing Eqs. (10.17) and (10.18) into Eq. (10.19) and solving for F yields

$$F = YA\frac{\partial \xi}{\partial x}. \tag{10.20}$$

For the case of a rod or wire in equilibrium with one end fixed at point O (Fig. 10-12) and subject at the other end B to a force F, the force at each section must be the same and equal to F. Then integrate Eq. (10.20) with F constant to obtain the deformation at each section:

$$\int_0^\xi d\xi = \frac{F}{YA}\int_0^x dx \qquad \text{or} \qquad \xi = \frac{F}{YA}x.$$

In particular the deformation l at the free end B is obtained by making $x = L$ so that $l = FL/YA$. This relation provides the basis for the experimental measurement of Young's modulus.

When the rod is not in equilibrium, the force is not the same along the rod. As a result, a section of the rod of thickness dx is subject to a net or resultant force. For example in Fig. 10-11 the side B' of the section of thickness dx is subject to a force F' toward the right because of the pull of the right part of the rod; and the side B is subject to a force F pointing to the left because of the pull of the left part of the rod. The net force to the right on the section is $F' - F = dF = (\partial F/\partial x)\,dx$. Given that ρ is the density of the material of the rod, the mass of the section is $dm = \rho\,dV = \rho A\,dx$ where $A\,dx$ is the volume of the section. The acceleration of this mass is $\partial^2\xi/\partial t^2$. Therefore applying the dynamical relation force = mass × acceleration gives the equation of motion of the section as

$$\frac{\partial F}{\partial x}dx = (\rho A\,dx)\frac{\partial^2 \xi}{\partial t^2} \qquad \text{or} \qquad \frac{\partial F}{\partial x} = \rho A\frac{\partial^2 \xi}{\partial t^2}. \tag{10.21}$$

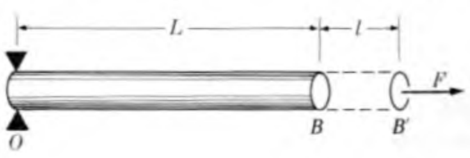


Fig 10-12

In this problem there are *two* fields: one is the displacement ξ of each section of the rod where ξ is a function of position and time; and the other is the force F at each section where F is also a function of position and time. These two fields are related by Eqs. (10.20) and (10.21). These two equations, called the differential equations of the *elastic field* of a deformed rod, describe the physical conditions of the problem. These equations are the mathematical equivalent of Maxwell's equations for the electromagnetic field. Taking the derivative of Eq. (10.20) with respect to x gives

$$\frac{\partial F}{\partial x}=YA\frac{\partial^2 \xi}{\partial x^2}.$$

This equation can be combined with Eq. (10.21) to give the wave equation for the elastic field as

$$\frac{\partial^2 \xi}{\partial t^2}=\frac{Y}{\rho}\frac{\partial^2 \xi}{\partial x^2}. \tag{10.22}$$

This equation is similar to Eq. (10.15), and therefore one may conclude that the deformation field ξ propagates along the rod with a phase velocity

$$v=\sqrt{\frac{Y}{\rho}}, \tag{10.23}$$

a result confirmed experimentally by independently measuring the three quantities. Equation (10.23) checks dimensionally since Y is expressed in Pa and ρ in kg m^{-3}. Therefore their ratio is (Pa) (kg m^{-3})$^{-1}$=(N m^{-2}) (kg m^{-3})$^{-1}$=m^2 s^{-2}, which is the square of a velocity. In view of relation (10.20), it can be shown (see Example 10.5) that the force field F satisfies a similar equation,

$$\frac{\partial^2 F}{\partial t^2}=\frac{Y}{\rho}\frac{\partial^2 F}{\partial x^2}, \tag{10.24}$$

indicating that the force field propagates along the rod with the same phase velocity as the displacement field.

The wave described by Eqs. (10.22) and (10.24) corresponds to physical properties, deformation ξ and force F, oriented along the direction of propagation of the wave; i.e., along the X-axis. This kind of wave motion in which the field variations are in the same direction as the direction of propagation is called *longitudinal wave motion.*

It is important to recognize that the field equations (10.20) and (10.21) imply the wave equations (10.22) and (10.24), but the reverse is not true since other field equations may also imply a wave equation. Therefore the fundamental field equations of the problem are (10.20) and (10.21); the wave equations (10.22) and (10.24) are a consequence of the field equations.

Example 10.4. Velocity of propagation of longitudinal elastic waves in a steel bar.

▼ Using the values of Table 10-1 and a value of 7.8×10^3 kg m^{-3} for the density of steel, we have from Eq. (10.23) that

$$v=\sqrt{\frac{Y}{\rho}}=\sqrt{\frac{2.0 \times 10^{11}\ \text{Pa}}{7.8 \times 10^3\ \text{kg m}^{-3}}}=5.06 \times 10^3\ \text{m s}^{-1}.$$

The experimental value is 5.10×10^3 m s^{-1} at 0°C. This value should be compared with the velocity of sound in air: about 340 m s^{-1}. ▲

Example 10.5. The wave equation for the force field $F(x, t)$ that propagates along a rod.

▼ Beginning with the relation [given by Eq. (10.20)] for the applied force as a function of the variation in the displacement, we may find that the second derivative of F with respect to time is

$$\frac{\partial^2 F}{\partial t^2}=YA\frac{\partial^2}{\partial t^2}\left[\frac{\partial \xi}{\partial x}\right]=YA\frac{\partial}{\partial x}\left[\frac{\partial^2 \xi}{\partial t^2}\right].$$

From Eq. (10.21) we find the acceleration $\partial^2\xi/\partial t^2$ as $(1/\rho A)\partial F/\partial x$, which when substituted into the right-hand side of the equation above yields

$$\frac{\partial^2 F}{\partial t^2}=YA\frac{\partial}{\partial x}\left[\frac{1}{\rho A}\frac{\partial F}{\partial x}\right]=\frac{Y}{\rho}\frac{\partial^2 F}{\partial x^2}.$$

This expression is Eq. (10.24); thus both the force field and the displacement field are shown to propagate along the rod with the phase velocity $\sqrt{Y/\rho}$. ▲

Example 10.6. Longitudinal waves in a spring.

When a disturbance is produced in a stretched spring, and ξ is the displacement suffered by a section of the spring, the force at that section is $F=K(\partial\xi/\partial x)$ where K is the elastic modulus of the spring. This equation is the equivalent of Eq. (10.20) for a bar. The coefficient K should not be confused with the elastic constant k of the spring. To obtain the relation between K and k, note that if the spring of length L is stretched slowly until its length increases by l, the force F must be the same at all points of the spring in equilibrium. Thus $\partial\xi/\partial x=l/L$ and $F=(K/L)l$. Therefore $k=K/L$ or $K=kL$. Now consider a section of the spring of length dx: σ is the mass per unit length of the spring, and $\sigma\, dx$ is the mass of the section. The same logic used to obtain Eq. (10.21) produces

$$\sigma\frac{\partial^2\xi}{\partial t^2}=\frac{\partial F}{\partial x}=K\frac{\partial^2\xi}{\partial x^2} \qquad \text{or} \qquad \frac{\partial^2\xi}{\partial t^2}=\frac{K}{\sigma}\frac{\partial^2\xi}{\partial x^2},$$

which has the form of the wave equation (10.15). Therefore the velocity of propagation of the longitudinal wave along the spring is

$$v=\sqrt{K/\sigma}=\sqrt{kL/\sigma}.$$

This result shows that the velocity of propagation of a disturbance in a spring depends on its mass per unit length (σ). ▲

10.6 Pressure Waves in a Gas Column

Elastic waves in a gas result from pressure variations in the gas. Sound is the most important example of this type of wave. There is an important difference between elastic waves in a gas and elastic waves in a solid rod: gases are very compressible; and when pressure fluctuations are set up in a gas, the density of the gas will suffer the same kind of fluctuations as the pressure.

Consider waves propagated in a gas within a cylindrical pipe or tube. Call p_0 and ρ_0 the equilibrium pressure and density in the gas. Under equilibrium conditions, p_0 and ρ_0 are the same throughout the volume of the gas; that is, they are independent of x. If the pressure of the gas is disturbed, a volume element such as $A\,dx$ in Fig. 10-13 is set in motion because the pressures p and p' on either side of the element differ and give rise to a net force. As a result, section A is displaced an amount ξ; and section A', an amount ξ' so that the thickness of the volume element after the deformation is $dx+(\xi'-\xi)=dx+d\xi$. So far, all seems identical to the case of the solid rod. However because of the change in volume, there is now also a change in density because of the greater compressibility of the gas. The mass within the undisturbed volume element is $\rho_0 A\,dx$. If ρ is the density of the disturbed gas, the mass of the disturbed volume element is $\rho A(dx+d\xi)$. The conservation of matter requires that both masses be equal; that is,

$$\rho A(dx+d\xi)=\rho_0 A\,dx \qquad \text{or} \qquad \rho\left(1+\frac{\partial\xi}{\partial x}\right)=\rho_0,$$

so that

$$\rho=\frac{\rho_0}{1+\partial\xi/\partial x}.$$

Since in general $\partial\xi/\partial x$ is small, $(1+\partial\xi/\partial x)^{-1}$ may be replaced by $1-\partial\xi/\partial x$ using the binomial expansion (M.28). This results in $\rho=\rho_0(1-\partial\xi/\partial x)$ or

$$\rho-\rho_0=-\rho_0\left(\frac{\partial\xi}{\partial x}\right). \tag{10.25}$$

The pressure p is related to the gas density ρ by an equation of state $p=f(\rho)$. If we

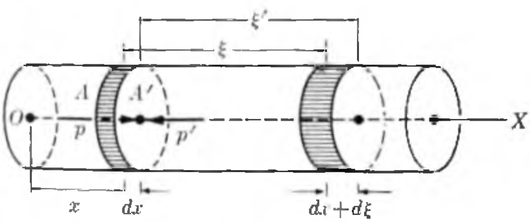


Fig. 10-13. Compressional wave in a gas column.

apply the Taylor expansion technique of (M.31), this function may be written as

$$p=p_0+(\rho-\rho_0)\left(\frac{dp}{d\rho}\right)_{\rho=\rho_0}+\frac{1}{2}(\rho-\rho_0)^2\left(\frac{d^2p}{d\rho^2}\right)_{\rho=\rho_0}+\cdots.$$

For small changes in density we need keep only the first two terms and write

$$p=p_0+(\rho-\rho_0)\left(\frac{dp}{d\rho}\right)_{\rho=\rho_0}. \tag{10.26}$$

The *bulk modulus of elasticity* is defined by the quantity

$$\kappa=\rho_0\left(\frac{dp}{d\rho}\right)_{\rho=\rho_0} \tag{10.27}$$

The bulk modulus is expressed in Pa, the same units used to express pressure. Then after a little manipulation, we may write Eq. (10.26) as

$$p=p_0+\kappa\left(\frac{\rho-\rho_0}{\rho_0}\right). \tag{10.28}$$

This expression corresponds to Hooke's law for fluids. Using Eq. (10.25) to eliminate $(\rho-\rho_0)/\rho_0$ gives

$$p=p_0-\kappa\frac{\partial\xi}{\partial x}. \tag{10.29}$$

This expression relates the pressure at any point in the gas to the deformation at the same point. [This expression is equivalent to Eq. (10.20) for an elastic rod.]

Next we need to determine the equation of motion of the volume element. The mass of the volume element is $\rho_0 A\ dx$ and its acceleration is $\partial^2\xi/\partial t^2$. The gas on the left of the volume element pushes to the right with a force pA, and the gas on the right pushes to the left with a force $p'A$. Therefore the resultant force in the $+X$-direction is $(p-p')A=-A\ dp$ since $dp=p'-p$. The equation of motion is then

$$-A\ dp=(\rho_0 A\ dx)\frac{\partial^2\xi}{\partial t^2}$$

or

$$\frac{\partial p}{\partial x}=-\rho_0\frac{\partial^2\xi}{\partial t^2}. \tag{10.30}$$

Again we have two fields: the displacement field ξ and the pressure field p. Expressions (10.29) and (10.30) are the equations relating both fields. These expressions can be combined in the following way: taking the derivative of Eq. (10.29) with respect to x and remembering that p_0 is constant throughout the gas gives

$$\frac{\partial p}{\partial x}=-\kappa\frac{\partial^2\xi}{\partial x^2}.$$

When this equation is compared with Eq. (10.30), we see that

$$\frac{\partial^2 \xi}{\partial t^2}=\frac{\kappa}{\rho_0}\frac{\partial^2 \xi}{\partial x^2}. \tag{10.31}$$

Once more the displacement wave equation obtained is similar to Eq. (10.15). We conclude that the displacement caused by a pressure disturbance in a gas propagates with a velocity

$$v=\sqrt{\frac{\kappa}{\rho_0}}. \tag{10.32}$$

The pressure also obeys a wave equation like Eq. (10.31) as the student may verify by using Eq. (10.29) with Eq. (10.30) in the same way as the wave equation for the force field was developed in Example 10.5. The pressure wave equation may therefore be written as

$$\frac{\partial^2 p}{\partial t^2}=\frac{\kappa}{\rho_0}\frac{\partial^2 p}{\partial x^2}. \tag{10.33}$$

This equation explains why elastic waves in a gas are often described as *pressure waves*. Sound is simply a pressure wave in air. An explosion, which is a sudden local increase in pressure, sets up a *blast pressure wave*; but in this case the fluctuations in density are usually so large that the approximations made in the theoretical development above are no longer valid; and a more complex equation results such that the wave does not propagate undistorted.

Similarly the student [combining Eq. (10.25) with Eq. (10.31)] may verify that the gas density obeys a wave equation of the same form; that is,

$$\frac{\partial^2 \rho}{\partial t^2}=\frac{\kappa}{\rho_0}\frac{\partial^2 \rho}{\partial x^2}.$$

Therefore when referring to a gas, one may speak of a displacement wave, a pressure wave, and a density wave. Displacement waves resemble the picture of waves on a liquid surface (i.e., the motion of matter in bulk). Although they do not correspond to such a physical picture, pressure and density waves also describe a physical disturbance propagated in a gas.

Wave motion in gases is generally an *adiabatic* process, which means that no energy is exchanged in the form of heat by a volume element of the gas. Under adiabatic conditions $p=C\rho^{\gamma}$ where γ is a quantity characteristic of each gas. For most diatomic gases the value of γ is very close to 1.4. Then $dp/d\rho=\gamma C\rho^{\gamma-1}$, and $\kappa=\rho_0(dp/d\rho)_0=\gamma C\rho_0^{\gamma}=\gamma p_0$. Then dropping the subscript 0 and substituting into Eq. (10.32), the velocity of sound in a gas may be written

$$v=\sqrt{\frac{\gamma p}{\rho}}. \tag{10.34}$$

The wave associated with the ξ field is again a longitudinal wave since the displacement is parallel to the direction of propagation. Since the pressure p is not a vector, no direction is associated with the pressure wave. Therefore the wave motion corresponding to the pressure field is a *scalar* wave. The associated direction is that of the *force* produced by the pressure difference; that is, normal to the cross-sectional area of the cylindrical tube. (The wave corresponding to the density ρ is also scalar.)

All waves discussed so far in this chapter fall into the category of *elastic waves*, in which the disturbance—whether it is a strain, a pressure, or a bulk displacement involving many atoms—propagates with a velocity depending on the elastic properties of the medium.

Elastic waves are also called acoustic waves. Whenever an elastic wave propagating through a gas, a liquid, or a solid reaches the ear, the wave produces vibrations in the ear's membrane resulting in a sensation called *sound*. The response of the human nervous system constitutes the process known as *hearing*. The human nervous system responds to sound frequencies between approximately 16 Hz and 20,000 Hz. Outside these limits, sounds are inaudible to human beings, but the elastic waves are still called sound. (The frequency range is different for other animals.) Elastic waves with frequencies above 20,000 Hz are called *ultrasonic*; the elastic wave region below 16 Hz is called *infrasonic*.

Sound involves the displacement of atoms and molecules of the medium through which sound propagates, but this displacement is an ordered collective motion in which all atoms in a small volume suffer essentially the same displacement. This ordered motion is then superposed on the random or disordered molecular agitation of liquids and gases. The net result is that the intensity of sound decreases or is *attenuated* while the sound wave propagates because some of the wave's energy is taken away by the molecules of the medium after collisions. This attenuation results in an increase in the molecular internal energy, mainly rotational molecular motion, or in the translational kinetic energy. In liquids the viscosity, which in essence is an effect of the molecular motion, also plays an important role in sound attentuation.

The velocity of propagation of sound is practically independent of frequency for a very large range of frequencies, extending up to more than 10^8 Hz. The value of this velocity for different substances is given in Table 10-2. The velocity of propagation is, however, rather sensitive to temperature and pressure changes because of the dependence of velocity on density. Many of the wave phenomena that will be described in succeeding chapters apply to sound waves.

Table 10-2. Velocity of Sound, m s^{-1}

Solids (20°C)		Liquids (25°C)		Gases (0°C)	
Granite	6000	Fresh water	1493.2	Air	331.45
Iron	5130	Sea water	1532.8	Hydrogen	1269.5
Copper	3750	(3.6% salinity)		Oxygen	317.2
Aluminum	5100	Kerosene	1315	Nitrogen	339.3
Lead	1230	Mercury	1450	Steam (100°C)	404.8

Example 10.7. The relation between the velocity of a pressure wave in a gas and the temperature of the gas.

▼ The relation between pressure and volume in a gas is $pV=NRT$ where N is the number of moles of gas. Since $\rho=m/V$, we may write $p/\rho=NRT/m=RT/M$ where $M=m/N$ is the mass of one mole of the gas, expressed in kg. Therefore the ratio p/ρ is proportional to the temperature; and from Eq. (10.34)

$$v=\sqrt{\gamma p/\rho}=\sqrt{\gamma RT/M}=\alpha\sqrt{T} \tag{10.35}$$

where $\alpha=\sqrt{\gamma R/M}$. From experimental measurements it has been found that at $T=273.15$ K (or 0° C) the velocity of sound in air is 331.45 m s^{-1}; therefore the coefficient α has the value 20.055 m s^{-1} $T^{-\frac{1}{2}}$. The velocity of sound in air at any temperature (measured in K) is therefore $v=20.055\sqrt{T}$ m s^{-1}, a result that agrees with experiment over a fairly large temperature range. ▲

Example 10.8. The relation between the amplitudes of the displacement waves and the pressure waves in a gas column.

▼ Suppose that the displacement waves are harmonic, expressed by $\xi=\xi_0 \sin(kx-\omega t)$. Substituting this result in Eq. (10.29) gives

$$p-p_0=-\kappa\frac{\partial\xi}{\partial x}=-\kappa k\xi_0\cos(kx-\omega t).$$

Thus the pressure wave oscillates about its average value with an amplitude given by $\mathscr{P}_0=\kappa k\xi_0$; or using Eq. (10.32) to eliminate κ

$$\mathscr{P}_0=v^2\rho_0 k\xi_0.$$

An alternative expression for $\mathscr{P}_0$ is obtained by using the relation given in Eq. (10.6), namely $k=\omega/v$. Then

$$\mathscr{P}_0=v\rho_0\omega\xi_0=2\pi\rho_0 v\nu\xi_0.$$

These relations are extremely useful in acoustic calculations. For example at a frequency of 400 Hz, the faintest sound that can be heard corresponds to a pressure amplitude of about 8×10^{-5} Pa. The corresponding displacement amplitude with an air density of 1.29 kg m^{-3} and a velocity of sound of 345 m s^{-1} is

$$\xi_0=\frac{\mathscr{P}_0}{2\pi v\rho_0\nu}=7.15\times10^{-11}\text{ m.} \quad ▲$$

10.7 Transverse Waves in a String

Consider next a wave in a string subject to a tension T. Under equilibrium conditions the string is straight. Suppose now that the string is displaced sidewise or perpendicular to its length by an amount small relative to the length as shown in Fig. 10-14.

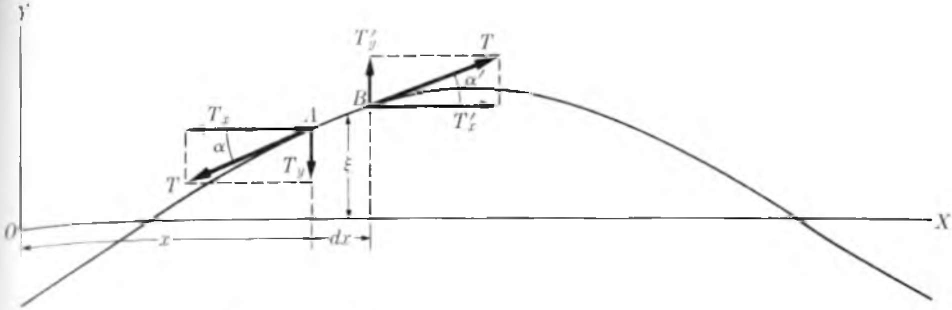


Fig. 10-14. Forces on a section of a transversely displaced string.

Consider a section AB of the string, of length dx, that has been displaced a distance ξ from the equilibrium position. On each end of section AB, a tangential force T is acting: the one at B is produced by the pull of the string on the right; and the one at A, by the pull of the string on the left. Because of the curvature of the string, the two forces are not directly opposed. The vertical or Y-component of each force is $T_y' = T \sin \alpha'$ and $T_y = -T \sin \alpha$. The resultant normal force on the section AB of the string is

$$F_y = T(\sin \alpha' - \sin \alpha).$$

If the curvature of the string is not very large, the angles α and α' are small; and the sines can be replaced by the tangents. So the normal or transverse force in the upward direction is

$$F_y = T(\tan \alpha' - \tan \alpha),$$

which may also be written as

$$F_y = T\, d(\tan \alpha) = T \frac{\partial}{\partial x} (\tan \alpha)\, dx$$

where the partial derivative is used because $\tan \alpha$ depends not only on the position x but also on the time t. However, $\tan \alpha$ is the slope of the string; and by definition the slope is equal to $\partial\xi/\partial x$. Then

$$F_y = T \frac{\partial}{\partial x}\left(\frac{\partial \xi}{\partial x}\right) dx = T \frac{\partial^2 \xi}{\partial x^2}\, dx.$$

This force must be equal to the mass of the section AB times its upward acceleration $\partial^2\xi/\partial t^2$. Given that σ is the *linear density* of the string, or mass per unit length, expressed in kg m^{-1}, the mass of the section AB is $\sigma\, dx$; and based on the relation force = mass × acceleration, the equation of motion of this section of the string may be written as

$$(\sigma\, dx) \frac{\partial^2 \xi}{\partial t^2} = T \frac{\partial^2 \xi}{\partial x^2}\, dx$$

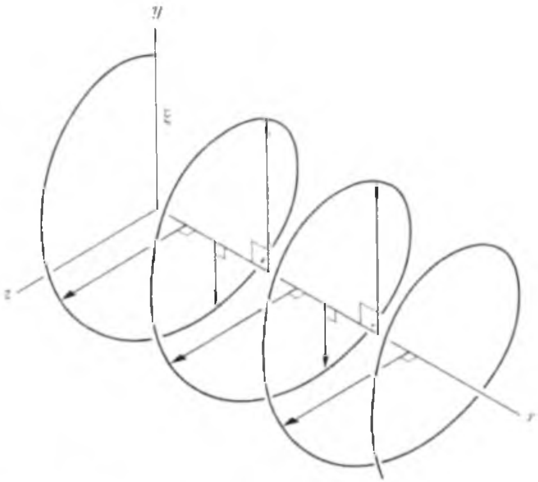


Fig. 10-15. Nonpolarized transverse wave in a string.

or

$$\frac{\partial^2 \xi}{\partial t^2} = \frac{T}{\sigma} \frac{\partial^2 \xi}{\partial x^2}. \tag{10.36}$$

Once more we obtain Eq. (10.15), verifying that a transverse disturbance in a string propagates along the string with a velocity

$$v = \sqrt{\frac{T}{\sigma}} \tag{10.37}$$

provided that the amplitude is small. The student should check the consistency of the units in this equation.

This example differs from the previous one in two important respects. One difference is that there is only one field, the displacement ξ; and the wave equation (10.36) is a direct result of the equation of motion. The second and more important difference is that the wave motion is *transverse*; that is, the displacement ξ is perpendicular to the direction of the wave's propagation, which is along the X-axis. However there are many directions of displacement perpendicular to the X-axis. If two mutually perpendicular directions Y and Z are chosen as references, the transverse displacement ξ (which must be considered a vector) may be expressed in terms of its components along the Y- and Z-axes. While the disturbance propagates, the direction of ξ may change from point to point; and a twisting of the string results (Fig. 10-15). However if all the displacements are in the same direction, say along the Y-axis, the string is always in the XY-plane, and the wave motion is *linearly polarized* (Fig. 10-16). A transverse wave can always be considered as the combination of two waves linearly polarized in perpendicular directions. When ξ has a constant magnitude

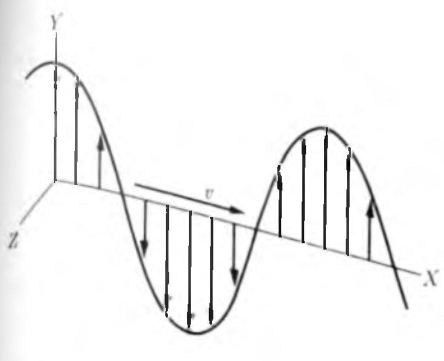


Fig. 10-16. Linearly polarized transverse wave in a string.

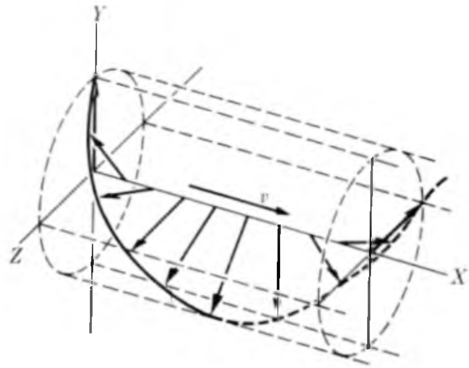


Fig. 10-17. Circularly polarized transverse wave in a string.

but changes in direction so that the string lies over a circular cylindrical surface (Fig. 10-17), the wave is *circularly polarized*. In this case, each portion of the string moves in a circle around the X-axis. The polarization of transverse waves is a very important subject and will be discussed in more detail in Chapter 14.

Note that Eq. (10.36) takes into account only the transverse motion of the string because there is essentially no motion along the string. To see this point, consider the resultant force parallel to the X-axis:

$$F_x = T\cos\alpha' - T\cos\alpha = T(\cos\alpha' - \cos\alpha).$$

When the angle is very small, the cosine is essentially one. Therefore to the first order of approximation, $\cos\alpha' \approx \cos\alpha$ and $F_x = 0$ so that there is no net force parallel to the X-axis and therefore no resultant motion of the string in the X-direction.

Example 10.9. Transverse elastic waves in a bar. A shear wave.

▼ In Section 10.5 longitudinal elastic waves in a solid bar were studied. Now we shall analyze transverse elastic waves. Consider a bar that in its undistorted state is represented by the dashed horizontal lines in Fig. 10-18. If we start the bar vibrating by hitting it transversely, at a particular instant, it adopts the shape of the solid curved lines; and we may assume that each section of the

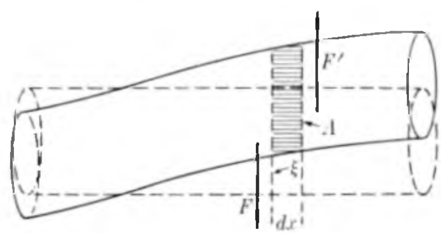


Fig. 10-18. Transverse or shear wave in a clamped rod.

bar moves up and down with no horizontal motion. Call ξ the transverse displacement of a section dx at a particular time. This displacement must also be a function of position because if it were constant, it would correspond to a parallel displacement of the bar. The quantity $\gamma = \partial\xi/\partial x$, which is the change of transverse displacement per unit length along the bar, is defined as the *shearing strain*.

As a result of the deformation, each section of thickness dx is subject to the opposing forces F and F'; these forces are tangent to the surface (compare with the situation in Fig. 10-11) and are produced by those portions of the bar on each side of the section. The tangential force per unit area, $\mathscr{S} = F/A$, is defined as the *shearing stress*.

Similar to Hooke's law, Eq. (10.19), that expresses the relationship between normal stress and normal strain, there is a relation between the shearing stress and the shearing strain given by

$$\mathscr{S} = G\gamma, \tag{10.38}$$

where G is a coefficient characteristic of the material and is called the *shear modulus*. Therefore using the expressions for $\mathscr{S}$ and γ we have

$$F = AG\frac{\partial\xi}{\partial x}. \tag{10.39}$$

The resultant force on the section shown in Fig. 10-18 is $F' - F = dF = (\partial F/\partial x)\,dx$. If ρ is the density of the material, the mass of the section is $\rho A\,dx$; and its equation of motion in the transverse direction is

$$\frac{\partial F}{\partial x}dx = (\rho A\,dx)\frac{\partial^2\xi}{\partial t^2}$$

or

$$\frac{\partial F}{\partial x} = \rho A\frac{\partial^2\xi}{\partial t^2}. \tag{10.40}$$

However from Eq. (10.39) after taking the derivative with respect to x, we have

$$\frac{\partial F}{\partial x} = AG\frac{\partial^2\xi}{\partial x^2};$$

when this expression is substituted in Eq. (10.40) and the common factor A is canceled, the result is

$$\frac{\partial^2\xi}{\partial t^2} = \frac{G}{\rho}\frac{\partial^2\xi}{\partial x^2}. \tag{10.41}$$

Again we obtain the differential equation of wave motion (10.15), indicating that the transverse deformation propagates along the bar with a velocity given by

$$v = \sqrt{\frac{G}{\rho}}. \tag{10.42}$$

More properly, the wave should be called a *shear wave*. ▲

Example 10.10. Transverse elastic waves in a bar. A torsional wave.

▼ Another example of a shear wave is a *torsional wave*. Suppose that a variable torque is applied at the free end of a rod clamped at one point. This torque produces a twisting of the rod (Fig. 10-19). If the torque is time dependent, the angle of twist, which is called torsion, changes with time;

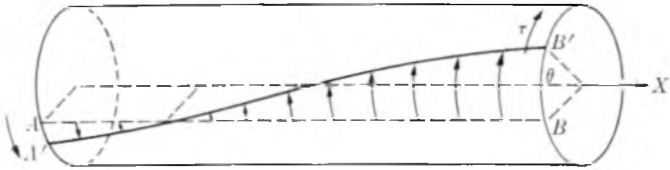


Fig. 10-19. Torsional wave in a clamped rod.

the result is a torsional wave propagated along the rod. A mathematical analysis of the problem shows that irrespective of the shape of the cross section of the rod, the velocity of propagation of the torsional wave is also given by Eq. (10.42). It is not surprising that the shear and torsional waves in a bar propagate with the same velocity since both processes are essentially due to the same internal phenomenon in the material of the rod. Another interesting aspect of torsional waves is that they correspond not to displacements parallel to or transverse to the axis of the rod, but to rotations around the axis without change in shape. These various examples have been given to help the student understand the great variety of elastic-wave phenomena: all have different internal dynamics; but under the approximations here used, all are described mathematically by the same equation: Eq. (10.15). ▲

10.8 Surface Waves in a Liquid

As a final example of wave motion in one direction, consider waves on the surface of a liquid. These are the most familiar kinds of waves; they are the waves observed on the oceans and lakes, or simply when a stone drops in a pond. The mathematical aspect, however, is more complicated than that in the previous examples, and will be omitted. Instead we present a description of such waves with a simplified mathematical discussion deferred until Examples 10.11 and 10.12.

The undisturbed surface of a liquid is plane and horizontal. A disturbance of the surface produces a displacement of all molecules directly underneath the surface (Fig. 10-20). Each volume element of the liquid describes a closed path. The amplitude of the horizontal and the vertical displacements of a volume element of a fluid varies in general with the depth. Of course the molecules at the bottom suffer hardly any vertical displacement since they must remain close to the bottom. At the surface of the liquid certain forces enter into play in addition to that due to the atmospheric pressure. One force, due to the *surface tension* of the liquid, gives an upward force on an element of the surface similar to that found in the case of a string. Another force is the weight of the liquid above the undisturbed level. The resulting equation for the surface displacement is not exactly of type (10.15), but slightly more complicated. However for harmonic waves of wavelength λ, the velocity of propagation of the

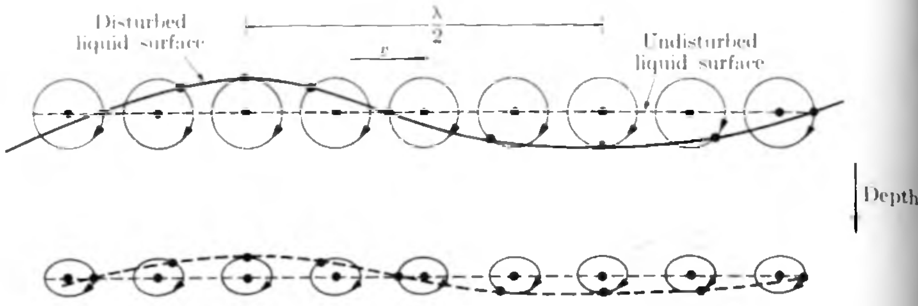


Fig. 10-20. Molecule displacement resulting from a surface wave in a liquid.

surface wave is given by

$$v=\sqrt{\frac{g\lambda}{2\pi}+\frac{2\pi\mathscr{T}}{\rho\lambda}} \tag{10.43}$$

where ρ is the density of the liquid. $\mathscr{T}$ is the surface tension, and g is the acceleration of gravity. In Eq. (10.43) the surface tension is expressed in N m^{-1}. This expression is valid only when the depth is very great compared with the wavelength λ. Otherwise a different expression results (see Example 10.11).

The most interesting aspect of Eq. (10.43) is that *the velocity of propagation depends on the wavelength*, a situation not encountered in any of the previous examples. Since the frequency is related to the wavelength and the velocity of propagation through $\nu=v/\lambda$, we conclude that *the velocity of propagation is a function of the frequency of the wave.*

When λ is large enough so that the second term in Eq. (10.43) can be neglected,

$$v=\sqrt{\frac{g\lambda}{2\pi}}. \tag{10.44}$$

The waves in this case are called *gravity waves*. In this approximation the velocity of propagation is independent of the nature of the liquid since no factor pertaining to the liquid (such as its density or its surface tension) appears in Eq. (10.44). In this case the velocity of propagation is proportional to the square root of the wavelength; and the longer the wavelength, the faster the propagation. For this reason a strong, steady wind produces waves of longer wavelength than a swift, gusty wind.

When the wavelength is very small, the dominant term is the second in Eq. (10.43), which gives the velocity of propagation as

$$v=\sqrt{\frac{2\pi\mathscr{T}}{\rho\lambda}}. \tag{10.45}$$

These waves are called *ripples* or *capillary waves*. They are the waves observed when a very gentle wind blows over the water, or when a liquid in a container is subject to vibrations of high frequency and small amplitude. In this case the longer the wavelength, the slower the propagation.

When the velocity of propagation of a wave motion depends on the wavelength or the frequency, we say that there is *dispersion*. If a wave motion resulting from the superposition of several harmonic waves of different frequencies impinges on a *dispersive* medium, the wave is distorted since each of its component waves propagates with a different velocity. Dispersion is a very important phenomenon present in several types of wave propagation. In particular dispersion appears in the case of electromagnetic waves propagating through matter as we shall see in the next chapter.

Example 10.11. The propagation of waves in a liquid of finite depth.

▼ The general expression for the velocity of propagation of surface waves in a liquid is

$$v=\sqrt{\left(\frac{g\lambda}{2\pi}+\frac{2\pi\mathcal{T}}{\rho\lambda}\right)\tanh\frac{2\pi h}{\lambda}} \tag{10.46}$$

where h is the depth of the liquid. When the depth h is very large compared with the wavelength (that is, the quantity $2\pi h/\lambda$ is large compared with unity), the value of the hyperbolic tangent is very close to 1, and therefore the last factor in Eq. (10.46) can be replaced by 1 without great error. In this approximation Eq. (10.46) becomes Eq. (10.43).

On the other hand when the depth h is very small compared with the wavelength λ, the quantity $2\pi h/\lambda$ is very small compared with unity; and since $\tanh x \sim x$ when x is very small, the last factor in Eq. (10.46) may be replaced by $2\pi h/\lambda$. Also, the term $2\pi\mathcal{T}/\rho\lambda$ can be neglected when the wavelength is large. Then

$$v=\sqrt{\frac{g\lambda}{2\pi}\cdot\frac{2\pi h}{\lambda}}=\sqrt{gh}. \tag{10.47}$$

In these circumstances the velocity of propagation is independent of the wavelength and is a function only of depth. ▲

Example 10.12. Surface waves in a liquid when the wavelength is very large and the amplitude is very small compared with the depth.

▼ Consider a liquid in a channel of depth h and width L. If the surface of the liquid is perturbed with waves of small amplitude and large wavelength (compared with h), a particular vertical section of the liquid of width dx suffers some displacement in both the vertical and the horizontal directions. As a result of these displacements, the width of the section changes from dx to $dx+d\xi$ (Fig. 10-21); and its height, from h to $h+\eta$. Assuming the liquid to be incompressible, the volume of the section must remain constant. Therefore

$$\begin{aligned} Lh\,dx &= L(h+\eta)(dx+d\xi) \\ &= L(h\,dx+\eta\,dx+h\,d\xi+\eta\,d\xi). \end{aligned}$$

Because η is very small compared with h, and $d\xi$ is very small compared with dx, the last term

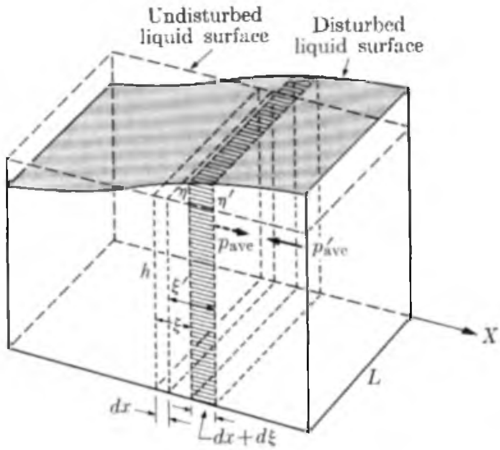


Figure 10-21

$\eta\, d\xi$ may be neglected. After cancellation of equivalent terms we have

$$\eta\, dx + h\, d\xi = 0 \qquad \text{or} \qquad \eta = -h\frac{\partial \xi}{\partial x}. \tag{10.48}$$

Equation (10.48) relates the vertical surface displacement to the horizontal displacement for an incompressible liquid.

Because the disturbed level is not horizontal, the average pressure on each side of the fluid section is different as shown in the figure. If $A = hL$ is the area of the cross section of the undisturbed liquid, the net force to the right on the section is

$$p_{ave}A - p'_{ave}A = -(p'_{ave} - p_{ave})A = -A\, dp_{ave}.$$

The horizontal motion of the section is thus

$$(\rho A\, dx)\frac{\partial^2 \xi}{\partial t^2} = -A\, dp_{ave} \qquad \text{or} \qquad \rho\frac{\partial^2 \xi}{\partial t^2} = -\frac{\partial p_{ave}}{\partial x}.$$

However from the hydrostatic pressure relation, $p = \rho g z$, the pressure difference is

$$dp_{ave} = \rho g(\eta' - \eta) = \rho g\frac{\partial \eta}{\partial x}\, dx$$

so that $\partial p_{ave}/\partial x = \rho g\, \partial\eta/\partial x$, and the preceding equation may be written as

$$\frac{\partial^2 \xi}{\partial t^2} = -g\frac{\partial \eta}{\partial x}.$$

Differentiation of Eq. (10.48) with respect to x gives

$$\frac{\partial \eta}{\partial x} = -h\frac{\partial^2 \xi}{\partial x^2}.$$

Therefore elimination of $\partial\eta/\partial x$ between these two equations finally yields

$$\frac{\partial^2 \xi}{\partial t^2}=gh\frac{\partial^2 \xi}{\partial x^2}.$$

This expression is again the wave equation (10.15) corresponding to waves propagating with velocity $v=\sqrt{gh}$, and agrees with the result obtained in Eq. (10.47) for the circumstances assumed there. Because of relation (10.48), the vertical displacement at the surface satisfies a similar equation: that is,

$$\frac{\partial^2 \eta}{\partial t^2}=gh\frac{\partial^2 \eta}{\partial x^2}. \quad \blacktriangle$$

10.9 What Propagates in Wave Motion

It is important to understand what is propagated as a wave in wave motion. Generally what propagates is a physical disturbance generated at some place; as a consequence of the nature of the phenomenon, the condition may be transmitted to other regions. Since this explanation is somewhat abstract, let us try to formulate it in more concrete terms.

Consider the different kinds of waves discussed in the previous sections. All these waves correspond to certain kinds of motion of atoms or molecules of the medium through which the wave propagates; but on the average the atoms remain at their equilibrium positions (Fig. 10-22). Then it is not matter that propagates but the *state of motion* of matter. It is a dynamic condition that is transferred from one region to another. We are in the habit of describing a dynamic condition in terms of momentum and energy. Therefore, we say that

energy and momentum are transferred or propagated in wave motion.

Look, for example, at the case of longitudinal elastic waves propagating along a rod. At a particular section that is displaced with velocity $\partial\xi/\partial t$ (Fig. 10-11), the right side of the rod pulls on the left side with a force F and the left side pulls on the right side with a force $-F$. Then the power (or work per unit time) that the left side transmits to the right side at that section is

$$\frac{\partial W}{\partial t}=(-F)\frac{\partial \xi}{\partial t}.$$

Therefore when the disturbance passes from one section to another, this power must be transmitted. If the wave propagates from left to right, energy must be fed into the left end of the rod. If the energy is fed in during a short time interval, a disturbance of limited length, or a transient pulse, is produced. If there is to be a continuous train of waves, then energy must be supplied continuously at the left end.

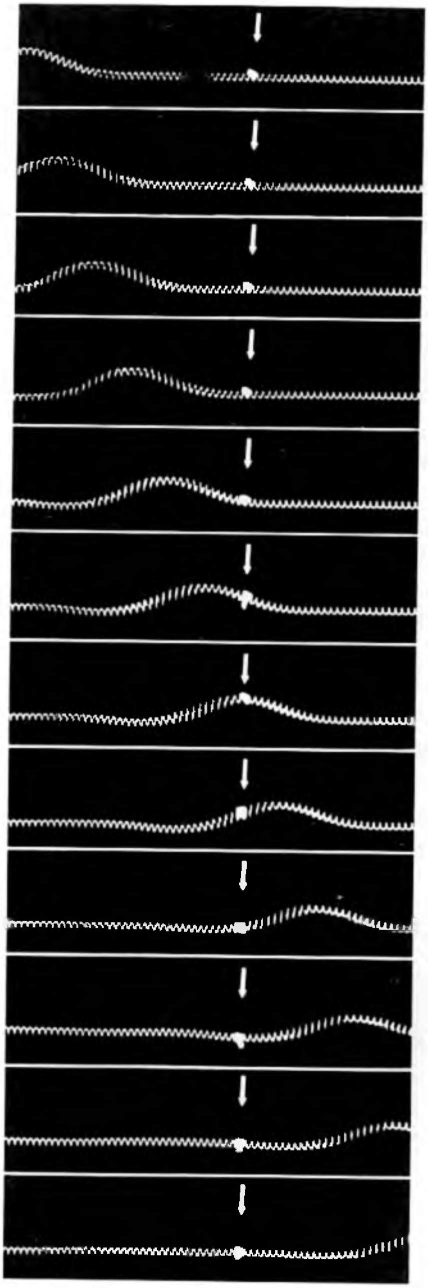


Fig. 10-22. Propagation of a pulse on a spring. The sections of the spring move up and down as the pulse moves from left to right.

To see the problem in more detail, consider the case of a sinusoidal elastic wave, $\xi = \xi_0 \sin(kx - \omega t)$. Taking the appropriate derivatives, we find that

$$\frac{\partial \xi}{\partial t} = -\omega \xi_0 \cos(kx - \omega t)$$

and

$$F = YA \frac{\partial \xi}{\partial x} = YAk\xi_0 \cos(kx - \omega t)$$

where A is the cross-sectional area of the rod. Then, with the relations $\omega = kv$ and $v = \sqrt{Y/\rho}$, the power is

$$\frac{\partial W}{\partial t} = F\left(\frac{\partial \xi}{\partial t}\right) = YA\omega k \xi_0^2 \cos^2(kx - \omega t)$$

$$= (\rho v^2) A(\omega^2/v) \xi_0^2 \cos^2(kx - \omega t)$$

$$= vA[\rho\omega^2 \xi_0^2 \cos^2(kx - \omega t)].$$

The presence of the factor $\cos^2(kx - \omega t)$ assures us that $\partial W/\partial t$ is always positive, although fluctuating. Since $\partial W/\partial t$ depends on $kx - \omega t$, it also satisfies the wave equation and corresponds to an *energy wave*. The average power is

$$\left(\frac{\partial W}{\partial t}\right)_{\text{ave}} = vA\{\rho\omega^2 \xi_0^2 [\cos^2(kx - \omega t)]_{\text{ave}}\}.$$

However $[\cos^2(kx - \omega t)]_{\text{ave}} = \frac{1}{2}$ so that

$$\left(\frac{\partial W}{\partial t}\right)_{\text{ave}} = vA(\tfrac{1}{2}\rho\omega^2 \xi_0^2). \tag{10.49}$$

Now remembering Eq. (12.11) of Volume I, which gives the total energy of an oscillator as $\frac{1}{2}m\omega^2\xi_0^2$, and that instead of the mass m we have the density ρ, we see that

$$\mathrm{E} = \tfrac{1}{2}\rho\omega^2\xi_0^2 \tag{10.50}$$

is the energy per unit volume, or the *energy density* in the rod due to the oscillations resulting from the wave motion. Substituting Eq. (10.50) into Eq. (10.49) gives

$$\left(\frac{\partial W}{\partial t}\right)_{\text{ave}} = vA\mathrm{E}. \tag{10.51}$$

Since v is the velocity of propagation, $v\mathrm{E}$ is the energy flow through the unit area per unit time. Multiplying this quantity by the area A gives the energy flow per unit time through a cross section of the rod. Thus Eq. (10.51) may be interpreted as indicating an average energy flow along the rod as a result of the wave motion.

Expressed in W m^{-2}, the average energy flow per unit area and unit time is

$$I = \frac{1}{A}\frac{\partial W}{\partial t} = v\mathrm{E}. \tag{10.52}$$

This quantity is called the *intensity* of the wave. From Eq. (10.50) for the energy density, the intensity is proportional not only to the square of the frequency of oscillation but also to the square of the amplitude. (The student may verify that similar results hold for pressure waves in a gas and for transverse waves in a string.)

In conclusion we may again say that in all wave motions, energy and momentum are transferred from one place to another along the wave.

Example 10.13. Intensity of the waves in a gas column expressed in terms of the amplitude of the pressure wave.

▼ From Example 10.8 at the end of Section 10.6, the amplitudes of the pressure and displacement waves were found to be related by $\mathscr{P}_0 = 2\pi v \rho_0 \nu \xi_0$. Therefore the energy density of the wave is

$$E = \tfrac{1}{2}\rho_0 \omega^2 \xi_0^2 = 2\pi^2 \rho_0 \nu^2 \xi_0^2 = \frac{\mathscr{P}_0^2}{2v^2 \rho_0};$$

and according to Eq. (10.52), the intensity of the wave is

$$I = vE = \frac{\mathscr{P}_0^2}{2v\rho_0}.$$

The sensitivity of the human ear is such that for each frequency there is a minimum intensity or *threshold of hearing*, below which sound is not audible, and a maximum intensity or *threshold of feeling*, above which sound produces discomfort or pain. This situation is illustrated for each frequency by the two curves of Fig. 10-23, which indicates also the intensity and pressure amplitudes. Note that intensity is also expressed by means of another unit called the *decibel*. The

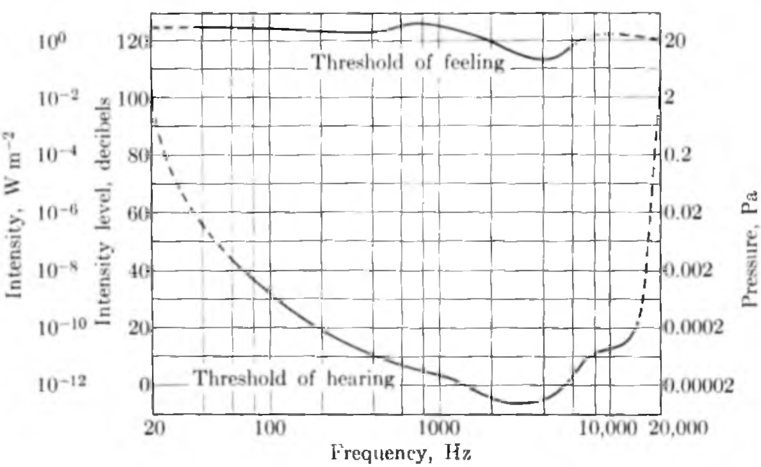


Fig. 10-23. Average auditory range for the human ear.

intensity level of sound (or of any wave motion) is indicated by B and is expressed in decibels (db) according to the definition

$$B = 10 \log \frac{I}{I_0} \tag{10.53}$$

where I_0 is a reference intensity. For the case of sound in air the reference level has been arbitrarily chosen as 10^{-12} W m^{-2}. For example to the pressure amplitude given in Example 10.8 for the faintest sound that can be heard at 400 Hz, there corresponds an intensity of 7.2×10^{-12} W m^{-2} and an intensity level of 8.57 db. ▲

10.10 Group Velocity

The velocity $v = \omega/k$, given by Eq. (10.6) for a harmonic wave of angular frequency ω and wavelength $\lambda = 2\pi/k$, is called the *phase* velocity. However, this velocity is not necessarily that with which a wave motion travels. For a continuous harmonic wave (also called a wave train of infinite length), the wave has only a single wavelength and a single frequency. A wave of this nature is not adequate for transmitting a signal because a signal implies something that begins at a certain time and ends at a certain later time; i.e., the wave must have a shape similar to that indicated in Fig. 10-24. A wave with such a shape is called a *pulse* or a *wave packet*. Measuring the time for this signal to travel between two points implies measuring the velocity with which this pulse travels.

As a first consideration, we may say that this velocity is just the phase velocity $v = \omega/k$ since we kept saying in all previous sections that this is the velocity of propagation of the waves. However an important factor enters here. The wave or pulse depicted in Fig. 10-24 is *not* harmonic since its amplitude is not constant along the X-axis. Thus we must make a Fourier analysis of the wave. When we do, it is clear that the pulse actually contains several frequencies and wavelengths. Of course if the velocity of propagation is independent of the frequency (i.e., if there is no dispersion), then all Fourier components of the wave travel with the same speed, and it is correct to say

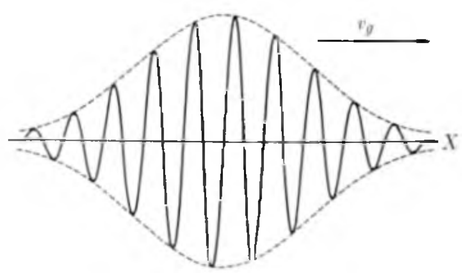


Fig. 10-24. Wave pulse or wave packet.

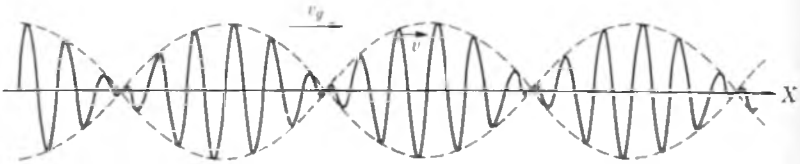


Fig. 10-25. Phase and group velocity.

that the velocity of the pulse is the same as the phase velocity. However in a dispersive medium each Fourier component has its own velocity of propagation, and therefore the situation requires more careful examination.

For simplicity consider a case in which the wave motion may be broken down into only two frequencies ω and ω', which are almost equal so that $\omega' - \omega$ is very small. Assume also that their amplitudes are the same. Then a linear superposition of the two waves gives

$$\begin{aligned}\xi &= \xi_0 \sin(kx - \omega t) + \xi_0 \sin(k'x - \omega' t)\\ &= \xi_0[\sin(kx - \omega t) + \sin(k'x - \omega' t)]\\ &= 2\xi_0 \cos \tfrac{1}{2}[(k' - k)x - (\omega' - \omega)t] \sin \tfrac{1}{2}[(k' + k)x - (\omega' + \omega)t].\end{aligned}$$

Since ω and ω' as well as k and k' are almost equal, $\frac{1}{2}(\omega' + \omega)$ may be replaced by ω, and $\frac{1}{2}(k' + k)$ by k so that

$$\xi = 2\xi_0 \cos \tfrac{1}{2}[(k' - k)x - (\omega' - \omega)t] \sin(kx - \omega t). \tag{10.54}$$

Equation (10.54) represents a wave motion whose amplitude is *modulated*. The modulation is given by the factor

$$2\xi_0 \cos \tfrac{1}{2}[(k' - k)x - (\omega' - \omega)t].$$

This modulation has been indicated in Fig. 10-25. The modulating amplitude itself corresponds to a wave motion propagating with a velocity

$$v_g = \frac{\omega' - \omega}{k' - k} = \frac{d\omega}{dk}, \tag{10.55}$$

which is called *group velocity*. This velocity is that with which the amplitude wave, represented by the dashed line in Fig. 10-25, propagates. It may also be concluded that the maximum of the pulse in Fig. 10-25 propagates with the group velocity v_g. Therefore in a dispersive medium the signal velocity is the group velocity. Since $\omega = kv$, Eq. (10.55) becomes

$$v_g = v + k\frac{dv}{dk}. \tag{10.56}$$

If the phase velocity is independent of the wavelength, $dv/dk = 0$ and $v_g = v$. Therefore in nondispersive media there is no difference between phase velocity and group velocity as was previously inferred; but in a dispersive medium the group velocity may be larger or smaller than the phase velocity.

Although derived for the case of only two frequencies. Eq. (10.56) also holds true for the case of a pulse containing frequencies in the range from $\omega-\Delta\omega$ to $\omega+\Delta\omega$. However, this matter is really more complex than has been shown, and a thorough discussion of it is beyond the scope of this book.

Example 10.14. Group velocity for surface waves.

▼ Consider the case of surface waves in a liquid in the long wave approximation. In this case the phase velocity is given by Eq. (10.44); and since $k=2\pi/\lambda$, $v=\sqrt{g\lambda/2\pi}=\sqrt{g/k}$. Then

$$\frac{dv}{dk}=-\frac{1}{2k}\sqrt{\frac{g}{k}}=-\frac{v}{2k},$$

and Eq. (10.56) gives $v_g=\frac{1}{2}v$ so that the group velocity is just half the phase velocity. Thus if a long wave disturbance is produced in water, the initial disturbance is distorted in such a way that the components of longer wavelength "escape" from the disturbance by moving faster than the group velocity, which is the velocity of the peak of the disturbance. ▲

10.11 The Doppler Effect

When an observer and the source of a wave are in relative motion with respect to the material medium in which the waves propagate, the frequency of the waves observed is different from the frequency of the source. This phenomenon is called the *Doppler effect*, after the German-born Austrian physicist C. J. Doppler (1803–1853), who first noticed it in sound waves.

Suppose that a wave source, such as a vibrating body, moves in the positive X-direction (Fig. 10-26) with velocity v_S through a still medium such as air or water. Observing the waves emitted at several positions 1, 2, 3, 4, . . . , we note that after a time t counted from the time when the source was at position 1, the waves emitted at the several positions occupy the spheres 1, 2, 3, 4, . . . that are not concentric. The waves are more closely spaced on the side in which the body is moving and are more widely separated on the opposite side. To an observer at rest on either side, this spacing corresponds respectively to a shorter and a longer effective wavelength or to a larger and a smaller effective frequency. However if the observer is also in motion with velocity v_O, the waves will be observed arriving at a different rate. For example if the observer is approaching from the right of the source, an even shorter wavelength or higher frequency will be observed since the observer is moving into the

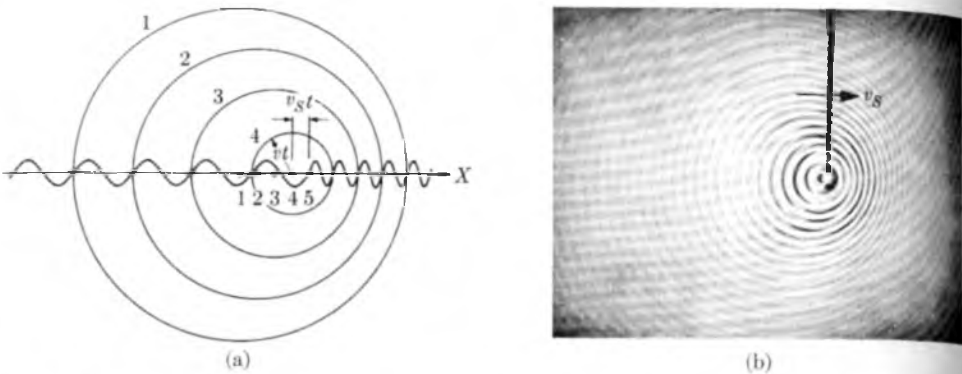


Fig. 10-26. Doppler effect produced by a moving source. The photograph illustrates the Doppler effect on a liquid surface.

waves. The opposite happens if the observer is receding from the source and therefore moving away from the waves.

To obtain the relation between the frequency ν of the waves produced by the source and the frequency ν' recorded by the observer, we use the following reasoning. For simplicity assume that the source and the observer are moving along the same line. Suppose that at time $t=0$ when the distance AB (Fig. 10-27) between the source and the observer is l, the source emits a wave that reaches the observer at a later time t. In that time the observer has moved the distance $v_0 t$, and the total distance traversed by the wave in the time t has been $l+v_0 t$. If v is the velocity of propagation of the wave, this distance is also vt. Then

$$vt = l + v_0 t \qquad \text{or} \qquad t = \frac{l}{v - v_0}.$$

At time $t=\tau$ the source is at A', and the wave emitted at that instant will reach the observer at point B' at a time t', measured from the same time origin as before. The total distance traveled by the wave from the time it is emitted at A' until it is received by the observer is $(l - v_S\tau) + v_0 t'$. The actual travel time of the wave has been $t' - \tau$ and the distance traveled is $v(t'-\tau)$. Therefore

$$v(t'-\tau) = l - v_S\tau + v_0 t' \qquad \text{or} \qquad t' = \frac{l + (v - v_S)\tau}{v - v_0}.$$

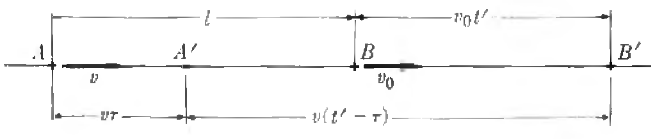


Figure 10-27

The time interval between the two waves emitted by the source at A and at A' is reckoned by the observer as

$$\tau' = t' - t = \frac{v - v_S}{v - v_0}\tau.$$

Now if ν is the frequency of the source, the number of waves emitted by the source in time τ is $\nu\tau$. Since the number of waves received by the observer in the time τ' must equal the number emitted by the source, the frequency observed is $\nu' = \nu\tau/\tau'$ or

$$\nu' = \left(\frac{v - v_0}{v - v_S}\right)\nu. \tag{10.57}$$

This equation gives the relation between the frequency ν of the source and the frequency ν' measured by the observer when both are moving along the same line.

When both v_0 and v_S are very small compared with v, expression (10.57) can be simplified. First we write

$$\nu' = \left(\frac{1 - v_0/v}{1 - v_S/v}\right)\nu = \left(1 - \frac{v_0}{v}\right)\left(1 - \frac{v_S}{v}\right)^{-1}\nu.$$

Remembering the binomial expansion, Eq. (M.28), we may write $(1 - v_S/v)^{-1} \approx 1 + v_S/v$, and

$$\nu' = \left(1 - \frac{v_0}{v}\right)\left(1 + \frac{v_S}{v}\right)\nu = \left(1 - \frac{v_0}{v} + \frac{v_S}{v} - \frac{v_0 v_S}{v^2}\right)\nu.$$

Then if the term $v_0 v_S/v^2$ is neglected, the frequency measured by the observer is

$$\nu' = \left(1 - \frac{v_0 - v_S}{v}\right)\nu = \left(1 - \frac{v_{0S}}{v}\right)\nu \tag{10.58}$$

where $v_{0S} = v_0 - v_S$ is the velocity of the observer relative to the source. Because $\omega = 2\pi\nu$, the angular frequency detected by the observer may be written as

$$\omega' = \left(1 - \frac{v_{0S}}{v}\right)\omega. \tag{10.59}$$

If v_{0S} is positive, the observer is receding from the source; and the frequency observed is smaller; but if v_{0S} is negative, the observer and the source are approaching each other and the frequency observed is larger.

When v_{0S} is not along the direction of propagation but makes an angle with it, Eq. (10.59) must be replaced by

$$\omega' = \left(1 - \frac{v_{0S}\cos\theta}{v}\right)\omega, \tag{10.60}$$

since $v_{0S}\cos\theta$ is the component of the velocity of the observer relative to the source along the direction of propagation.

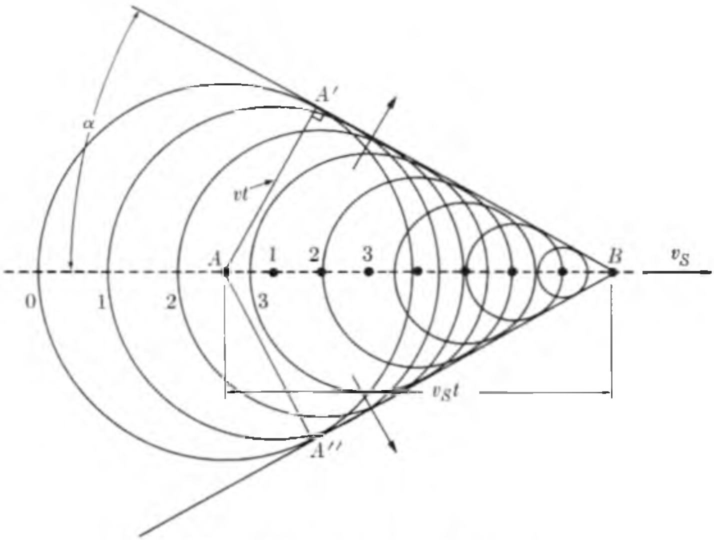
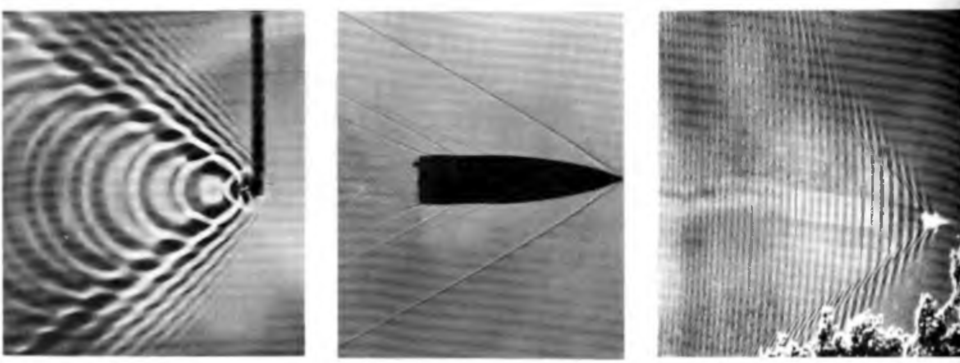


Fig. 10-28. Mach or shock wave.

A special situation occurs when the source moves with a velocity larger than v_S, the velocity of propagation of sound in the medium. Then in a given time the source advances more than the wave front; for example if in time t the source moves from A to B (Fig. 10-28), the wave emitted at A has traveled only from A to A' and A''. The surface tangent to all successive waves is a cone, whose axis is the line of motion of the source and whose aperture α is given by

$$\sin \alpha = \frac{v}{v_S}. \tag{10.61}$$



(a) (b) (c)

Fig. 10-29. Different examples of Mach (shock) waves produced by (a) a moving, vibrating reed touching a water surface, (b) a bullet in air, and (c) a speeding boat.

The resultant wave motion is then a *conical* wave that propagates as indicated by the arrows in Fig. 10-28. This wave is sometimes called a *Mach wave* or a *shock wave*, and is the sudden and violent sound heard when a supersonic plane passes nearby. These waves are also observed in the wakes of boats moving faster than the speed of surface waves on the water (Fig. 10-29).

10.12 Waves in Two and Three Dimensions

Although $\xi = f(x - vt)$ represents a wave motion propagating in the direction of the positive X-axis, the wave need not necessarily be interpreted as concentrated *on* the X-axis. If the physical disturbance described by ξ is extended over all space, then at a given time t, the function $\xi = f(x - vt)$ takes the same value at all points having the same x. But $x = \text{const}$ represents a plane perpendicular to the X-axis (Fig. 10-30). Therefore in three dimensions $\xi = f(x - vt)$ describes a *plane wave* propagating parallel to the X-axis. If ξ is a displacement (or a vector field), the wave is longitudinal when ξ is parallel to the direction of propagation or X-axis as indicated by the arrow labeled L. When ξ is perpendicular to the direction of propagation (i.e., parallel to the YZ-plane), the wave is transverse. In this case ξ can also be expressed as the superposition of two displacements along mutually perpendicular directions as indicated by the arrows labeled T and T'.

Note that what is relevant in a plane wave is the direction of propagation, indicated by a unit vector $\boldsymbol{u}$ perpendicular to the plane of the wave (that is, its *wave front*), and that the orientation of the coordinate axes is a more or less arbitrary matter. It is therefore convenient to express the plane wave $\xi = f(x - vt)$ in a form that is independent of the orientation of the wave relative to the axes. In Fig. 10-30 the unit vector $\boldsymbol{u}$

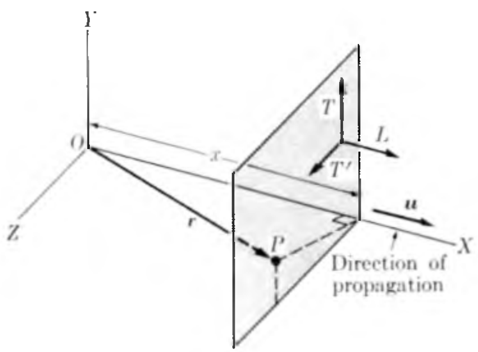


Fig. 10-30. Plane wave propagating along the X-axis.

is parallel to the X-axis. If $\boldsymbol{r}$ is the position vector of *any* point P on the wave front, $x = \boldsymbol{u} \cdot \boldsymbol{r}$ and we may write

$$\xi = f(\boldsymbol{u} \cdot \boldsymbol{r} - vt). \tag{10.62}$$

If $\boldsymbol{u}$ is pointing in an arbitrary direction (Fig. 10-31), the quantity $\boldsymbol{u} \cdot \boldsymbol{r}$ is still a distance measured from an origin O along the direction of propagation. Therefore Eq. (10.62) represents a plane wave propagating in the direction of $\boldsymbol{u}$. In the case of a harmonic or sinusoidal plane wave propagating in the direction of $\boldsymbol{u}$ we write

$$\xi = \xi_0 \sin k(\boldsymbol{u} \cdot \boldsymbol{r} - vt).$$

It is convenient to define a vector $\boldsymbol{k} = k\boldsymbol{u}$. This vector has a length $k = 2\pi/\lambda = \omega/v$, and points in the direction of propagation. This vector is usually called the *propagation vector* (or, sometimes, the *wave vector*). Then since $\omega = kv$, a plane harmonic wave is expressed as

$$\xi = \xi_0 \sin(\boldsymbol{k} \cdot \boldsymbol{r} - \omega t) = \xi_0 \sin(k_x x + k_y y + k_z z - \omega t) \tag{10.63}$$

where k_x, k_y, and k_z are the components of k satisfying

$$k_x^2 + k_y^2 + k_z^2 = k^2 = \omega^2/v^2. \tag{10.64}$$

When the propagation is in three-dimensional space, the wave equation (10.15) must be modified accordingly. It then becomes

$$\frac{\partial^2 \xi}{\partial t^2} = v^2 \left(\frac{\partial^2 \xi}{\partial x^2} + \frac{\partial^2 \xi}{\partial y^2} + \frac{\partial^2 \xi}{\partial z^2} \right), \tag{10.65}$$

a result that was to be expected from symmetry conditions alone. It may be verified by direct substitution that expression (10.63) for a harmonic plane wave satisfies the general wave equation (10.65). This check is left to the student. [*Hint:* It is necessary to use Eq. (10.64).]

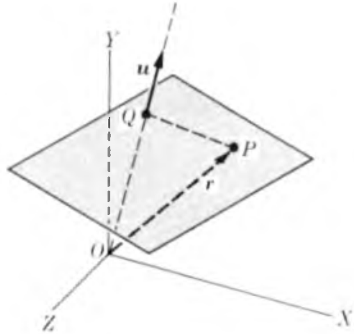


Fig. 10-31. Plane wave propagating in an arbitrary direction.

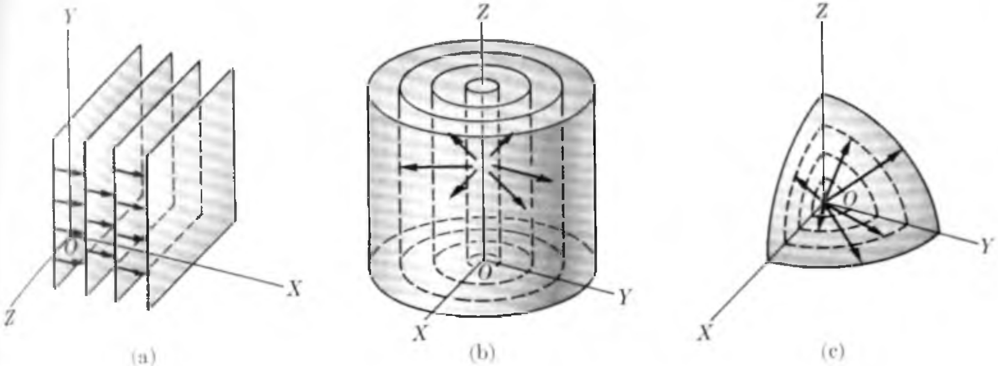


Fig. 10-32. (a) Plane, (b) cylindrical, and (c) spherical waves.

Although they contain the three coordinates x, y, z, the plane waves (10.62) or (10.63) are really one-dimensional problems since the propagation is along one particular direction and the physical situation is the same in all planes perpendicular to the direction of propagation (Fig. 10-32a). In nature, however, there are waves that propagate in several directions. The two most interesting cases are *cylindrical* and *spherical* waves. It can be proved that these more general waves are also solutions of the three-dimensional wave equation (10.65). In the case of cylindrical waves the wave fronts are surfaces parallel to a given line, say the Z-axis, and thus perpendicular to the XY-plane (Fig. 10-32b). The disturbance propagates in all directions perpendicular to the Z-axis. This type of wave is produced, for example, if a series of sources uniformly distributed along the Z-axis all oscillate in phase.

If a disturbance originates at a certain point and propagates with the same velocity in all directions (i.e., the medium is *isotropic*; from *isos*, the same, plus *tropos*, direction), spherical waves result. The wave fronts are spheres concentric with the point at which the disturbance originated (Fig. 10-32c). Such waves are produced, for example, when there is a sudden change of pressure at a point in a gas.

Sometimes the velocity of propagation is not the same in all directions, in which case the medium is called *anisotropic*. For example a gas in which there is a temperature gradient, a solid under certain conditions of strain, or a large crystal may have different elastic properties in different directions; the result is a different velocity of propagation for each direction. In these media, the waves are not spherical.

Even if a wave is spherical, it may not have the same amplitude or intensity in all directions because the source of the disturbance may produce different effects in different directions. For example a musician blowing a horn produces a pressure wave at the open end of the instrument. However because of the shape of the tube at the end, a listener does not hear the sound with the same intensity in all directions although the sound propagates with the same velocity in all directions (Fig. 10-33).

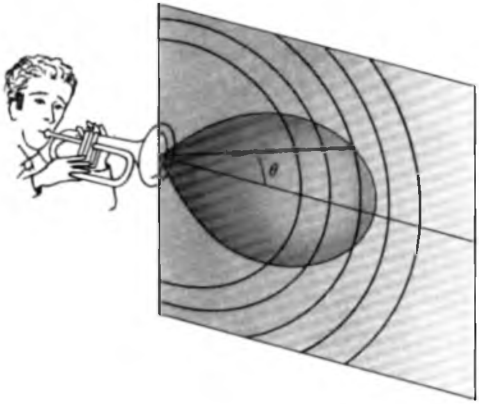
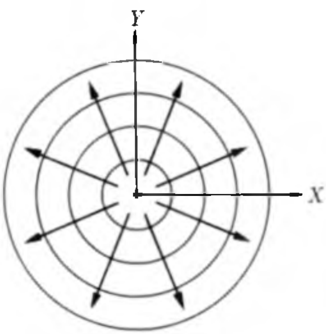


Fig. 10-33. Angular distribution of the intensity of sound produced by a horn.

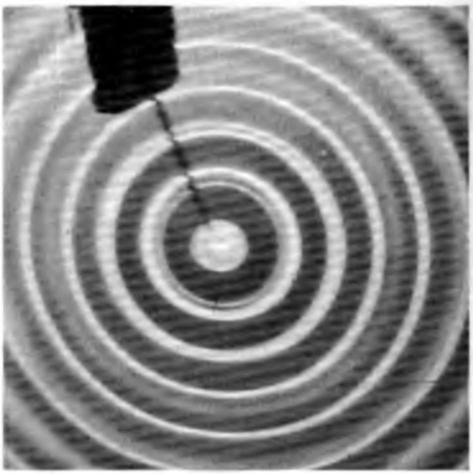
In some instances a wave propagates over a surface such as a membrane or the free surface of a liquid. If a disturbance is produced at a certain point of the surface, the disturbance propagates in all directions along the surface with the same velocity; the result is a series of *circular* waves (Fig. 10-34). This two-dimensional wave requires only two space coordinates to describe it. The equation for this wave is not Eq. (10.65) but

$$\frac{\partial^2 \xi}{\partial t^2} = v^2 \left(\frac{\partial^2 \xi}{\partial x^2} + \frac{\partial^2 \xi}{\partial y^2} \right) \tag{10.66}$$

since the z-coordinate is not required to describe the process.



(a)



(b)

Fig. 10-34. Circular waves on a liquid surface.

Example 10.15. Elastic waves on the surface of a stretched membrane.

▼ Consider a thin, stretched membrane assumed for simplicity to be rectangular (Fig. 10-35). The membrane is mounted in a frame that exerts a force *per unit length* $\mathscr{T}$ on the membrane. If the membrane is deformed at a particular point, and suffers a displacement in the perpendicular direction, this deformation propagates along the membrane and results in a two-dimensional surface wave.

To obtain the equation of motion for this situation, consider a small rectangular section of the membrane with sides dx and dy (Fig. 10-36). At a particular instant it suffers an upward displacement ξ. Because the surface is curved, the displacement ξ is a function of coordinates x and y, and the forces on the sides of the section are not directly opposite. To obtain the resultant vertical force on the section, use the same logic applied in Section 10.7 where transverse waves in a string were discussed. Following that reasoning, the sides parallel to the Y-axis are subject to forces $\mathscr{T}\,dy$ with a resultant vertical force of

$$(\mathscr{T}\,dy)\frac{\partial^2\xi}{\partial x^2}\,dx=\mathscr{T}\frac{\partial^2\xi}{\partial x^2}\,dx\,dy.$$

Similarly the sides parallel to the X-axis are subject to forces $\mathscr{T}\,dx$ with a vertical resultant force of

$$(\mathscr{T}\,dx)\frac{\partial^2\xi}{\partial y^2}\,dy=\mathscr{T}\frac{\partial^2\xi}{\partial y^2}\,dx\,dy.$$

Therefore the total vertical force is the sum of these two, or

$$F_z=\mathscr{T}\left(\frac{\partial^2\xi}{\partial x^2}+\frac{\partial^2\xi}{\partial y^2}\right)dx\,dy.$$

If σ is the mass per unit *area* of the membrane (or *surface mass density*), the mass of the section is $\sigma\,dx\,dy$; and since the vertical acceleration is $\partial^2\xi/\partial t^2$, the equation of motion of this section of the membrane is

$$(\sigma\,dx\,dy)\frac{\partial^2\xi}{\partial t^2}=\mathscr{T}\left(\frac{\partial^2\xi}{\partial x^2}+\frac{\partial^2\xi}{\partial y^2}\right)dx\,dy$$

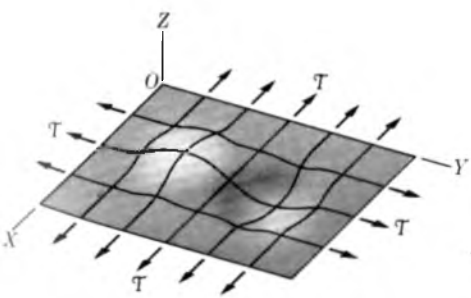


Fig. 10-35. Surface wave on a stretched membrane.

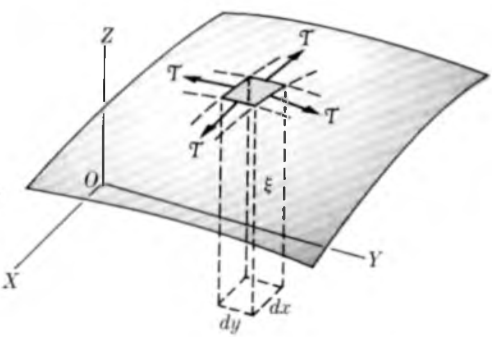


Fig. 10-36. Forces on a surface element of a stretched membrane.

or

$$\frac{\partial^2 \xi}{\partial t^2} = \frac{\mathscr{T}}{\sigma}\left(\frac{\partial^2 \xi}{\partial x^2} + \frac{\partial^2 \xi}{\partial y^2}\right).$$

Since this equation is similar to Eq. (10.66), we may conclude that the disturbance propagates along the membrane as a wave with a velocity $v = \sqrt{\mathscr{T}/\sigma}$. (The student should verify that the expression for v is dimensionally correct.) ▲

10.13 Spherical Waves in a Fluid

Consider a pressure wave in a homogeneous, isotropic fluid that has been generated by a point source such that the wave is spherical. At first we may be tempted to say that with r as the distance from the origin and p_0 the normal pressure, the pressure wave can be written in the form $p - p_0 = f(r - vt)$ since r now plays the role of x in a plane wave. However this expression is not correct and the situation requires more careful examination.

Observe that while a spherical wave propagates, the wave surface becomes larger and larger (increasing as r^2). A wave propagating within the solid angle Ω (Fig. 10-37) has a wave surface at a distance r from the source whose surface area is A; the wave surfaces at $2r, 3r, \ldots, nr$, are $4A, 9A, \ldots, n^2 A$. This situation suggests that the amplitude of the pressure wave must drop as the distance from the source increases since the pressure acts over a larger area. This result is confirmed experimentally and predicted by a more detailed theoretical analysis than that presented here. For example if the fluid is isotropic and the wave has the same amplitude in all directions at a distance r_0 from the source, it can be proved that the pressure wave is given by the expression

$$p - p_0 = \frac{1}{r} f(r - vt). \tag{10.67}$$

The geometrical factor $1/r$ that was not present in a plane wave accounts for a decrease in pressure with the distance from the source. When the amplitude (or intensity) is

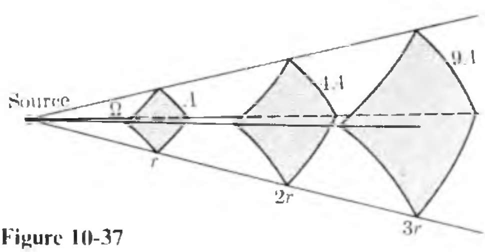


Figure 10-37

different in each direction, a more complicated expression may result. Equation (10.67) represents an *outgoing* spherical wave. An *incoming* spherical wave may be expressed by

$$p-p_0=\frac{1}{r}f(r+vt).$$

The velocity of propagation is given by the same expression obtained for plane waves, Eq. (10.32). That is,

$$v=\sqrt{\frac{\kappa}{\rho_0}}. \tag{10.68}$$

A particularly interesting case is that of a spherical harmonic pressure wave expressed by

$$p=p_0+\frac{\mathscr{P}_0}{r}\sin(kr-\omega t). \tag{10.69}$$

The amplitude of the pressure wave at a distance r from the source is $\mathscr{P}_0/r$ and decreases with the distance from the source. The displacement corresponding to this pressure wave is given by a more complicated expression; but at large distances from the source, the displacement may be expressed to a good approximation by

$$\xi=\frac{\xi_0}{r}\cos(kr-\omega t) \tag{10.70}$$

where

$$\xi_0=\frac{\mathscr{P}_0}{v\rho_0\omega},$$

a relation identical to that for plane waves (Example 10.8). Note that the amplitude of the displacement wave also decreases with the distance from the source as $1/r$.

The energy per unit volume at large distances is given, according to Eq. (10.50), by

$$E=\frac{1}{2}\frac{\rho_0\omega^2\xi_0^2}{r^2}=\frac{\mathscr{P}_0^2}{2v^2\rho_0 r^2},$$

and decreases as $1/r^2$ because the amplitude is now ξ_0/r rather than ξ_0. If we use Eq. (10.51) with $A=4\pi r^2$, the energy flowing per unit time through a spherical surface of radius r is

$$\left(\frac{\partial W}{\partial t}\right)_{\text{ave}}=v(4\pi r^2)\left(\frac{1}{2}\frac{\rho_0\omega^2\xi_0^2}{r^2}\right)=2\pi v\rho_0\omega^2\xi_0^2=\frac{2\pi\mathscr{P}_0^2}{\rho_0 v}. \tag{10.71}$$

Note that the factor r^2 has canceled from the expression above; the result is a value independent of the radius. This result should have been expected since the conservation of energy requires that on the average the same amount of energy flows per unit time through any spherical surface concentric with the source, irrespective of the radius. The appearance of the factor $1/r$ in Eqs. (10.69) and (10.70) now seems reasonable.

The intensity of the spherical wave at a distance r from the source (or the average energy crossing the unit area per unit time) is, according to Eq. (10.52),

$$I = v\mathrm{E} = \frac{\mathscr{P}_0^2}{2v\rho_0 r^2} = \frac{I_0}{r^2} \tag{10.72}$$

where

$$I_0 = \frac{\mathscr{P}_0^2}{2\rho_0 v}, \tag{10.73}$$

a result identical to that of Example 10.13. We conclude then that

> *in a spherical wave both the energy density and the intensity decrease as the inverse of the square of the distance from the source,*

a result of great application in both acoustics and optics. Again this result is consistent with the conservation of energy since if the energy flowing through each spherical surface has to be the same, and if the area of the sphere changes as r^2, the energy flowing through the unit area in unit time must change as $1/r^2$.

The spherical waves just discussed apply only to the case of perfect fluids that cannot sustain a shear stress. However in an elastic solid, two kinds of waves are possible: *irrotational* and *solenoidal* waves. In the case of plane waves the two kinds correspond essentially to the longitudinal and the transverse waves discussed in Sections 10.5 through 10.7. The respective velocities of propagation of irrotational and solenoidal waves are

$$v_l = \sqrt{\frac{\kappa + \frac{4}{3}G}{\rho}}, \qquad v_t = \sqrt{\frac{G}{\rho}} \tag{10.74}$$

where κ is the bulk modulus and G the shear modulus of the solid.

Note that if $G=0$, there are only longitudinal waves with a velocity equal to result (10.68). On the other hand, no stable medium can have $v_l=0$ and sustain only transverse waves because that situation would require $\kappa = -\frac{4}{3}G =$ a negative number. Such a value for κ would mean that an increase in pressure would result in an increase in volume, a situation contrary to both experience and intuition.

Problems

10.1 By rocking a boat, a man produces surface waves on a quiet lake. He observes that the boat performs 12 oscillations in 20 seconds, each oscillation producing a wave crest. It takes 6 s for a given crest to reach the shore 12 m away. Calculate the wavelength of the surface waves.

10.2 The equation of a certain wave is $\xi = 0.10 \sin 2\pi(2x - 100t)$ m, where x is in meters and t is in seconds. Determine (a) the amplitude,

(b) the wavelength, (c) the frequency, and (d) the velocity of propagation of the wave. (e) Sketch the wave and show the amplitude and the wavelength.

10.3 Given the wave

$$\xi = 0.02 \sin 2\pi(0.1x - 5t) \text{ m}$$

where x is in meters and t in seconds, determine (a) the wavelength, (b) the frequency, (c) the period, (d) the velocity of propagation, (e) the amplitude, and (f) the direction of propagation. (g) Give the expression for a wave that is identical but propagates in the opposite direction.

10.4 Given the wave

$$\xi = 0.02 \sin 2\pi(0.5x - 10t) \text{ m}$$

where t is in seconds and x is in meters, plot ξ over several wavelengths for (a) $t=0$ and (b) $t=\frac{1}{40}$ s. Repeat the problem for

$$\xi = 0.02 \sin 2\pi(0.5x + 10t) \text{ m},$$

for (c) $t=0$ and (d) $t=\frac{1}{40}$ s. (e) Compare results.

10.5 A harmonic wave,

$$\xi = A \sin 2\pi\left(\frac{x}{\lambda} - \frac{t}{P}\right),$$

propagates to the right. On the same set of axes plot the disturbance at time 0, $\frac{1}{4}P$, $\frac{1}{2}P$, $\frac{3}{4}P$, and P, all at $x=0$. Label each curve.

10.6 Plot the displacement ξ of the previous problem, all at $t=0$, for 13 positions $x=0$, $\lambda/12$, $\lambda/6$, $\lambda/4$, $5\lambda/12, \ldots \lambda$ on the same set of axes. Label each curve.

10.7 Assuming that the wave in Problem 10.5 corresponds to an elastic transverse wave, (a) plot the velocity $\partial\xi/\partial t$ and (b) the acceleration $\partial^2\xi/\partial t^2$ at $t=0$, $\frac{1}{4}P$, $\frac{1}{2}P$, $\frac{3}{4}P$ and P, all at $x=\lambda/4$.

10.8 Given the equation for a wave in a string,

$$\xi = 0.03 \sin(3x - 2t) \text{ m},$$

where ξ and x are in meters and t is in seconds, answer the following questions. (a) At $t=0$, what is the displacement at $x=0$, 0.1 m, 0.2 m and 0.3 m? (b) At $x=0.1$ m, what is the displacement at $t=0$, 0.1 s, and 0.2 s? (c) What is the equation for the velocity of oscillation of the particles of the string? (d) What is the maximum velocity of oscillation? (e) What is the velocity of propagation of the wave?

10.9 A certain wave is excited by a source whose motion can be represented by

$$\xi = \frac{8}{\pi^2} A\left[\sin \omega t - \frac{1}{3^2}\sin 3\omega t + \frac{1}{5^2}\sin 5\omega t - \cdots\right].$$

(a) Construct the approximate wave form by adding the first three terms graphically. The infinite series for the wave form leads to a shape called a "saw-tooth" curve. (b) Express a traveling wave having the same shape and propagating to the right with velocity v, independent of the frequency. (Note that $1+(\frac{1}{3})^2+(\frac{1}{5})^2+\cdots=\pi^2/8$ and that when $\sin \omega t=1$, $\sin 3\omega t=-1$.)

10.10 Repeat Problem 10.9 for a source whose motion is of the form

$$\xi = \frac{4}{\pi} A(\sin \omega t + \tfrac{1}{3}\sin 3\omega t + \tfrac{1}{5}\sin 5\omega t + \cdots).$$

(Note that $1-\frac{1}{3}+\frac{1}{5}-\cdots=\pi/4$.)

10.11 Consider longitudinal waves along a rod (Section 10.5) and assume that the deformation at each point is

$$\xi = \xi_0 \sin 2\pi\left(\frac{x}{\lambda} - \frac{t}{P}\right).$$

(a) Using relation (10.20), obtain the expression for the force on each section. (b) Show that the ξ and F waves have a phase difference of one-quarter wavelength. (c) Plot ξ and F against x at a given time for a distance of several wavelengths.

10.12 A spring having a normal length of 1 m and a mass of 0.2 kg is elongated 4×10^{-2} m when it is stretched by a force of 10 N. Find the velocity of propagation of longitudinal waves along the spring.

10.13 A steel spring has a normal length of 4 m and a mass of 0.2 kg. When the spring is supported vertically and a 0.1 kg body is attached to the lower end, the spring stretches 5×10^{-2} m. Find the velocity of longitudinal waves in the spring.

10.14 Compute the velocity of propagation of sound in (a) hydrogen, (b) nitrogen, and (c) oxygen at 0°C. Compare with the experimental results. For the three gases, assume $\gamma = 1.40$.

10.15 Find the change of sound velocity in air per unit change in temperature at 300 K (27°C).

10.16 From the value given in Example 10.7 for the coefficient $\alpha = \sqrt{\gamma R/M}$ for air, obtain the effective molecular mass of air and compare with the result obtained by other means. Assume that for air $\gamma = 1.40$.

10.17 How is the velocity of propagation of a transverse wave along a string modified if the tension is (a) doubled or (b) halved? How must the tension be changed to (c) double or (d) halve the velocity of propagation?

10.18 A steel wire having a diameter of 2×10^{-4} m is subject to a tension of 200 N. Determine the velocity of propagation of transverse waves along the wire.

10.19 A string of length 2 m and mass 4×10^{-3} kg is held horizontal, with one end fixed and with a mass of 2 kg supported at the other end. Find the velocity of transverse waves in the string.

10.20 One end of a horizontal string is attached to a prong of an electrically driven tuning fork whose frequency of vibration is 240 Hz. The other end passes over a pulley and supports a weight of 3 kg. The mass per unit length of the string is 2×10^{-2} kg m^{-1}. (a) What is the speed of a transverse wave in the string? (b) What is the wavelength?

10.21 A rubber tube is fastened at one end to a fixed support. The other end passes over a pulley at 5 m from the fixed end and carries a load of 2 kg. The mass of the tube between the fixed end and the pulley is 0.6 kg. (a) Find the velocity of propagation of transverse waves along the tube. (b) Suppose that a harmonic wave of amplitude 10^{-3} m and wavelength 0.3 m propagates along the tube; find the maximum transverse velocity of any point of the tube. (c) Write the equation of the wave.

10.22 A vibrating source at the end of a stretched string has a displacement given by the equation $\xi = 0.1 \sin 6t$ m, where ξ is in meters and t is in seconds. The tension in the string is 4 N and the mass per unit length is 10^{-2} kg m^{-1}. (a) What is the wave velocity in the string? (b) What is the frequency of the wave? (c) What is the wavelength? (d) What is the equation of the displacement at a point 1 m from the source? at 3 m? (e) Make a graph of ξ versus t at $x = 3$ m. (f) What is the amplitude of motion? (g) Make a graph of ξ versus x at $t = \pi/12$ s.

10.23 A steel wire having a length of 2 m and a radius of 5×10^{-4} m hangs from the ceiling. (a) If a body having a mass of 100 kg is hung from the free end, find the elongation of the wire. (b) Also find the displacement and the downward pull at a point at the middle of the wire. (c) Determine the velocity of longitudinal and transverse waves along the wire when the mass is attached.

10.24 Compare the relative importance of the two terms in the velocity of surface waves in deep water [Eq. (10.43)] for the following wavelengths: (a) 10^{-3} m, (b) 10^{-2} m, (c) 1 m. (d) At what wavelength are the two terms equal? For water the surface tension is about 7×10^{-2} N m^{-1}.

10.25 Consider a canal of rectangular cross section having a depth of 4 m. Determine the velocity of propagation of waves having a wavelength of (a) 10^{-2} m, (b) 1 m, (c) 10 m, (d) 100 m. In each case use the formula that corresponds best to the order of magnitude of the quantities involved. The water in the canal has a surface tension of 7×10^{-2} N m^{-1}.

10.26 In Section 10.9 we obtained the energy flow of a longitudinal wave in a solid rod. Repeat the calculation for transverse waves in a string; show that the average power is $v(\frac{1}{2}m\omega^2\xi_0^2)$. Note that the quantity inside the parentheses now corresponds to energy per unit length. (*Hint:* Compute the rate of work done by the force perpendicular to the string: that is, $F \sin\alpha \approx F(\partial\xi/\partial x)$ of Fig. 10-14.)

10.27 The faintest sound that can be heard has a pressure amplitude of about 2×10^{-5} Pa, and the loudest that can be heard without pain has a pressure amplitude of about 28 Pa. In each case, determine (a) the intensity of the sound both in W m^{-2} and in db, and (b) the amplitude of the oscillations if the frequency is 500 Hz. Assume an air density of 1.29 kg m^{-3} and a velocity of sound of 345 m s^{-1}.

10.28 Two sound waves have intensity levels differing by (i) 10 db, and (ii) 20 db. Find the ratio of (a) their intensities and of (b) their pressure amplitudes.

10.29 (a) How is the intensity of a sound wave changed when the pressure amplitude is doubled? (b) How must the pressure amplitude change to increase the intensity by a factor of 10?

10.30 Express in db the difference in intensity levels of two sound waves if (a) the intensity of one wave is twice the intensity of the other, and (b) the pressure amplitude of one is twice that of the other.

10.31 Two sound waves, one in air and one in water, have the same intensity. (a) What is the ratio of the pressure amplitude of the wave in water to that of the wave in air? (b) What would be the ratio of their intensities if the pressure amplitudes were the same?

10.32 The pitch of the whistle of a locomotive is 500 Hz. Determine the frequency of the sound heard by a person standing at the station if the train is moving with a velocity of 72 km hr^{-1} (a) toward and (b) away from the station.

10.33 A sound source has a frequency of 10^3 Hz and moves at 30 m s^{-1} relative to the air. Assuming that the velocity of sound relative to still air is 340 m s^{-1}, find the effective wavelength and the frequency perceived by an observer who is at rest relative to the air and who sees the source (a) receding, and (b) approaching.

10.34 (a) Repeat Problem 10.33 if the source is at rest relative to the air but the observer moves at 30 m s^{-1}. (b) From your results do you conclude that it is immaterial which is in motion, the source or the observer?

CHALLENGING PROBLEMS

10.35 A train moving with speed v sounds its whistle, which has a frequency f. If the speed of sound in air is c_s, determine the wavelength of the sound heard by an observer O directly in front of the train. (AP-B; 1971)

10.36 The horn on a very fast racing car is blown as the car moves past an observer beside the road. After the car passes, the pitch of the sound heard by the observer is an octave lower than the pitch of the sound heard by the observer as the car approached; i.e., the frequency of the sound goes down by a factor of 2. If the speed of sound is 340 m per second, how fast is the car traveling? (AP-B; 1972)

10.37 Obtain the velocity of shear waves in steel. Compare with the result for longitudinal waves given in Example 10.4.

10.38 (a) Show that the energy wave discussed in Section 10.9 can be written in the form

$$\frac{\partial W}{\partial t} = v\{\rho\omega^2\xi_0^2[\tfrac{1}{2}+\tfrac{1}{2}\cos 2(kx-\omega t)]\}.$$

(b) Obtain from it its average value. (c) Show that the frequency of the energy wave is twice, and the wavelength one-half, that of the displacement wave. (d) Plot $\partial W/\partial t$ as a function of x at a given time.

10.39 A pendulum consists of a steel wire 2.00 m long carrying a mass of 20 kg. If the pendulum is released from a position that makes an angle of 60° with the vertical, find the difference in length of the wire when the bob is at the initial position and when the bob passes through the lowest point.

10.40 A steel rod is forced to transmit longitudinal waves by means of an oscillator coupled to one end. The rod has a diameter of 4×10^{-3} m. The amplitude of the oscillations is 10^{-4} m and the frequency is 10 oscillations per second. Find (a) the equation of the waves along the rod, (b) the energy per unit volume of the rod, (c) the average energy flow per unit time across any section of the rod, and (d) the power required to drive the oscillator.

10.41 A rope of length L and mass M hangs freely from the ceiling. (a) Show that the velocity of a transverse wave as a function of position along the rope is $v=\sqrt{gx}$ where x is the distance from the lower end. (b) Show that a transverse pulse will traverse the rope in a time $2\sqrt{L/g}$. Note that the results are independent of the mass of the rope.

10.42 (a) Show that a transverse elastic wave propagating along the X-axis and corresponding to a displacement ξ having as components

$$\xi_y = \xi_0 \sin (kx - \omega t)$$

and

$$\xi_z = \xi_0 \cos (kx - \omega t)$$

is circularly polarized. (b) Determine the sense of rotation of ξ as seen by an observer on the X-axis. (c) Write the expressions for ξ_y and ξ_z for a wave having an opposite polarization.

10.43 Referring to pressure waves in a gas column (Section 10.6), assume that the pressure changes in the form

$$p - p_0 = \mathscr{P}_0 \sin 2\pi \left(\frac{x}{\lambda} - \frac{t}{P} \right).$$

(a) Using Eqs. (10.25) and (10.29), obtain the expressions for the density and displacement waves in the gas. (b) Show that the pressure and density waves are in phase but that the displacement wave has a phase difference of one-quarter wavelength. (c) Plot the three waves against x for a given time over a length of several wavelengths.

10.44 A plane harmonic sound wave in air at 293 K and standard pressure has a frequency of 500 Hz and a displacement amplitude of 10^{-8} m. (a) Write the expression describing the displacement wave. (b) Plot the displacement wave form at $t=0$ s over a few wavelengths. (c) Write the expression describing the pressure wave. (d) Plot the pressure wave form at $t=0$ s over a few wavelengths and compare with the plot in (b). (e) Express in db the intensity level of this wave.

10.45 Two harmonic waves of the same frequency and amplitude propagate with the same velocity in opposite directions. (a) Determine the resultant wave motion. (b) Assuming that the resultant wave corresponds to a transverse wave in a string, plot the displacement of the points of the string at different times.

10.46 Two waves of equal amplitude, velocity, and frequency, but with a phase difference of $\pi/4$, run in the same direction on a string. Add the two and show that the result is a running wave of the same velocity and frequency.

10.47 Two waves of the same amplitude and velocity but of different frequencies, equal to 1000 and 1010 Hz, travel in the same direction at 10 m s^{-1}. (a) Write equations for the separate waves and for their sum. (b) Make a sketch of the resultant wave form.

10.48 Two waves, plane polarized in perpendicular planes, travel in the OX-direction at the same velocity. Find the resultant wave motion if (a) $A_1 = 2A_2$ and the phases are the same, (b) $A_1 = 2A_2$ and the phases differ by $\pi/2$, and (c) $A_1 = A_2$ and the phases differ by $\pi/2$.

10.49 In the discussion of longitudinal waves in a rod (Section 10.5), we neglected the lateral strain that accompanies the longitudinal strain. When this effect is taken into account, it can be shown that the phase velocity of harmonic longitudinal waves of wavelength λ propagating along a cylindrical rod of radius R is

$$v_p = \sqrt{\frac{Y}{\rho}}(1 - \pi^2\sigma^2R^2/\lambda^2)$$

where σ is a coefficient called *Poisson's ratio* (see Problem 10.55). (a) Find the group velocity of waves along the rod and express it in terms of v_p. (b) Obtain the limiting value of the group velocity for the case of R much smaller than λ. (c) Discuss the variation in v_p and v_g as a function of R/λ.

10.50 The phase velocity of a harmonic *flexural wave* in a solid rod is

$$v_p = \frac{v}{\sqrt{1 + \lambda^2/4\pi^2K^2}}$$

where $v = \sqrt{Y/\rho}$ is the phase velocity for longitudinal waves, λ the wavelength, and K the radius of gyration of the cross section of the rod about an axis through the center normal to the rod's longitudinal axis. (a) Find the group velocity for flexural waves and express it in terms of the phase velocity. (b) Specialize to the case of a rod of circular cross section. (c) Obtain the group velocity when λ is much larger than $2\pi K$. (*Note:* A *flexural wave* is a wave propagated along a loaded rod; that is, a rod subject to a transverse force (such as its own weight) uniformly distributed along its length.)

10.51 Equation (10.58) for the Doppler effect was derived on the assumption that the medium through which the waves propagate remains at rest. Show that if the medium has a velocity v_m along the line joining the source and the observer, the equation becomes

$$\nu' = \frac{\nu(v - v_0 + v_m)}{(v - v_s + v_m)}.$$

10.52 The *volume strain* of a body is defined by the relation $\epsilon_V = dV/V$ where dV is the change in volume resulting from the forces applied to the body of volume V. (a) Show that

$$\epsilon_V = -\frac{d\rho}{\rho}$$

where ρ is the density of the body. (*Hint:* Note that $\rho V = m = \text{const.}$) (b) Show also that the bulk modulus defined by Eq. (10.27) can be expressed in the alternative form

$$\kappa = -V\left(\frac{dp}{dV}\right)$$

where dV is the change in volume resulting from the change in pressure dp.

10.53 Using the values of the bulk modulus for iron and for lead (Table 10-1), compute the percentage change in density and volume of each substance for a change in pressure equal to 10^5 Pa.

10.54 The *linear strain* is defined by the relation

$$\epsilon_L = \frac{dL}{L}$$

where L is the distance between any two points in the body in the undeformed state and dL is their distance change resulting from the deformation. By considering a cube of side L, show that $\epsilon_V = 3\epsilon_L$.

10.55 When a wire is stretched in the direction of its length, the diameter D of the wire is decreased. The result is a *lateral strain* defined by

$$\epsilon_D = \frac{dD}{D}.$$

The *Poisson ratio* is defined by

$$\sigma = \frac{dD}{dL}.$$

Show that if a rectangular parallelepiped is subject to a normal stress S on each surface, the net linear strain of each side is

$$\epsilon_L = S(1 - 2\sigma)/Y.$$

(*Hint:* Note that the normal stress on each pair of surfaces of the parallelepiped results in opposite lateral strains on the other pairs of surfaces.)

10.56 (a) Using the results of Problems 10.54 and 10.55, show that

$$Y = 3\kappa(1 - 2\sigma).$$

(b) Solve this relation for σ; and using the values of Table 10-1, compute the Poisson ratio for some materials.

10.57 By a logic similar to that of Problem 10.56, it can be shown that

$$Y=2G(1+\sigma).$$

(a) By eliminating σ between this expression and that of Problem 10.56, show that

$$Y=\frac{3\kappa G}{(\kappa+\frac{1}{3}G)}.$$

(b) Using the values given in Table 10-1, for some of the materials listed, verify the extent of the validity of this theoretical expression for Y.

10.58 For a certain substance $G=1.24\times10^{10}$ Pa and $Y=3.20\times10^{10}$ Pa. (a) Calculate the value of the bulk modulus and Poisson's ratio for this substance. (b) Do the same for quartz, which has $Y=5.18\times10^{10}$ Pa and $G=2.88\times10^{10}$ Pa. (c) Discuss the physical implications of your results.

10.59 It can be shown that for a spring the constant K that was introduced in Example 10.6 is given by $\pi GR^4/2a^2$ where R is the wire radius and a the spring radius. (a) Find the value of K for a steel spring that has a radius of 10^{-2} m and is made of steel wire having a radius of 10^{-3} m. (b) If the unstretched length of the spring is 0.5 m, find its elongation when a force of 50 N is applied to it.

10.60 Assume a field ξ having a propagation equation

$$\frac{\partial^2\xi}{\partial t^2}=a\frac{\partial^4\xi}{\partial x^4}$$

where a is a certain constant. (a) Does it admit an expression of the form

$$\xi=\xi_0\sin k(x\pm vt)$$

as a solution? If so, what is the value of v? (b) Does it admit $\xi=f(x\pm vt)$ as a solution? (c) From the preceding results, do you conclude that this field propagates undistorted?

10.61 A rod of circular cross section with radius R is twisted as a result of torques applied around its axis. Show that if θ is the torsion angle at a point x on the abscissa, the torque there is

$$\tau=\tfrac{1}{2}AGR^2\left(\frac{\partial\theta}{\partial x}\right)$$

where $A=\pi R^2$ is the area of the cross section.

10.62 Using the result of the previous problem, show that the velocity of propagation of torsional waves along the rod is

$$v=\sqrt{G/\rho}.$$

(*Hint:* Consider a section of thickness dx and note that the net torque on it is $(\partial\tau/\partial x)\,dx$.)

10.63 It can be shown that a spherical isotropic wave satisfies the differential equation

$$\frac{\partial^2(r\xi)}{\partial t^2}=v^2\frac{\partial^2(r\xi)}{\partial r^2}.$$

(a) Verify that the solution of this equation is

$$\xi=\left(\frac{1}{r}\right)f(r\pm vt).$$

(b) Compare with the discussion in Section 10.13 for pressure waves in a fluid.

10.64 Show that when the amplitude is large, the equation for transverse waves in a string becomes

$$\frac{\partial^2\xi}{\partial t^2}=\frac{T}{m}\frac{\partial^2\xi}{\partial x^2}\left[1-\frac{3}{2}\left(\frac{\partial\xi}{\partial x}\right)^2\right].$$

Observe that this equation is not linear and reduces to Eq. (10.36) when $(\partial\xi/\partial x)^2$ is negligible. (*Hint:* Note that $\sin\alpha=\tan\alpha/\sqrt{1+\tan^2\alpha}=\tan\alpha-\frac{1}{2}\tan^3\alpha+\cdots$.)

CHAPTER ELEVEN

ELECTRO-MAGNETIC WAVES

11.1 Introduction

In Section 4.5 we suggested that the electromagnetic field may propagate in vacuum with a velocity

$$c = \frac{1}{\sqrt{\epsilon_0 \mu_0}} \approx 3 \times 10^8 \text{ m s}^{-1},$$

a velocity that corresponds to the velocity of light in vacuum. In Section 9.6 while dealing with the phenomenon of electromagnetic induction, we indicated the possibility of transmitting a signal from one place to another using a time-dependent electromagnetic field. Near the end of the nineteenth century, the German physicist Heinrich Hertz (1857–1894) proved beyond any doubt that the electromagnetic field does propagate in vacuum with a velocity equal to c.* The properties of these electromagnetic waves discovered by Hertz have been examined experimentally with great care. The large body of information that has accumulated about the properties of electromagnetic waves, such as their production, propagation, and absorption, has opened the door to the world of communications today. Before Hertz performed his experiments, the existence of electromagnetic waves had been predicted by Maxwell as a result of a careful analysis of the equations of the electromagnetic field. The development of our knowledge of electromagnetic waves is another example of the close relationship between theory and experiment in the evolution of physical ideas.

In this chapter we examine Maxwell's equations (which describe the time-dependent electromagnetic field) to see how to interpret the propagation of this field in the form of waves. In order to do that, we must show that the electric and magnetic fields satisfy a wave equation of the form of Eq. (10.15); that is, $\partial^2\xi/\partial x^2 = v^2(\partial^2\xi/\partial t^2)$. We then discuss a simple solution to these equations: plane electromagnetic waves. Finally some mechanisms for emitting, absorbing, and scattering electromagnetic radiation are explored.

11.2 Plane Electromagnetic Waves

We shall show that Maxwell's equations for the electromagnetic field admit as a special solution an electric field $\mathscr{E}$ and a magnetic field $\mathscr{B}$ perpendicular to each other. Assume an $\mathscr{E}$-field parallel to the Y-axis and orient a $\mathscr{B}$-field parallel to the

*Hertz's experiments are described in Section 16.6.

Z-axis. In this special case,

$$\mathscr{E}_x=0, \qquad \mathscr{E}_y=\mathscr{E}, \qquad \mathscr{E}_z=0,$$

and

$$\mathscr{B}_x=0, \qquad \mathscr{B}_y=0, \qquad \mathscr{B}_z=\mathscr{B}.$$

Assume also that the field is in vacuum; that is, that there are no free charges or currents, the implication being that $\rho=0$ and $j=0$ in Maxwell's equations.

Under these conditions Eqs. (8.34) become

(a) Gauss's law for the electric field, div $\mathscr{E}=0$,

$$\frac{\partial \mathscr{E}}{\partial y}=0. \tag{11.1}$$

(b) Gauss's law for the magnetic field, div $\mathscr{B}=0$,

$$\frac{\partial \mathscr{B}}{\partial z}=0. \tag{11.2}$$

(c) Faraday-Henry's law, curl $\mathscr{E}=-\partial\mathscr{B}/\partial t$,

$$\frac{\partial \mathscr{E}}{\partial z}=0, \tag{11.3}$$

$$\frac{\partial \mathscr{E}}{\partial x}=-\frac{\partial \mathscr{B}}{\partial t}. \tag{11.4}$$

(d) Ampère-Maxwell's law, curl $\mathscr{B}=\epsilon_0\mu_0(\partial\mathscr{E}/\partial t)$,

$$\frac{\partial \mathscr{B}}{\partial y}=0, \tag{11.5}$$

$$-\frac{\partial \mathscr{B}}{\partial x}=\epsilon_0\mu_0\frac{\partial \mathscr{E}}{\partial t}. \tag{11.6}$$

Equations (11.1), (11.2), (11.3), and (11.5) indicate that neither $\mathscr{E}$ nor $\mathscr{B}$ depends on y or z. Therefore the fields $\mathscr{E}$ and $\mathscr{B}$ depend only on x and t so that at each instant the fields have the same value at all points of planes perpendicular to the X-axis (Fig. 11-1). The two equations (11.4) and (11.6) can now be used to obtain the dependence of $\mathscr{E}$ and $\mathscr{B}$ on x and t. Taking the derivative of Eq. (11.4) with respect to x, we get

$$\frac{\partial^2 \mathscr{E}}{\partial x^2}=-\frac{\partial^2 \mathscr{B}}{\partial x\,\partial t}.$$

Similarly taking the derivative of Eq. (11.6) with respect to t yields

$$-\frac{\partial^2 \mathscr{B}}{\partial t\,\partial x}=\epsilon_0\mu_0\frac{\partial^2 \mathscr{E}}{\partial t^2}.$$

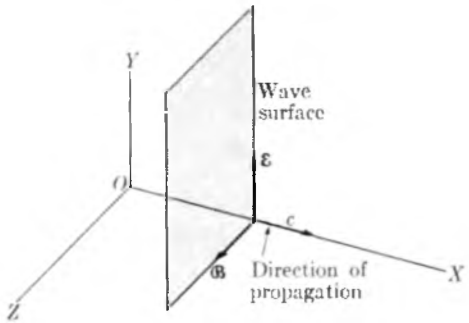


Fig. 11-1. Orientation of the electric and the magnetic fields relative to the direction of propagation of a plane electromagnetic wave.

Combining these two results gives

$$\frac{\partial^2 \mathscr{E}}{\partial t^2}=\frac{1}{\epsilon_0\mu_0}\frac{\partial^2 \mathscr{E}}{\partial x^2}. \tag{11.7}$$

This equation is the same as Eq. (10.15), and indicates that the electric field $\mathscr{E}$ propagates along the X-axis with a velocity

$$c=\frac{1}{\sqrt{\epsilon_0\mu_0}}, \tag{11.8}$$

and can be expressed as

$$\mathscr{E}=\mathscr{E}(x-ct). \tag{11.9}$$

By taking the time derivative of Eq. (11.4) and the space derivative of Eq. (11.6), we get

$$\frac{\partial^2 \mathscr{B}}{\partial t^2}=\frac{1}{\epsilon_0\mu_0}\frac{\partial^2 \mathscr{B}}{\partial x^2} \tag{11.10}$$

so that the magnetic field $\mathscr{B}$ also propagates along the X-axis with velocity c and can be expressed as

$$\mathscr{B}=\mathscr{B}(x-ct). \tag{11.11}$$

Consider the particular case of harmonic waves of frequency $\nu=\omega/2\pi$ and wavelength $\lambda=2\pi/k$. In such a case.

$$\mathscr{E}=\mathscr{E}_0 \sin k(x-ct)=\mathscr{E}_0 \sin (kx-\omega t)$$

and (11.12)

$$\mathscr{B}=\mathscr{B}_0 \sin k(x-ct)=\mathscr{B}_0 \sin (kx-\omega t).$$

In writing these equations we employed the relation $\omega=kc$, given by Eq. (10.6).

The amplitudes $\mathscr{E}_0$ and $\mathscr{B}_0$ are not independent since Eqs. (11.4) and (11.6) must be satisfied simultaneously. Now

$$\frac{\partial \mathscr{E}}{\partial x}=k\mathscr{E}_0 \cos k(x-ct) \qquad \text{and} \qquad \frac{\partial \mathscr{B}}{\partial t}=-kc\mathscr{B}_0 \cos k(x-ct).$$

Substituting these in Eq. (11.4) gives

$$\mathscr{E}_0=c\mathscr{B}_0 \qquad \text{or} \qquad \mathscr{B}_0=\frac{1}{c}\mathscr{E}_0. \tag{11.13}$$

The student may verify that the same result is obtained by using Eq. (11.6) instead of (11.4). Because of their special form, the instantaneous values given by Eq. (11.12) also have the relationship

$$\mathscr{E}=c\mathscr{B} \qquad \text{or} \qquad \mathscr{B}=\frac{1}{c}\mathscr{E}. \tag{11.14}$$

This relationship is not necessarily true for more general cases.

From Eq. (11.14) the $\mathscr{E}$ and $\mathscr{B}$ fields are in phase and reach their zero and maximum values at the same time. The electromagnetic wave described by Eq. (11.12) is represented in Fig. 11-2. The electric field oscillates in the XY-plane; and the magnetic field, in the XZ-plane. This situation corresponds to a *plane* or *linearly* polarized wave. The *plane of polarization* is defined as the plane in which the electric field oscillates, in this case the XY-plane. Thus an electromagnetic wave actually consists of two coupled waves: the electric wave and the magnetic wave.

Equation (11.12) is not the only plane wave solution of Eqs. (11.7) and (11.10). For example the electric field could have been oriented along the Z-axis, in which case the magnetic field would be along the $-Y$-axis. That is,

$$\mathscr{E}_z=\mathscr{E}_0 \sin(kx-\omega t), \qquad \mathscr{B}_y=-\mathscr{B}_0 \sin(kx-\omega t).$$

In addition, in both this wave and the wave of Eq. (11.12), it is possible to use cosines instead of sines, or even to add an arbitrary constant phase.

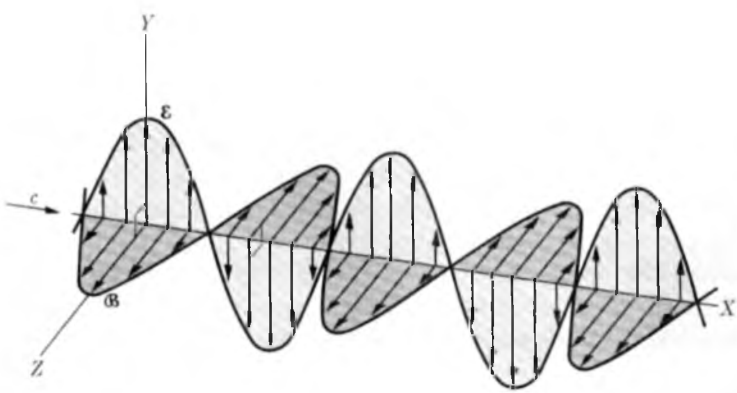


Fig. 11-2. Electric and magnetic fields in a harmonic plane electromagnetic wave.

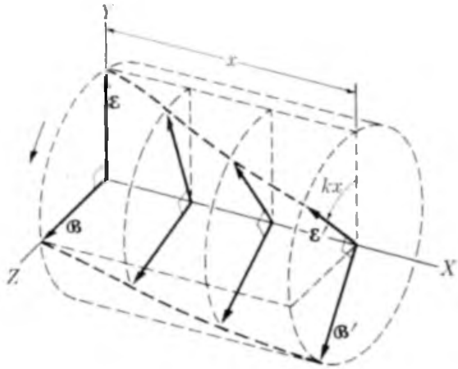


Fig. 11-3. Circularly polarized electromagnetic wave. The $\mathscr{E}$ and $\mathscr{B}$ fields rotate around the direction of propagation.

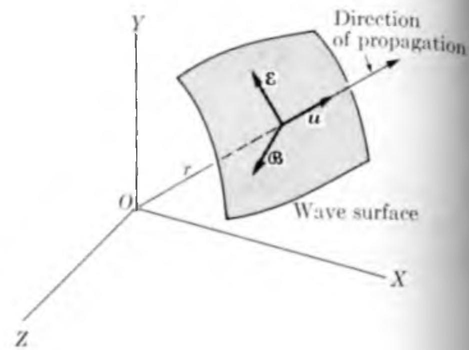


Fig. 11-4. Spherical electromagnetic wave at a large distance from the source.

Another plane-wave solution is one in which the electric and the magnetic fields remain constant in magnitude but rotate around the direction of propagation, and result in a *circularly polarized* wave (Fig. 11-3). This new solution is obtained by combining two linearly polarized solutions for each of the $\mathscr{E}$ and $\mathscr{B}$ fields, previously discussed in Chapter 12 of Volume I, with equal amplitudes for each field and with the appropriate phase difference. (This combination is possible because Maxwell equations are linear in the $\mathscr{E}$ and $\mathscr{B}$ fields.) The circular polarization may be right- or left-handed according to the sense of rotation of the fields. For example we could take the components of the electric and the magnetic fields along two mutually perpendicular axes as

$$\mathscr{E}_y = \mathscr{E}_0 \sin(kx - \omega t), \qquad \mathscr{E}_z = \pm \mathscr{E}_0 \cos(kx - \omega t),$$

and

$$\mathscr{B}_y = \mp \mathscr{B}_0 \cos(kx - \omega t), \qquad \mathscr{B}_z = \mathscr{B}_0 \sin(kx - \omega t),$$

corresponding to a phase difference of $\pm\pi/2$ between the components of each field, in accordance with Section 12.9 of Volume I, with the magnetic field $\mathscr{B}$ perpendicular to $\mathscr{E}$ at each instant. If the amplitudes of the two rectangular components of each field are different, *elliptical polarization* results. In addition, other plane-wave solutions of Maxwell's equations are possible that do not correspond to any particular state of polarization; the requirement is that the solutions satisfy the relations (11.1) through (11.6). However, we shall not discuss these other solutions here since for most applications the basic understanding of plane and circularly polarized waves is enough.

The choice of the relation between the $\mathscr{E}$ and $\mathscr{B}$ fields relative to the XYZ-axes was a matter of convenience; the plane-wave solutions of Maxwell's equations we obtained are completely general and illustrate a key feature of these waves:

> *plane electromagnetic waves are transverse with the $\mathscr{E}$ and $\mathscr{B}$ fields perpendicular both to each other and to the direction of propagation of the waves.*

This theoretical prediction of Maxwell's equations has been amply confirmed by experiment and results in several phenomena that will be considered in subsequent chapters. Besides plane-wave solutions to Maxwell's equations, there are also cylindrical and spherical electromagnetic waves. Practically speaking, at a large distance from the source a limited portion of a cylindrical or a spherical wave can be considered as plane. In fact all solutions to Maxwell's equations have the electric and the magnetic fields perpendicular to each other as well as to the direction of propagation (i.e., radial) as indicated in Fig. 11-4.

11.3 Energy and Momentum of an Electromagnetic Wave

From Eq. (2.40) the energy density associated with the electric field of an electromagnetic wave in vacuum is

$$\mathrm{E}_e = \tfrac{1}{2}\epsilon_0 \mathscr{E}^2.$$

Similarly when Eqs. (11.14), $\mathscr{B} = \mathscr{E}/c$, and (11.8), $c = 1/\sqrt{\epsilon_0\mu_0}$, are used, the magnetic energy density given by Eq. (9.13) is

$$\mathrm{E}_m = \frac{1}{2\mu_0}\mathscr{B}^2 = \frac{1}{2\mu_0 c^2}\mathscr{E}^2 = \tfrac{1}{2}\epsilon_0\mathscr{E}^2$$

so that $\mathrm{E}_e = \mathrm{E}_m$. That is, the electric energy density of an electromagnetic wave is equal to the magnetic energy density. The total energy density is

$$\mathrm{E} = \mathrm{E}_e + \mathrm{E}_m = \epsilon_0\mathscr{E}^2. \tag{11.15}$$

The intensity of the electromagnetic wave (that is, the energy passing through the unit area in the unit time) is, from Eq. (10.52),

$$I = \mathrm{E}c = c\epsilon_0\mathscr{E}^2. \tag{11.16}$$

The average intensity of the electromagnetic wave is $I_{\text{ave}} = c\epsilon_0(\mathscr{E}^2)_{\text{ave}}$. In the case of a harmonic electromagnetic wave,

$$(\mathscr{E}^2)_{\text{ave}} = \mathscr{E}_0^2[\sin^2 k(x-ct)]_{\text{ave}} = \tfrac{1}{2}\mathscr{E}_0^2$$

so that the average intensity is

$$I_{\text{ave}} = \tfrac{1}{2}c\epsilon_0\mathscr{E}_0^2. \tag{11.17}$$

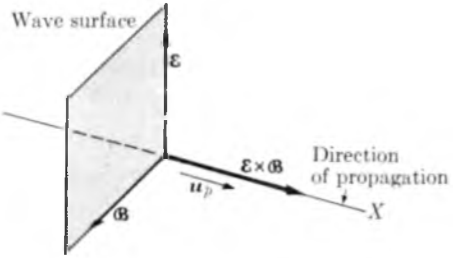


Fig. 11-5. Definition of the direction of energy flow in an electromagnetic wave.

Because it has special significance, let us compute the vector product $\mathscr{E} \times \mathscr{B}$ for a plane electromagnetic wave. The direction of $\mathscr{E} \times \mathscr{B}$ is perpendicular to the wave front and is therefore pointing in the direction of propagation of the wave (Fig. 11-5). The magnitude of the product is

$$|\mathscr{E} \times \mathscr{B}| = \mathscr{E}\mathscr{B} = \frac{1}{c}\mathscr{E}^2,$$

and so the vector $c\mathscr{E} \times \mathscr{B}$ has magnitude $\mathscr{E}^2$. Then the vector $c^2\epsilon_0\mathscr{E} \times \mathscr{B}$, called the *Poynting vector*, has a magnitude equal to I. Therefore the energy crossing a surface S per unit time is found by calculating the flux of the Poynting vector across the surface S. Thus,

$$\frac{dE}{dt} = \int_S c^2\epsilon_0(\mathscr{E} \times \mathscr{B}) \cdot \boldsymbol{u}_N \, dS. \tag{11.18}$$

We know from the special theory of relativity that energy and momentum are closely related, and that they form a four-vector (see Section 11.7 of Volume I). We may then expect that an electromagnetic wave carries a certain momentum in addition to its energy. Since electromagnetic radiation propagates with velocity c, the energy-momentum relation, $p = E/c$, given by the special theory of relativity [see Eq. (11.18) of Volume I], may be used to obtain the momentum P per unit volume associated with an electromagnetic wave. Thus

$$\text{P} = \frac{\text{E}}{c} = \frac{\epsilon_0\mathscr{E}^2}{c} = \epsilon_0|\mathscr{E} \times \mathscr{B}|. \tag{11.19}$$

(The student should verify that $\epsilon_0|\mathscr{E} \times \mathscr{B}|$ has the dimensions of $\text{m}^{-2}\ \text{kg s}^{-1}$, which corresponds to momentum per unit volume.) Since momentum is a vector quantity, P must have the same direction as the direction of propagation of the wave. The equation above may then be written in vector form as

$$\mathbf{P} = \frac{\text{E}}{c}\,\boldsymbol{u} = \epsilon_0\mathscr{E} \times \mathscr{B}$$

where $\boldsymbol{u}$ is the unit vector in the direction of propagation.

If an electromagnetic wave has momentum, the wave may also have an angular momentum. The angular momentum per unit volume relative to a point O is

$$\mathbf{L} = \mathbf{r} \times \mathbf{P} = \epsilon_0 \mathbf{r} \times (\mathscr{E} \times \mathscr{B})$$

where $\mathbf{r}$ is the vector from O to the point of the wave where $\mathbf{P}$ is evaluated. This value could be called the "orbital" angular momentum of radiation because of the similarity to the angular momentum of an orbiting particle. In addition, electromagnetic radiation possesses an intrinsic angular momentum, or spin, similar to the spin of fundamental particles (recall Section 7.5). For circularly polarized plane waves it can be shown that the spin component along the direction of propagation is equal to $\mp E/\omega$, depending on whether the polarization is clockwise or counterclockwise. For a linearly polarized wave, the average value of the spin component along the direction of propagation is zero because such a wave may be considered as the combination of two oppositely rotating, circularly polarized waves. In summary

> *an electromagnetic wave carries momentum and angular momentum as well as energy.*

This result is not surprising since an electromagnetic field describes the electromagnetic interaction between electric charges, and this interaction means an exchange of energy and momentum between the charges. This exchange is accomplished by means of the electromagnetic field, which is the carrier of the energy and momentum exchanged. The existence of a momentum associated with the electromagnetic field has already been suggested in Section 4.8. The relation $p = E/c$ between the energy and momentum of electromagnetic radiation is particularly important. We shall have the opportunity to refer to the relation again several times, and shall indicate the experimental evidence that supports this assumption. Also, when a charged particle absorbs or emits electromagnetic radiation, not only do the energy and momentum of the particle change but also its angular momentum changes accordingly, a result that has been verified experimentally.

Example 11.1. Radiation pressure.

▼ If electromagnetic waves carry momentum, they must give rise to a pressure when they are reflected or absorbed at the surface of a body. The basic principle is the same as that in the case of the pressure exerted by a gas on the walls of a container.

Consider first some simple cases. Suppose that a plane electromagnetic wave falls perpendicularly on a perfectly absorbing surface (Fig. 11-6). The incident momentum per unit volume is P; and the amount of momentum in the radiation falling on the surface A per unit time is obtained by multiplying P by the volume cA: that is, $\mathrm{P}cA$. If the radiation is completely absorbed by the surface, this expression is also the momentum absorbed per unit time by the surface A; that is, the force on A. Dividing by A gives the pressure caused by the radiation:

$$P_{\text{rad}} = c\mathrm{P} = \mathrm{E} = \epsilon_0 \mathscr{E}^2.$$

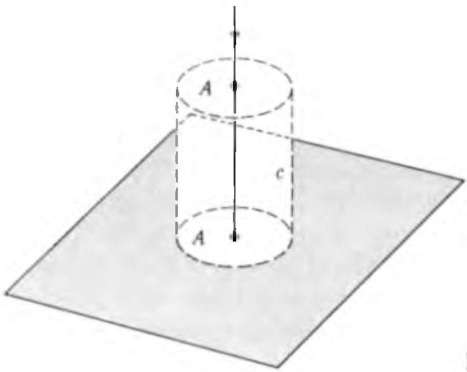


Fig. 11-6. Radiation pressure at normal incidence.

Thus for normal incidence the radiation pressure on a perfect absorber is equal to the energy density in the wave.

On the other hand if the surface is a perfect reflector, the radiation after reflection has a momentum equal in magnitude but opposite in direction to the incident radiation. The change in momentum per unit volume is thus 2P, and the radiation pressure is accordingly

$$P_{\text{rad}} = 2c\text{P} = 2\text{E} = 2\epsilon_0\mathscr{E}^2.$$

These results can be generalized to the case of oblique incidence (Fig. 11-7), in which case the change in momentum of the radiation per unit volume at the perfectly reflecting surface is $2\text{P} \cos\theta$ and the volume is $cA\cos\theta$. The corresponding radiation pressure is

$$P_{\text{rad}} = 2c\text{P}\cos^2\theta = 2\text{E}\cos^2\theta.$$

This result is identical to the pressure produced by a stream of particles falling on a surface if c is replaced by v, the molecular velocity, and P is replaced by nmv. If the radiation propagates in all directions, one must integrate over all directions to obtain the result

$$P_{\text{rad}} = \tfrac{2}{3}c\text{P} = \tfrac{2}{3}\text{E}.$$

When the surface is a perfect absorber, the change in momentum normal to the surface is reduced

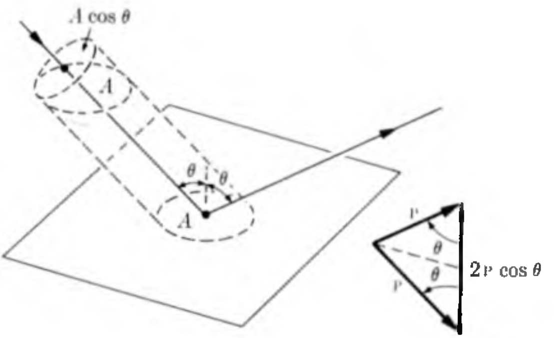


Fig. 11-7. Radiation pressure at oblique incidence. The momentum diagram is shown at the right.

to one-half the previously determined value (because there is no reflected wave carrying momentum), and results in

$$P_{\text{rad}} = \tfrac{1}{3}E.$$

To estimate the radiation pressure on the earth's surface from the sun, we must consider that the incident energy is about 1.4×10^3 W m^{-2}, corresponding to an energy density equal to 4.7×10^{-6} J m^{-3} (when we divide by c). On the assumption that the earth is a perfect absorber and that the radiation comes from all directions, the radiation pressure is $P_{\text{rad}} = \frac{1}{3}E = 1.6 \times 10^{-6}$ Pa. This pressure should be compared with atmospheric pressure, which is about 10^5 Pa. ▲

11.4 Radiation from an Oscillating Electric Dipole

So far, electromagnetic waves have been considered without mentioning how they are produced; in other words, without explaining what the *sources* of the electromagnetic waves are. The sources of electromagnetic waves are clearly the same as the sources of the electromagnetic field; that is, moving electric charges. Given a set of charges in motion, Maxwell's equations give (in principle) the electromagnetic field the charges produce, and therefore the nature of the resulting electromagnetic waves. Instead of considering the general solution of Maxwell's equations for charges in arbitrary motion (which is a very important theoretical problem, but too complicated to be discussed in this book), we shall concentrate on several special but important cases: an oscillating electric dipole, an oscillating magnetic dipole, and finally radiation from accelerated charges in general.

The case of an oscillating electric dipole arises when the motion of the charges can be described collectively by an electric dipole whose moment changes with time according to the law $\Pi = \Pi_0 \sin \omega t$.* This situation could be the case, for example, of an oscillating current in a linear antenna of a broadcasting station or of an electron within an atom when the orbiting motion of the electron is perturbed. When the electric dipole moment is constant, the only field produced is electric as explained in Section 1.9; but when the electric dipole moment is oscillating, the electric field is also oscillating and is therefore time dependent. Therefore a magnetic field is also present (the Ampère-Maxwell law). The magnetic field arises because an oscillating electric dipole is equivalent to a linear oscillating current, and an electric current always produces a magnetic field.

The solution of Maxwell's equations for the case of an oscillating electric dipole is too difficult a mathematical problem to be presented here, but we may use our

*In this chapter the symbol Π is used for the electric dipole moment to avoid confusion with momentum and pressure.

physical intuition to determine its main characteristics. At points very close to the electric dipole, the effect of retardation caused by the finite velocity of propagation of the electromagnetic waves is negligible because the distance r is very small (remember the discussion in Section 4.8). The electric field is then similar to the field created by a static electric dipole, varying as $1/r^3$, as computed in Section 1.9; and the magnetic field is negligible. At large distances, however, the finite propagation of the waves produces a modification in the field. The solution of the wave equation for spherical waves of equal amplitude in all directions was given in Section 10.13 and suggests that in this case (although there is no spherical symmetry but rather axial symmetry around the axis of oscillation of the dipole) the electromagnetic field may depend asymptotically on the distance by a factor $1/r$, instead of $1/r^3$ as for small distances. (This suggestion is corroborated by the actual solution of Maxwell's equations.) In addition at large distances when a small portion of the wave front looks like a plane wave, the electric field is perpendicular to the direction of propagation, which is along the radius vector r, so that $\mathscr{E}_r=0$.

If we assume that the Z-axis is oriented parallel to the oscillating electric dipole, the magnitude of the electric field is then found to be

$$\mathscr{E}=\frac{\Pi_0 \sin\theta}{4\pi\epsilon_0 r}\left(\frac{\omega}{c}\right)^2 \sin(kr-\omega t), \tag{11.20}$$

and has the orientation indicated in Fig. 11-8. On the other hand since the oscillating dipole corresponds to a current along the Z-axis, the magnetic field is parallel to the XY-plane. From the relation $\mathscr{B}=\mathscr{E}/c$, the magnitude of the magnetic field is

$$\mathscr{B}=\frac{1}{c}\mathscr{E}=\frac{\Pi_0 \sin\theta}{4\pi\epsilon_0 cr}\left(\frac{\omega}{c}\right)^2 \sin(kr-\omega t), \tag{11.21}$$

with the field oriented as indicated in Fig. 11-8. Again, $\mathscr{E}$ and $\mathscr{B}$ are perpendicular to each other and to the direction of propagation. Note that both $\mathscr{E}$ and $\mathscr{B}$ are zero

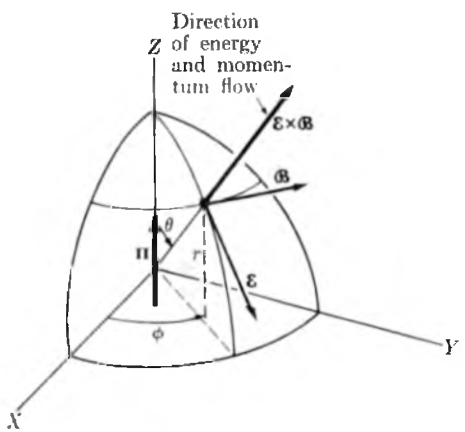


Fig. 11-8. Electric and magnetic fields produced by an oscillating electric dipole.

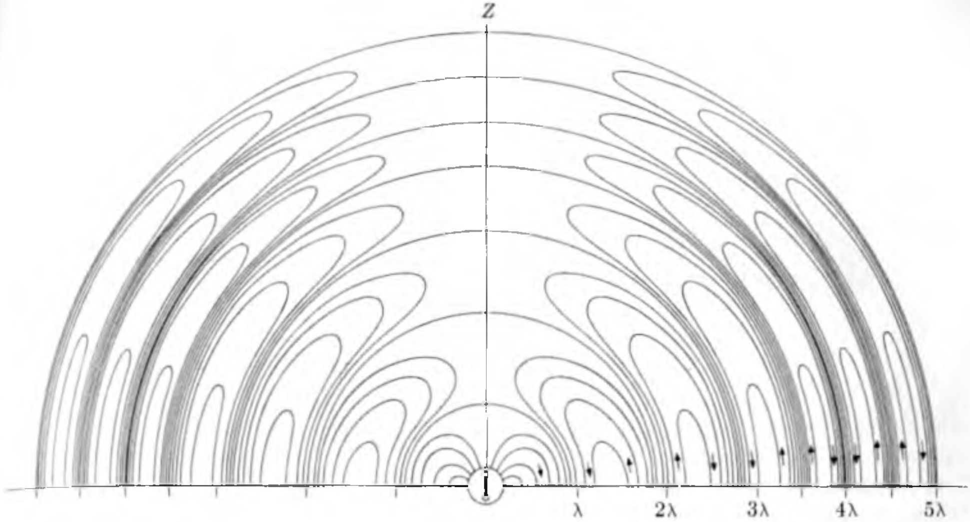


Fig. 11-9. Electric field lines produced by an oscillating electric dipole.

for $\theta = 0$ and π; that is, for points along the Z-axis. Thus the amplitude of the electromagnetic wave of an oscillating electric dipole is zero along the direction of oscillation. On the other hand, $\sin\theta$ has its maximum value for $\theta = \pi/2$, or points on the XY-plane. Therefore the electromagnetic wave of an oscillating electric dipole has its maximum intensity in the equatorial plane of the electric dipole. The waves are linearly polarized with the electric field oscillating in a meridian plane. Figure 11-9 is a description of the electric lines of force in such a meridian plane at a particular time. Each loop corresponds to one complete oscillation. The magnetic lines of force are circles parallel to the XY-plane with their centers on the Z-axis.

The vector $\mathscr{E} \times \mathscr{B}$ has the direction of r as shown in Fig. 11-8, and so energy and momentum flow away from the electric dipole in the radial direction; therefore to keep the electric dipole oscillating, energy must be supplied to it. If we use Eqs. (11.15) and (11.20), we find that the energy density in the wave at large distances from the oscillating electric dipole is

$$\mathrm{E} = \epsilon_0 \mathscr{E}^2 = \frac{\Pi_0^2 \sin^2\theta\, \omega^4}{16\pi^2 \epsilon_0 r^2\, c^4} \sin^2(kr - \omega t).$$

Since $[\sin^2(kr - \omega t)]_{\text{ave}} = \frac{1}{2}$, the average energy density is

$$\mathrm{E}_{\text{ave}}(\theta) = \frac{\Pi_0^2 \omega^4}{32\pi^2 c^4 \epsilon_0 r^2} \sin^2\theta. \tag{11.22}$$

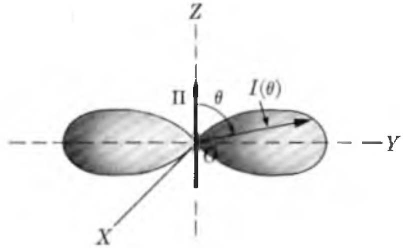


Fig. 11-10. Angular dependence of the intensity of the electromagnetic radiation produced by an oscillating electric dipole.

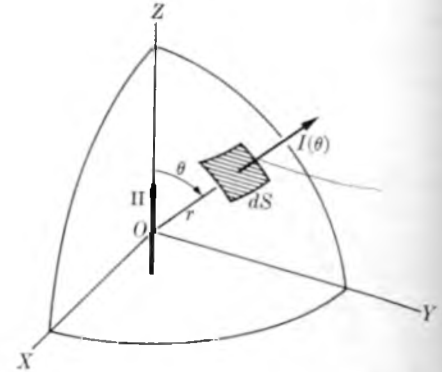


Fig. 11-11. Calculation of the total energy radiated per unit time by an oscillating electric dipole.

The intensity of the radiation from the oscillating electric dipole (that is, the energy passing per unit area and unit time in the direction of propagation) is

$$I(\theta)_{\text{ave}} = c\mathrm{E}_{\text{ave}}(\theta) = \frac{\Pi_0^2\omega^4}{32\pi^2 c^3 \epsilon_0 r^2}\sin^2\theta. \tag{11.23}$$

This intensity shows two interesting features. In the first place it exhibits the $1/r^2$ dependence expected from the discussion of spherical waves in Section 10.13. In addition the intensity has an angular dependence, proportional to $\sin^2\theta$. Therefore the intensity of electric dipole radiation is maximum in the equatorial plane and zero along the axis of the oscillating electric dipole; that is, *an oscillating electric dipole does not radiate energy along its axis.* The angular dependence of $I(\theta)$ is shown in Fig. 11-10.

The total energy radiated per unit time by the dipole is calculated as follows. Since the energy flows in the radial direction, we draw a sphere of very large radius around the dipole (Fig. 11-11). The average energy passing per unit time through the small area dS is $I(\theta)_{\text{ave}}\,dS$, and therefore the average energy radiated through the entire sphere per unit time is

$$\left\langle \frac{dE}{dt} \right\rangle = \int_{\text{Sphere}} I(\theta)_{\text{ave}}\,dS = \frac{\pi_0^2\omega^4}{32\pi^2 c^3 \epsilon_0 r^2}\int_{\text{Sphere}} \left[\sin^2\theta\right]_{\text{ave}} dS. \tag{11.24}$$

The computation of this integral is a mathematical exercise and will be omitted. The result is

$$\left\langle \frac{dE}{dt} \right\rangle = \frac{\Pi_0^2\omega^4}{12\pi\epsilon_0 c^3}. \tag{11.25}$$

Since the electric dipole moment may be written as qz where q is the oscillating charge, and $z = z_0 \sin\omega t$ is the displacement of the charge along the Z-axis, Π_0 may be replaced by qz_0 where z_0 is the amplitude of the oscillations. In many cases the oscillating charge is a proton within a nucleus or an electron within an atom so that q is equal to the fundamental charge $\pm e$. Then Eq. (11.25) becomes

$$\left\langle \frac{dE}{dt} \right\rangle = \frac{e^2 z_0^2 \omega^4}{12\pi\epsilon_0 c^3}. \tag{11.26}$$

In the case of an electron in an atom. the quantity z_0 is of the order of magnitude of the atomic size, or about 10^{-10} m. Introducing the values of the other constants, we see then that for atomic electric dipole radiation

$$\left\langle \frac{dE}{dt} \right\rangle \sim 10^{-74} \omega^4 \text{ W}.$$

In the optical region, ω is of the order of 10^{14} Hz, and therefore $\langle dE/dt \rangle \sim 10^{-18}$ W or 10 eV s^{-1}, a quantity small by engineering standards but appreciable from the atomic point of view.

Electric dipole radiation is one of the most effective ways for producing electromagnetic waves and constitutes the most important mechanism by which atoms, molecules, and nuclei emit (and absorb) electromagnetic radiation. However to discuss electric dipole radiation by atoms, molecules, and nuclei requires use of the methods of quantum mechanics. Therefore the results stated here and in succeeding sections give only a rough estimate of the orders of magnitude involved. One of the more important experimental results to be considered in the quantum-mechanical treatment is that an atom does emits radiation not continuously but in bursts. Another experimental result that must be accounted for is that the radiation emitted by atoms (or molecules or nuclei) is composed of a well-defined set of frequencies $\omega_1, \omega_2, \omega_3, \ldots$, characteristic of each atom, molecule, or nucleus; this set is called the *emission spectrum* of the substance, a fact mentioned in Section 2.7.

As pointed out in the discussion of Fig. 11-9, electric dipole radiation is polarized, with the electric field always in a meridian plane. However, the human eye does not seem to be sensitive to the direction of polarization of an electromagnetic wave. It is interesting that certain insects, however, do seem to be sensitive to polarization. In addition in most substances the radiating atomic dipoles are oriented at random, and no net polarization is observed in the total radiation from the substance.

Example 11.2. Power transmitted by the antenna of a radio station in the electric dipole approximation.

▼ An antenna in simplified form is just a wire of length z_0 in which an oscillating current is maintained. The current is related to the charges by $I = dq/dt$, and therefore the current amplitude is $I_0 = q\omega$. Therefore $\Pi_0 = qz_0 = I_0 z_0/\omega$. Introducing this relation into Eq. (11.25) gives

$$\left\langle \frac{dE}{dt} \right\rangle = \frac{I_0^2 \omega^2 z_0^2}{12\pi\epsilon_0 c^3}. \tag{11.27}$$

This expression gives the power required to broadcast at a frequency ω. From the discussion of Ohm's law applied to circuits with alternating currents (Section 9.5) the average power required to maintain a current is $\frac{1}{2}RI_0^2$ [Eq. (9.30)]. Accordingly we rewrite Eq. (11.27) in the form

$$\left\langle \frac{dE}{dt} \right\rangle = \frac{1}{2}\left(\frac{\omega^2 z_0^2}{6\pi\epsilon_0 c^3}\right) I_0^2. \tag{11.28}$$

and by analogy define

$$R = \frac{\omega^2 z_0^2}{6\pi\epsilon_0 c^3} = \frac{2\pi}{3}\sqrt{\frac{\mu_0}{\epsilon_0}}\left(\frac{z_0}{\lambda}\right)^2 \tag{11.29}$$

as the antenna's *radiation resistance*. It is expressed in ohms as may be verified from its dimensions in terms of fundamental units. (The total resistance of the antenna is, of course, the radiation resistance plus the conduction resistance.) Introducing numerical values in Eq. (11.29) gives $R = 787(z_0/\lambda)^2$ ohms. Note that both Eqs. (11.28) and (11.29) for a linear antenna have been derived by means of the electric dipole approximation, and thus these equations are valid only if the length z_0 is very small compared with the wavelength of the radiation.

For example consider a 30-m-long linear antenna that radiates electromagnetic waves of frequency 5×10^5 Hz with an rms (or root-mean-square) value of the current of 20 A. Using Eq. (11.29) with $\omega = 2\pi\nu = 3.14 \times 10^6\ \mathrm{s}^{-1}$ and $z_0 = 30$ m gives $R = 1.97\ \Omega$ for the radiation resistance. Since $I_{rms} = I_0/\sqrt{2}$ (remember Problem 9.28), $I_{rms}^2 = \frac{1}{2}I_0^2$. Therefore the power radiated is

$$\left\langle \frac{dE}{dt} \right\rangle = RI_{rms}^2 \sim 400\ \mathrm{W}.$$

Note that in this case $\lambda = c/\nu = 600$ m so that $z_0/\lambda \ll 1$, and this approximation can be used. ▲

11.5 Radiation from an Oscillating Magnetic Dipole

Another source of electromagnetic waves is an oscillating magnetic dipole. The interrelation between the fields of a magnetic dipole is similar to that of an electric dipole, except that the roles of the electric and magnetic fields are interchanged.

We have defined a magnetic dipole as a small current loop, the magnetic moment being $\mathscr{M} = IA$ where I is the current and A the area of the loop. Suppose that the loop lies in the XY-plane with its center at the origin of a coordinate system (Fig. 11-12). If the current oscillates with frequency ω so that it is given by $I = I_0 \sin \omega t$, the magnetic moment is $\mathscr{M} = \mathscr{M}_0 \sin \omega t$ where $\mathscr{M}_0 = I_0 A$. A static magnetic dipole produces only a constant magnetic field; but when the magnetic dipole oscillates, its magnetic field at each point of space is also time dependent. Thus an electric field is also present as was the case when the Faraday-Maxwell law was discussed.

Just as for the oscillating electric dipole, the derivation of the exact expressions for the electric and magnetic fields will be omitted. At points close to the magnetic dipole

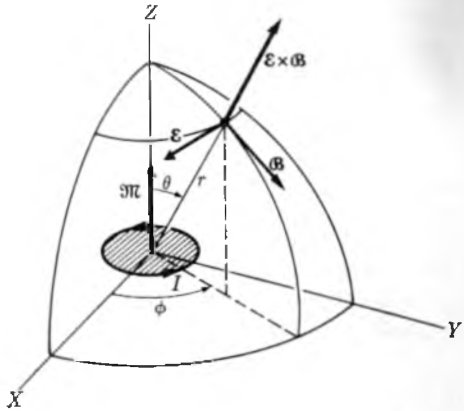


Fig. 11-12. Electric and magnetic fields produced by an oscillating magnetic dipole.

the effect of retardation caused by the finite velocity of propagation of the electromagnetic waves is negligible because the distance r is very small. The magnetic field is then similar to that of a static magnetic dipole varying as $1/r^3$ as explained in Section 5.3. and the electric field is negligible. At large distances, however, the finite propagation velocity of the waves produces a noticeable modification of the field. As in the case of an oscillating electric dipole, we may expect a solution depending asymptotically on $1/r$ instead of $1/r^3$ oscillating with the electric and the magnetic fields in a plane perpendicular to the direction of propagation of the waves. However for magnetic dipole radiation the magnetic field is in a meridian plane and the electric field in a transverse direction so that the electric lines of force are circles concentric with the Z-axis. In this approximation the fields are

$$\mathscr{E}=\frac{\mu_0 c}{4\pi}\frac{\mathscr{M}_0 \sin\theta}{r}\left(\frac{\omega}{c}\right)^2 \sin(kr-\omega t)$$

and (11.30)

$$\mathscr{B}=\frac{\mu_0}{4\pi}\frac{\mathscr{M}_0 \sin\theta}{r}\left(\frac{\omega}{c}\right)^2 \sin(kr-\omega t).$$

Note that the relation $\mathscr{B}=\mathscr{E}/c$ still holds. The relative orientation of the $\mathscr{E}$ and $\mathscr{B}$ fields for an oscillating magnetic dipole is illustrated in Fig. 11-12. Note that the vector $\mathscr{E}\times\mathscr{B}$ is still in the outward radial direction. The wave is plane polarized with the magnetic field oscillating in a meridian plane. In other words the plane of polarization is rotated 90° with respect to electric dipole waves.

By the same reasoning as that with the oscillating electric dipole, the average energy density of the radiation emitted by an oscillating magnetic dipole is

$$\mathrm{E}_{\mathrm{ave}}(\theta)=\epsilon_0\mathscr{E}^2_{\mathrm{ave}}=\frac{\mathscr{M}_0^2\omega^4}{32\pi^2\epsilon_0 c^6 r^2}\sin^2\theta \tag{11.31}$$

where $\mu_0 = 1/\epsilon_0 c^2$ has been used. The intensity of the radiation from the magnetic dipole, given by $I(\theta)_{ave} = c\text{E}_{ave}(\theta)$, is again zero along the axis of the dipole (Z-axis), and is maximum in the equatorial plane, a situation similar to that found for the oscillating electric dipole. The average energy radiated per unit time by the oscillating magnetic dipole is

$$\left\langle \frac{dE}{dt} \right\rangle = \frac{\mathscr{M}_0^2 \omega^4}{12\pi\epsilon_0 c^5}. \tag{11.32}$$

This result is obtained by following the same procedure used for the electric dipole.

In the case of an electron in an atom we have from Eq. (7.31) that $\mathscr{M}_0 = -(e/2m_e)L$ where L is the orbital angular momentum of the electron so that

$$\left\langle \frac{dE}{dt} \right\rangle = \frac{(e/2m_e)^2 L^2 \omega^4}{12\pi\epsilon_0 c^5}. \tag{11.33}$$

The quantity $e/2m_e$ is 8.794×10^{10} C kg^{-1}, and the angular momentum L is of the order of 10^{-34} J s^{-1} [see Eq. (7.24)] so that

$$\left\langle \frac{dE}{dt} \right\rangle \approx 10^{-79} \omega^4 \text{ W}.$$

When we compare this result with the corresponding result for an electric dipole, we conclude that for atoms (and also molecules) the ratio of the intensity of the magnetic dipole radiation to the electric dipole radiation is of the order of 10^{-5}. Therefore for the same frequency the magnetic dipole radiation from atoms is negligible compared with the electric dipole radiation, and must be taken into consideration only when the electric dipole radiation is absent. Actually since $L = mrv$ and r and z_0 are of the same order of magnitude, we find that

$$\left\langle \frac{dE}{dt} \right\rangle_{\text{Magnetic dipole}} \approx \left(\frac{v}{c}\right)^2 \left\langle \frac{dE}{dt} \right\rangle_{\text{Electric dipole}}$$

so that only for very fast electrons are the two comparable.

Example 11.3. Magnetic dipole radiation from an antenna.

▼ Equation (11.32), applied to a radiating circular antenna with $\mathscr{M}_0 = I_0 A$, gives the average power required to drive the antenna. This power is written as

$$\left\langle \frac{dE}{dt} \right\rangle = \frac{I_0^2 A^2 \omega^4}{12\pi\epsilon_0 c^5}. \tag{11.34}$$

Comparison with Eq. (11.27) for the radiation from an electric dipole antenna gives

$$\frac{\langle dE/dt \rangle_{\text{Magnetic dipole}}}{\langle dE/dt \rangle_{\text{Electric dipole}}} = \left(\frac{A\omega}{z_0 c}\right)^2.$$

However $\omega/c = k = 2\pi/\lambda$ and A is of the order of magnitude of z_0^2. Therefore

$$\frac{\langle dE/dt\rangle_{\text{Magnetic dipole}}}{\langle dE/dt\rangle_{\text{Electric dipole}}} \approx \left(\frac{2\pi z_0}{\lambda}\right)^2. \tag{11.35}$$

Since z_0 is normally much smaller than λ for broadcasting antennas, we conclude again that the magnetic mode of radiation is much weaker than the electric mode.

For an atom z_0 is of the order of 10^{-10} m, and in the optical region λ is about 10^{-7} m and gives a value of the order of 10^{-5} for the ratio appearing in Eq. (11.35), in agreement with the previous estimate. On the other hand for nuclei, z_0 is of the order of 10^{-14} m and λ is of the order of 10^{-12} m so that the ratio (11.35) is about 10^{-3}. ▲

Example 11.4. Radiation resistance of a loop antenna.

▼ Consider a circular, 30-m-long antenna that carries an rms current of 20 A oscillating with a frequency of 5×10^5 Hz. (These are the same values used in Example 11.2.) From Eq. (11.34)

$$\left\langle\frac{dE}{dt}\right\rangle = \frac{1}{2}\left(\frac{A^2\omega^4}{6\pi\epsilon_0 c^5}\right) I_0^2$$

where the terms have been arranged to give the form $\langle dE/dt\rangle = \frac{1}{2}RI_0^2$. The radiation resistance of a loop antenna is then

$$R = \frac{A^2\omega^4}{6\pi\epsilon_0 c^5} = \frac{8\pi^3}{3}\sqrt{\frac{\mu_0}{\epsilon_0}}\left(\frac{A}{\lambda^2}\right)^2 = 31{,}170\left(\frac{A}{\lambda^2}\right)^2 \text{ ohms.}$$

In this case the radius is $30/2\pi$ m and the area is $A = 900/4\pi = 71.6$ m^2. Therefore $R = 1.23 \times 10^{-3}\ \Omega$. The average power radiated is

$$\left\langle\frac{dE}{dt}\right\rangle = RI_{\text{rms}}^2 = 0.25 \text{ W}.$$

These results should be compared with the results of Example 11.2. ▲

11.6 Radiation from Higher-Order Oscillating Multipoles

In the two previous sections we have considered the radiation emitted from electric and magnetic dipoles; but Chapters 1 and 5 discussed higher-order multipoles, both electric and magnetic, related to different charge and current arrangements. If these multipoles oscillate, they produce electromagnetic waves that differ in their angular distribution and state of polarization from the dipole waves. In general the higher the order of the multipole, the lower the intensity of the radiation when compared with a dipole of similar dimensions and the same frequency; that is, electric dipole radiation is the most important mechanism for radiation in atomic systems. For example if r_0 is the order of magnitude of the dimensions of the system and λ the wavelength, the ratio between electric quadrupole radiation and electric dipole radiation is of the order of $(r_0/\lambda)^2$. For atoms, r_0 is of the order of 10^{-10} m; and for visible light, λ is of the order of 10^{-7} m so that $(r_0/\lambda)^2$ is about 10^{-6}. For nuclei, on the other hand, r_0 is of the order of 10^{-14} m and λ is of the order of 10^{-12} m so that $(r_0/\lambda)^2$ is about 10^{-4} and electric quadrupole radiation is relatively more important. Note that electric quadrupole radiation is of the same order as magnetic dipole radiation. Although higher-order multipole radiation is much weaker than dipole radiation, to explain some transitions in certain nuclei, the electric quadrupole and even the electric octupole radiation must be taken into consideration.

11.7 Radiation from an Accelerated Charge

The electromagnetic radiations studied in Sections 11.4 and 11.5 were related to two special charge configurations. We will now discuss the production of radiation in general, resulting from the acceleration of charged bodies. First, however, it is important to understand why an unaccelerated charge does not radiate.

Consider the case of a charge in uniform motion; that is, a charge moving with constant velocity $\boldsymbol{v}$. The electric and the magnetic fields of such a charge were discussed in Section 4.7. The electric field is radial, and the magnetic field is transverse with circular lines of force concentric with the line of motion. Figure 11-13 shows the electric field $\mathscr{E}$ and the magnetic field $\mathscr{B}$ at the four symmetric points P_1, P_2, P_3, and P_4. At each point the vector $\mathscr{E} \times \mathscr{B}$ has also been indicated. From the figure it can be seen that the contribution of the components of $\mathscr{E} \times \mathscr{B}$ perpendicular to the direction of motion cancel each other because of symmetry; the components of $\mathscr{E} \times \mathscr{B}$ parallel to the direction of motion are all in the forward direction and add to each other. There is therefore a net flux of energy in the direction in which the charge is moving; this is understandable from a physical point of view since the particle carries the field with itself (and therefore the field's energy and momentum as well). At points behind the moving charge in the laboratory frame of reference, the electromagnetic field is decreasing; at points ahead of the charge, the field is increasing by the same amount. The transfer of energy in the direction of motion of the charge gives rise to the energy flux.

To see if energy is radiated by a charge in uniform motion, we must compute the flux of the vector $\mathscr{E} \times \mathscr{B}$ through a closed surface surrounding the charge. Using Eq. (11.18) for a closed surface gives

$$\frac{dE}{dt} = c^2 \epsilon_0 \oint_S \mathscr{E} \times \mathscr{B} \cdot \boldsymbol{u}_N \, dS$$

when a sphere of radius r concentric with the charge is chosen for the closed surface. From Fig. 11-13 it is seen that the vector $\mathscr{E} \times \mathscr{B}$ is tangent to the spherical surface at

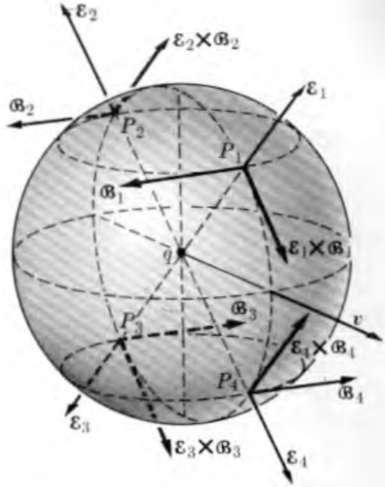


Fig. 11-13. Electric and magnetic fields of a uniformly moving charge.

all its points, and is therefore everywhere perpendicular to the unit vector $\boldsymbol{u}_N$ normal to the surface. Thus

$$\mathcal{E} \times \mathcal{B} \cdot \boldsymbol{u}_N = 0,$$

and the net energy flux across the spherical surface is zero. We conclude then that

> *a charge in uniform rectilinear motion does not radiate electromagnetic energy but carries the energy of the electromagnetic field along with itself.*

Since the field is static and the energy remains constant in the inertial frame of reference of the charge, the total energy must also remain constant in the laboratory frame of reference. There is merely a steady flow of energy along the direction of motion of the charge.

A very different situation exists for a charge that is in accelerated motion. The electric field of an accelerated charge is no longer radial and does not have the left-to-right symmetry present when the charge is in uniform motion. Since it is complicated, the expression for the field will not be given here; but its lines of force have a pattern similar to that shown in Fig. 11-14. When the particle moves, the field on the left decreases and the one on the right increases; but because of the acceleration, the increase of the field (corresponding to the new, larger velocity) is greater than the decrease of the field that existed previously (corresponding to an earlier, smaller velocity). Therefore a net excess energy must be transferred to all space to build up the field. Thus

an accelerated charge radiates electromagnetic energy.

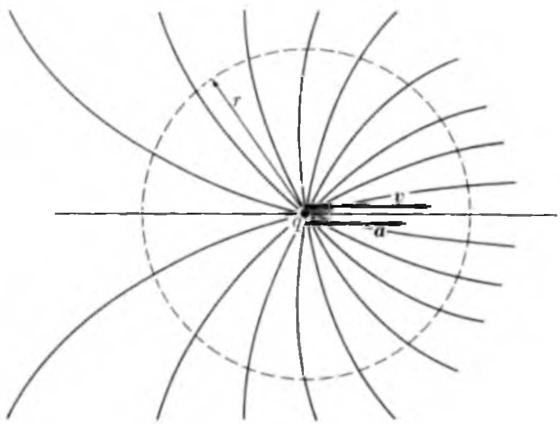


Fig. 11-14. Electric lines of force produced by an accelerated charge.

Accordingly, the equation of motion of a charged particle under an applied force must be modified to take into account the radiation of energy. This factor has not previously been considered because in most practical cases (in which the acceleration is small and the velocity is much less than c), the radiation is negligible.

By using the appropriate values for the fields $\mathscr{E}$ and $\mathscr{B}$, it may be proved that if the accelerated charge is momentarily at rest or is moving slowly relative to the observer (so that all retardation effects due to the finite velocity of propagation of the wave can be neglected), the energy radiated per unit time through a spherical surface of radius r around the charge is

$$\frac{dE}{dt}=\frac{q^2a^2}{6\pi\epsilon_0c^3} \tag{11.36}$$

where a is the acceleration of the charge. This result is called *Larmor's formula.* For a charge oscillating along the Z-axis with simple harmonic motion, the acceleration is given by $a=-\omega^2z$, and Larmor's formula for the oscillating charge is

$$\frac{dE}{dt}=\frac{q^2z^2\omega^4}{6\pi\epsilon_0c^3}.$$

To obtain the average energy radiated, we replace z^2 by $(z^2)_{\text{ave}}$, which is equal to $\frac{1}{2}z_0^2$. After this substitution, Larmor's formula becomes Eq. (11.26), the equation that gives the energy radiated by an oscillating electric dipole.

One important conclusion from the study of accelerated charges is that to maintain a charge in accelerated motion, energy must be supplied to compensate for the energy lost by radiation. Thus for example when an ion is accelerated in a linear accelerator, such as a Van de Graaff machine, a fraction of the energy supplied to the ion is lost as electromagnetic radiation. This loss is negligible, however, except at relativistic energies.

If the acceleration is parallel to the velocity, the angular distribution of the radiated energy is similar to that illustrated in Fig. 11-10 for an electric dipole as long as the velocity of the particle is small compared to the velocity of light. That is, using Eq. (11.23) with certain changes (replace Π_0^2 by $q^2 z_0^2$ and $z_0\omega^2$ by the acceleration a and note that using instantaneous values rather than average values necessitates removing a factor of $\frac{1}{2}$), the intensity of the radiation in the direction given by the angle θ with respect to the direction of the velocity can be expressed by

$$I(\theta)=\frac{q^2a^2}{16\pi^2c^3\epsilon_0 r^2}\sin^2\theta. \tag{11.37}$$

The angular distribution $I(\theta)$ is symmetric relative to a plane through the charge and perpendicular to the direction of motion as shown in Fig. 11-15. However in the high energy region the intensity of the energy radiated by an accelerated charge has its maximum over a conical surface oriented in the direction of motion of the particle as also indicated in Fig. 11-15. The angle of the cone decreases as the velocity of the particle increases.

If the particle is decelerated instead of accelerated, expression (11.36) still holds, and the energy radiated is that which the electromagnetic field has in excess at each moment as a result of the decrease in the velocity of the charge. This situation occurs, for example, when a fast charge, such as an electron or a proton, hits a target. A substantial part of the total energy of the charge goes off as radiation, called *deceleration radiation*, or more commonly *bremsstrahlung* (from the German *Bremsung*,

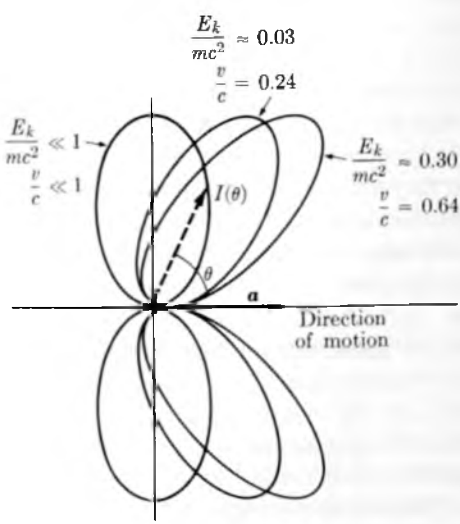


Fig. 11-15. Angular distribution of the radiation emitted by an accelerated charge for different values of v/c and with kinetic energy E_k.

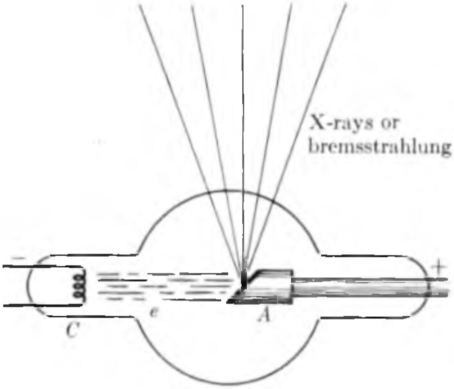


Fig. 11-16. Radiation emitted by a charge decelerated when hitting the target A in an X-ray tube. The target must be constructed from material with a high melting point and it must be continuously cooled.

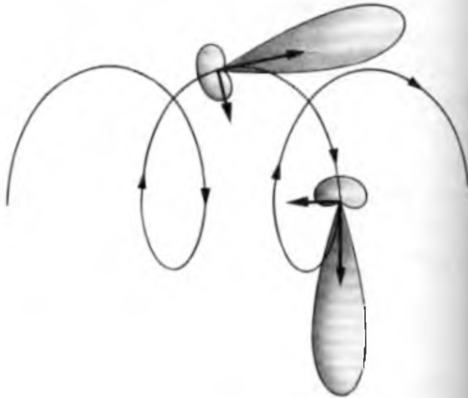


Fig. 11-17. Synchrotron radiation of a particle moving in a magnetic field. The angular distribution of the intensity is shown at two positions. The directions of the velocity and acceleration are also shown.

deceleration, and *Strahlung*, radiation) (Fig. 11-16). This process is the main mechanism by which radiation is produced in X-ray tubes used for physical, medical, and industrial applications.

Although Fig. 11-15 shows the case for which the acceleration is in the same direction as the motion, this discussion holds true for any kind of motion in which there is acceleration. For example a charged particle moving in a circular path has a centripetal acceleration, and hence emits radiation. Therefore when an ion is accelerated in a cyclical accelerator, such as a cyclotron, a betatron, or a synchrotron, a fraction of the energy applied to the ion is lost as electromagnetic radiation, an effect that is relatively more important in cyclical than in linear accelerators because the accelerations are generally larger. Our previous discussion of the cyclotron and the betatron did not take this fact into account because this omission is justified when the energy involved is not very great and the acceleration is small.

When particles reach high energies, as they do in synchrotrons, where the acceleration is large, the loss produced by radiation, called *synchrotron radiation*, becomes very important and constitutes a serious limitation in the construction of cyclic accelerators of very high energy. When a particle trapped in a magnetic field spirals as discussed in Section 4.3, the particle also emits synchrotron radiation. Since electromagnetic radiation is emitted preferentially in a direction perpendicular to the acceleration (see Fig. 11-15), and since the acceleration is pointing toward the axis of the helix and is perpendicular to the velocity, synchrotron radiation is emitted mainly in the direction of motion within a cone whose axis is tangent to the electron's path as indicated in Fig. 11-17. Radiation coming from the charged particles trapped in the earth's magnetic field, from sun spots, or from some more distant bodies (such

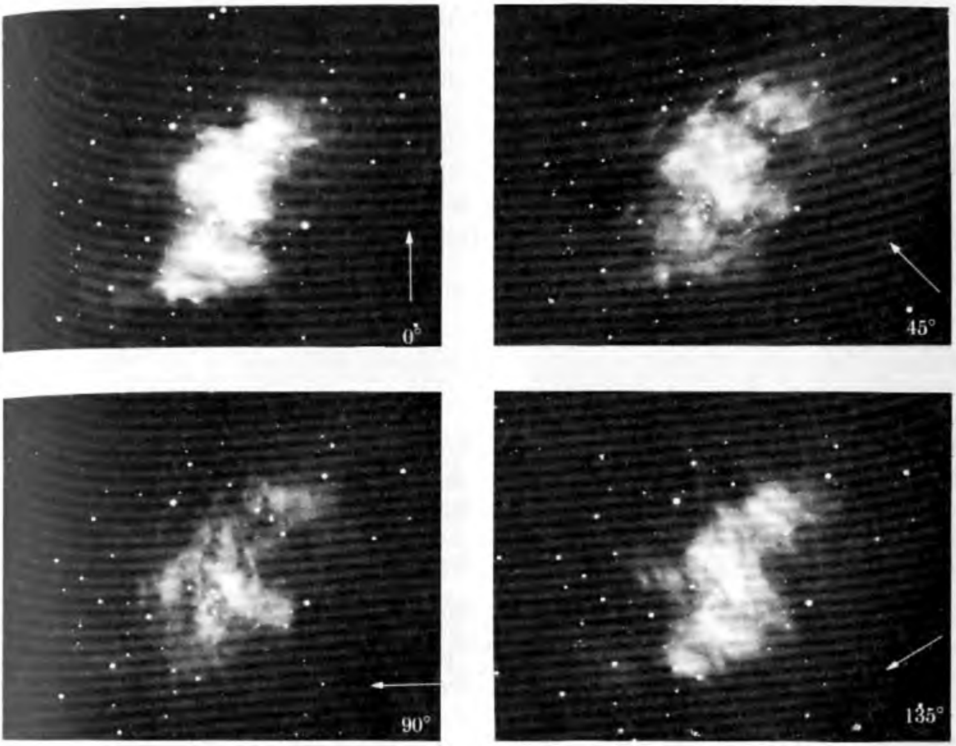


Fig. 11-18. Synchrotron radiation from the Crab Nebula. Each photograph was taken through a device that accepts only radiation with the electric vector as shown. The fact that the photographs are different indicates that the radiation is polarized. (Photographs courtesy of Mt. Wilson and Palomar Observatories.)

as certain nebulae) shows the same characteristics. Figure 11-18 shows four photographs of the Crab Nebula. The radiation received, which extends from radio frequencies to the extreme ultraviolet, is thought to be synchrotron radiation from electrons having an energy up to about 10^{12} eV and moving in circular or helical orbits in a magnetic field of the order of 10^{-8} T. The radiation shows a strong polarization as can be seen from the differences in the photographs, which were taken through a polarizing filter, allowing only radiation with the electric field in the specified direction to be photographed. The arrows indicate the direction of the electric field.

Another interesting consideration is related to atomic structure. In Section 7.3 we indicated that as a result of Rutherford's experiments on the scattering of alpha particles, we picture the atom as formed by a central nucleus, positively charged, with negatively charged electrons describing closed orbits around it. However, this conceptualization means that the electrons are moving with accelerated motion, and if the ideas expressed in this section are applied, all atoms would be radiating energy continuously. As a result of this loss of energy, the electron orbits would be shrinking, and there would be a corresponding contraction in size of all bodies. Of course if all bodies were identical, this contraction would be impossible to detect since it would affect the bodies measured and the measuring ruler equally. However since atoms of different elements are different, they would shrink at different rates, and the effect would be noticeable. However neither this contraction of matter nor the continuous radiation associated with it has been observed. Therefore it must be concluded that the motion of atomic electrons is governed by some additional principles that we have not yet considered. This behavior is explained by the principles of quantum mechanics.

Example 11.5. Energy radiated per unit time by an accelerated charge for any velocity of the charge and any direction of the acceleration.

▼ Larmor's formula, Eq. (11.36), is strictly correct only when the particle is momentarily at rest relative to an observer. To obtain the value of the energy radiated by the charge as measured by an observer who sees the particle moving with velocity v, we must simply make a Lorentz transformation of all quantities involved in that expression. Suppose that the charge is momentarily at rest relative to an observer O' who uses the frame of reference $X'Y'Z'$. Equation (11.36) is

$$\frac{dE'}{dt'}=\frac{q^2a'^2}{6\pi\epsilon_0c^3}.$$

For an observer O in the frame XYZ, relative to whom the particle has a velocity v, dE'/dt' is replaced by dE/dt where dt and dt' are the respective time intervals. Since these two time intervals correspond to the same point in $X'Y'Z'$, they are related by $dt=dt'/\sqrt{1-v^2/c^2}$ [see Eq. (6.36) of Volume I or the appendix]. Similarly dE and dE' (which are the changes in energy of a particle having zero momentum relative to $X'Y'Z'$) are related by $dE=dE'/\sqrt{1-v^2/c^2}$ [see Eq. (11.24) of Volume I or the appendix]. (An alternative logic would be to remember that E/c transforms as ct does since they are both the fourth component of four-vectors.) Therefore $dE/dt=dE'/dt'$.

To transform the right-hand side of Larmor's formula, the accelerations of the particle as measured by the observers are related through

$$a'^2=\frac{a^2-(\boldsymbol{v}\times\boldsymbol{a})^2/c^2}{(1-v^2/c^2)^3}$$

(see Problem 6.40 of Volume I). Therefore

$$\frac{dE}{dt}=\frac{q^2}{6\pi\epsilon_0c^3}\frac{a^2-(\boldsymbol{v}\times\boldsymbol{a})^2/c^2}{(1-v^2/c^2)^3}, \tag{11.38}$$

a result known as *Liénard's formula*, first derived by A. Liénard in 1898, before the theory of relativity was developed. It can be proved that Liénard's formula already incorporates the retardation effects caused by the finite velocity of propagation of electromagnetic radiation.

If the acceleration is parallel to the velocity, $\boldsymbol{v}\times\boldsymbol{a}=0$ and Eq. (11.38) reduces to

$$\left(\frac{dE}{dt}\right)_{\parallel}=\frac{q^2a^2}{6\pi\epsilon_0c^3(1-v^2/c^2)^3}. \tag{11.39}$$

This expression must be used to estimate radiation losses in linear accelerators. On the other hand when the acceleration is perpendicular to the velocity as in the case of a circular orbit, $(v \times a)^2 = v^2 a^2$ and Eq. (11.38) reduces to

$$\left(\frac{dE}{dt}\right)_\perp = \frac{q^2 a^2}{6\pi\epsilon_0 c^3 (1 - v^2/c^2)^2}. \tag{11.40}$$

This expression is used for computing synchrotron radiation. In both cases, dE/dt increases very rapidly as v approaches c. At very low velocity ($v \ll c$), both expressions (11.39) and (11.40) become identical with Eq. (11.36). ▲

Example 11.6. The energy radiated by a proton accelerated in a Van de Graaff accelerator.

▼ If t is the time required by the proton to travel the length of the accelerator tube (on the assumption that the motion is nonrelativistic) and v is the proton's final velocity, then $v = at$. For such an assumption, Eq. (11.36) applies. Noting that the acceleration a is constant, the total energy lost by the proton through radiation in the time t [with $v = at$ and $q = e$ in Eq. (11.36)] is

$$E_{\text{rad}} = \left(\frac{dE}{dt}\right) t = \frac{e^2 v^2}{6\pi\epsilon_0 c^3 t}.$$

However if s is the length of the accelerator tube, $s = \frac{1}{2}at^2 = \frac{1}{2}(at)t = \frac{1}{2}vt$. Thus $t = 2s/v$ and

$$E_{\text{rad}} = \frac{e^2 v^3}{12\pi\epsilon_0 c^3 s}.$$

On the other hand the kinetic energy gained by the proton in going through the potential difference V is $E_k = \frac{1}{2}m_p v^2 = eV$. Thus

$$\frac{E_{\text{rad}}}{E_k} = \frac{e^2 v}{6\pi\epsilon_0 c^3 m_p s} = \frac{e^2}{6\pi\epsilon_0 c^3 m_p s}\left(\frac{2eV}{m_p}\right)^{1/2}$$

since $v = (2eV/m_p)^{1/2}$. For an accelerator having a length of 2.0 meters and a potential difference of 5×10^5 V, we find $E_{\text{rad}}/E_k = 1.7 \times 10^{-20}$. Therefore radiation losses by protons can be considered negligible for this accelerator. ▲

Example 11.7. The energy radiated in one revolution by a proton accelerated in a cyclotron.

▼ The acceleration of the proton in a circular path of radius r is $a = \omega^2 r = 4\pi^2 \nu^2 r$, and relativistic effects may be neglected as long as the energy is small. Thus Eq. (11.36) with $q = e$ yields

$$\frac{dE}{dt} = \frac{e^2(4\pi^2\nu^2 r)^2}{6\pi\epsilon_0 c^3} = \frac{8\pi^3 e^2 \nu^4 r^2}{3\epsilon_0 c^3},$$

and the energy radiated in one revolution (the time for which is $1/\nu$) is

$$E_{\text{rad}} = \left(\frac{dE}{dt}\right)\frac{1}{\nu} = \frac{8\pi^3 e^2 \nu^3 r^2}{3\epsilon_0 c^3}.$$

This expression, of course, is not the total energy radiated by the proton since we have to add the energy radiated because of the acceleration experienced when the proton crosses the gaps; however, a simple calculation shows that this energy is, relatively speaking, much smaller than the "synchrotron" radiation. The maximum kinetic energy gained by the proton in each revolution is $E_k = 2eV_{max}$ since the proton crosses the dees' gap twice. Then

$$\frac{E_{rad}}{E_k} = \frac{8\pi^3 e^2 v^3 r^2}{3\epsilon_0 c^3 (2eV_{max})} = \frac{4\pi^3 e v^3 r^2}{3\epsilon_0 c^3 V_{max}}.$$

For a cyclotron with a radius of 0.92 m, an applied frequency of 1.5×10^7 Hz, and a peak value for the potential difference of 2×10^4 V, one finds that $E_{rad}/E_k = 4.0 \times 10^{-15}$. Here E_{rad} is still much smaller than E_k, but is relatively more important than in the previous example of the linear accelerator. ▲

Problems

11.1 Write the equations of the $\mathscr{E}$- and $\mathscr{B}$-fields, describing the following electromagnetic waves that propagate in the X-direction: (a) a linearly polarized wave whose plane of vibration lies at an angle of 45° with the XY-plane; (b) a linearly polarized wave whose plane of vibration lies at an angle of 120° with the XY-plane; (c) a wave with right-handed circular polarization; and (d) a wave with right-handed elliptical polarization, and with the major axis parallel to the Y-axis; the major axis is twice the minor axis.

11.2 (a) Describe the state of polarization of the waves represented by the following equations:

$$\text{(I)} \quad \mathscr{E}_y = A\cos\omega\left(t - \frac{x}{c}\right)$$

$$\mathscr{E}_z = A\sin\omega\left(t - \frac{x}{c}\right)$$

$$\text{(II)} \quad \mathscr{E}_y = A\cos\omega\left(t - \frac{x}{c}\right)$$

$$\mathscr{E}_z = -A\cos\omega\left(t - \frac{x}{c}\right)$$

$$\text{(III)} \quad \mathscr{E}_y = A\cos\omega\left(t - \frac{x}{c}\right)$$

$$\mathscr{E}_z = A\cos\left[\omega\left(t - \frac{x}{c}\right) - \frac{3\pi}{4}\right]$$

$$\text{(IV)} \quad \mathscr{E}_y = A\cos\omega\left(t - \frac{x}{c}\right)$$

$$\mathscr{E}_z = A\cos\left[\omega\left(t - \frac{x}{c}\right) + \frac{\pi}{4}\right]$$

(b) For each case represent the magnetic field showing how it changes as the wave progresses.

11.3 The electric field of a plane electromagnetic wave in vacuum is represented by

$$\mathscr{E}_x = 0,$$

$$\mathscr{E}_y = 0.5\cos\left[2\pi \times 10^8\left(t - \frac{x}{c}\right)\right] \text{NC}^{-1},$$

$$\mathscr{E}_z = 0.$$

(a) Determine the wavelength, the state of polarization, and the direction of propagation of the field. (b) Write down the magnetic field of the wave. (c) Compute the average intensity or energy flux per unit area.

11.4 Solve (a), (b), and (c) of Problem 11.3 for the wave represented by

$$\mathscr{E}_x = 0,$$

$$\mathscr{E}_y = 0.5\cos\left[4\pi \times 10^7\left(t - \frac{x}{c}\right)\right] \text{NC}^{-1},$$

$$\mathscr{E}_z = 0.5\sin\left[4\pi \times 10^7\left(t - \frac{x}{c}\right)\right] \text{NC}^{-1}.$$

11.5 Consider the wave represented by:

$$\mathscr{E}_y = \mathscr{E}_0 \cos 2\pi \left(\frac{t}{P} - \frac{x}{\lambda} \right),$$

$$\mathscr{E}_z = \mathscr{E}_0 \cos 2\pi \left(\frac{t}{P} - \frac{x}{\lambda} + \frac{1}{8} \right)$$

(a) Compute the magnitude of the electric vector and the angle formed by the electric vector with the Y-axis at the times (i) $t=0$ and (ii) $t=P/4$ and at the points $x=0$, $x=\lambda/4$, $x=\lambda/2$, $x=3\lambda/4$, $x=\lambda$. (b) For each case express the resultant magnetic field.

11.6 A plane sinusoidal linearly polarized light wave of wavelength $\lambda = 5.0 \times 10^{-7}$ m travels in vacuum. The average intensity is 0.1 W m^{-2}. The direction of propagation lies in the XY-plane at 45° with the X-axis. The electric field oscillates parallel to the Z-axis. Write the equations describing the electric and magnetic fields of this wave.

11.7 A plane sinusoidal linearly polarized electromagnetic wave of wavelength $\lambda = 5.0 \times 10^{-7}$ m travels in vacuum in the direction of the X-axis. The average intensity of the wave per unit area is 0.1 W m^{-2} and the plane of vibration of the electric field is parallel to the Y-axis. Write the equations describing the electric and magnetic fields of this wave.

11.8 The electric field of a plane electromagnetic wave has an amplitude of 10^{-2} NC^{-1}. Find (a) the magnitude of the magnetic field, and (b) the energy per unit volume of the wave. (c) If the wave is completely absorbed when it falls on a body, determine the radiation pressure. (d) Repeat the previous question if the body is a perfect reflector.

11.9 Electromagnetic radiation from the sun falls on the earth's surface at the rate of 1.4×10^3 W m^{-2}. Assuming that this radiation can be considered as a plane wave, calculate the magnitude of the electric and magnetic field amplitudes in the wave.

11.10 Assume that a 100-W lamp of 80% efficiency radiates all its energy isotropically. Compute the amplitude of the electric and magnetic fields 2 m from the lamp.

11.11 Radio waves received by a radio set have an electric field of maximum amplitude equal to 10^{-1} NC^{-1}. Assuming that the wave can be considered as plane, calculate (a) the amplitude of the magnetic field, (b) the average intensity of the wave, (c) the average energy density. (d) Assuming that the radio set is 1 km from the broadcasting station and that the station radiates energy isotropically, determine the power of the station.

11.12 Two harmonic electromagnetic waves, both of frequency ν and amplitude $\mathscr{E}_0$, travel in vacuum in the directions of the X-axis and Y-axis, respectively. The electric fields of both waves are parallel to the Z-axis. For the wave resulting from their superposition, compute (a) the components of the electric field $\mathscr{E}$, (b) the components of the magnetic field $\mathscr{B}$, (c) the energy density E, and (d) the components of the Poynting vector.

11.13 Show that the average value of the Poynting vector of a plane harmonic wave is $\frac{1}{2}c\epsilon_0\mathscr{E}_0^2$ or $\mathscr{E}_0\mathscr{B}_0/2\mu_0$. Compare with Eq. (11.17).

11.14 Show that if a system of oscillating charges radiates electromagnetic energy isotropically, the average value of the Poynting vector at a distance r is

$$\frac{1}{4\pi r^2}\left(\frac{dE}{dt}\right)_{\text{ave}}.$$

11.15 A system of oscillating charges concentrated around a point radiates energy at the rate of 10^4 W. Assuming that the energy is radiated isotropically, for a point at a distance of 1 m find (a) the average value of the Poynting vector, (b) the amplitude of the electric and magnetic fields, and (c) the energy and momentum densities. (*Hint:* Note that at large distances from the source, a small portion of the wave front can be considered as plane.)

11.16 A gaseous source emits light of wavelength 5×10^{-7} m. Assume that each molecule acts as an oscillator of charge e and amplitude 10^{-10} m. (a) Compute the average rate of energy radiation per molecule. (b) If the total rate of energy radiation of the source is 1 W, how many molecules are emitting simultaneously?

11.17 Estimate the value of $(dE/dt)_{ave}$ as given by Eq. (11.25) for a proton in a nucleus. Assume that z_0 is of the order of 10^{-15} m and ν about 5×10^{20} Hz for low-energy gamma rays.

11.18 Obtain an expression for the rate of energy radiated by a charged particle moving with velocity $\boldsymbol{v}$ perpendicular to a magnetic field $\mathscr{B}$.

11.19 (a) The electron in a hydrogen atom has a kinetic energy of 13.6 eV and a radius of 5.3×10^{-11} m. Assuming that the theory of Example 11.7 can be applied, calculate the energy radiated per second and per revolution. (b) Repeat the problem for a 50-keV electron on a 1-m circular path. (c) Repeat for a 50-keV proton on a 1-m circular path.

CHALLENGING PROBLEMS

11.20 The average power of a broadcasting station is 10^5 W. Assume that the power is radiated uniformly over any hemisphere concentric with the station. For a point 10 km from the source, find the magnitude of the Poynting vector and the amplitudes of the electric and magnetic fields. Assume that at that distance the wave is plane.

11.21 A radar transmitter emits its energy within a cone having a solid angle of 10^{-2} sterad. At a distance of 10^3 m from the transmitter the electric field has an amplitude of 10 NC^{-1}. Find the amplitude of the magnetic field and the power of the transmitter.

11.22 Show that if

$$\boldsymbol{V} = \boldsymbol{V}_0 \sin(\boldsymbol{k} \cdot \boldsymbol{r} - \omega t),$$

then the condition div $\boldsymbol{V} = 0$ implies that $\boldsymbol{k} \cdot \boldsymbol{V}_0 = 0$ or that $\boldsymbol{k}$ is perpendicular to $\boldsymbol{V}_0$. This proves, according to Eqs. (8.34), that in vacuum both $\mathscr{E}$ and $\mathscr{B}$ are perpendicular to $\boldsymbol{k}$, and the results of Section 11.2 are of general validity.

11.23 (a) Show that if

$$\boldsymbol{V} = \boldsymbol{V}_0 \sin(\boldsymbol{k} \cdot \boldsymbol{r} - \omega t),$$

then

$$\text{curl } \boldsymbol{V} = \boldsymbol{k} \times \boldsymbol{V}_0 \cos(\boldsymbol{k} \cdot \boldsymbol{r} - \omega t)$$

and Eqs. (8.34) imply that

$$\boldsymbol{k} \times \mathscr{B} = -\mu_0 \epsilon_0 \omega \mathscr{E}$$

and

$$\boldsymbol{k} \times \mathscr{E} = \omega \mathscr{B}.$$

(b) Show that the two results are compatible. (c) From the results of this and the preceding problem, discuss the relative orientation of the vectors $\boldsymbol{k}$, $\mathscr{E}$, and $\mathscr{B}$. (d) Compare with the results of Section 11.2.

11.24 Using the results of the previous problems, show that the $\mathscr{E}$ and $\mathscr{B}$ fields of a plane electromagnetic wave must be in phase.

11.25 Show that the Poynting vector may be written as $\mathscr{E} \times \mathscr{H}$. This expression is applicable to an electromagnetic wave propagating either in vacuum or in a material medium.

11.26 Compute the energy flux per unit area through a plane perpendicular to the velocity of a charge moving with constant velocity and passing through the charge. Assume that the charge has a radius R and use the nonrelativistic expression of the electric and magnetic fields. Discuss your result critically. (*Hint:* Use rings of radius r and width dr concentric with the charge as area elements for the flux.)

11.27 Expression (11.37) gives the intensity of the radiation from an accelerated charge in terms of the direction of the radiation. Obtain Eq. (11.36) from it by integration over all directions. (*Hint:* Multiply $I(\theta)$ by the area element $dS = 2\pi r^2 \sin\theta \, d\theta$, and integrate from 0 to π.)

11.28 Show that for a particle moving in a linear accelerator the rate of energy radiation is

$$\left(\frac{dE}{dt}\right)_{\text{rad}} = \left(\frac{q^2}{6\pi\epsilon_0 m_0^2 c^3}\right)\left(\frac{dE_k}{dx}\right)^2$$

where E_k is the kinetic energy of the particle.

11.29 Show that the rate of energy radiation in a circular accelerator is

$$\left(\frac{dE}{dt}\right)_{\text{rad}} = \left(\frac{q^2 c}{6\pi\epsilon_0 r^2}\right)\left(\frac{v}{c}\right)^4\left(\frac{E}{m_0 c^2}\right)^4.$$

CHAPTER TWELVE

INTERACTION OF ELECTROMAGNETIC RADIATION WITH MATTER

12.1 Introduction

In nature there is a continuous exchange of energy between atoms, molecules, and electromagnetic radiation. The sun is the main source of the electromagnetic radiation that reaches the earth. The interaction of electromagnetic radiation from the sun and the bodies on the earth's surface accounts for most of the phenomena observed daily, including life itself.

In the previous chapter we showed that electromagnetic radiation may be produced by oscillating electric and/or magnetic dipoles and that the electromagnetic field propagates through vacuum with the speed of light, c. In this chapter we will examine the effect matter has upon electromagnetic radiation and the effect of electromagnetic radiation on matter. In particular we will introduce the concept of the photon as the interaction carrier or "particle" that is needed to interpret the way matter and electromagnetic waves interact.

12.2 Absorption of Electromagnetic Radiation

The most important radiative mechanisms by which an electromagnetic wave can be produced were outlined in Chapter 11. We now analyze the reverse process and see what happens when an electromagnetic wave interacts with an atom or a system of charges so that energy from the wave is absorbed by the system. The absorption of energy from an electromagnetic wave is a complicated problem that requires extensive mathematical calculations and the use of quantum mechanics, but the fundamental ideas are easy to understand. When an electromagnetic wave impinges on an atom, both the electric and the magnetic fields of the wave interact with the electrons in the atom. The effect of the magnetic field can be neglected in a first approximation because the magnetic force is of the order of magnitude of $ev\mathscr{B}_0$, which may be written as $(v/c)e\mathscr{E}_0$ by using the relation $\mathscr{B}_0 = \mathscr{E}_0/c$. Since $e\mathscr{E}_0$ is the force of the electric field, the two forces are comparable only when $v \approx c$.

In a region of space small compared with the wavelength of the radiation (as within an atom), the electric field of the electromagnetic wave can be written as $\mathscr{E} = \mathscr{E}_0 \sin \omega t$ since the x part of the wave expression is practically constant over the dimensions of the small region where the electron moves. The electron performs forced oscillations under the influence of the electric force $-e\mathscr{E}$. Recalling the discussion of Section 12.13 of Volume I, we may see that the rate at which energy is absorbed by the electron (that is, the average power transferred to the oscillator by the wave's electric field) is maximum at energy resonance, which occurs when the frequency of the wave is equal to the natural frequency of the electron. A more detailed quantum-

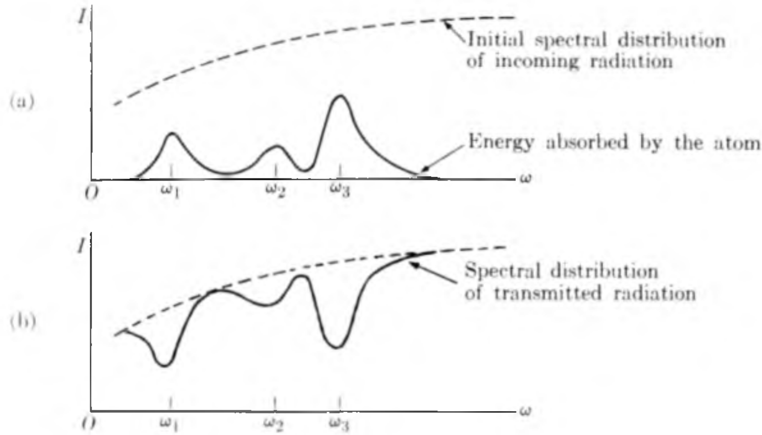


Fig. 12-1. Intensities of absorbed and transmitted radiation passing through a substance.

mechanical analysis, omitted here, shows that this frequency is any of the frequencies $\omega_1, \omega_2, \omega_3, \ldots$ of the emission spectrum of the atom (or molecule) to which the electron is bound. In other words

> *an atom or molecule absorbs electromagnetic radiation preferentially when the frequency of the electromagnetic wave coincides with one of the frequencies of the emission spectrum of the atom or molecule;*

or in a more synthetic form, the emission and absorption spectra of a substance are composed of the same frequencies. The intensity of absorption is greatest for those frequencies that involve a transition from the ground state since atoms are ordinarily in the ground state.

Figure 12-1(a) shows the intensity distribution of an incident wave and the energy absorbed by a substance as functions of the frequency. Figure 12-1(b) shows the intensity distribution of the transmitted radiation. Note the correspondence between the two curves since the transmitted radiation is depleted in the frequencies corresponding to favored absorption by the atom.

What is the result of energy absorption by the atom (or molecule)? This energy absorption results in an adjustment of the electronic motion to correspond to the new higher energy state of the atom (or molecule). The atom (or molecule) is then said to be in an *excited state.* An excited atom (or molecule) may in turn, by means of electric dipole radiation, re-emit the excess energy it has just absorbed. These are the processes we will discuss in this chapter.

12.3 Scattering of Electromagnetic Radiation by Bound Electrons

When an electromagnetic wave passes through an atom (or molecule), the wave disturbs the motion of the bound electrons; and the atom (or molecule) may be left in an excited state. By the reciprocal process since the electrons act as forced oscillating electric dipoles, the excited atom may emit electromagnetic radiation of the same frequency as the incident wave without an appreciable time delay. The energy the atom emits has been absorbed from the incident wave by the atom's bound electrons. This process is called *scattering*, and the radiation emitted is the *scattered wave* (Fig. 12-2).

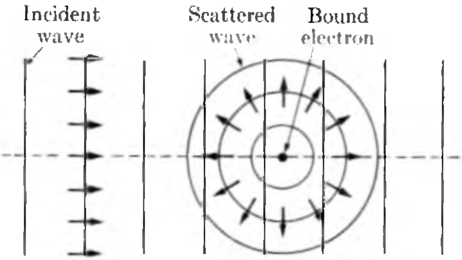


Fig. 12-2. Schematic description of scattering of radiation by a bound electron.

Scattering helps to decrease the intensity of the primary or incident wave because the energy absorbed from the wave is re-emitted in all directions, and the result is an effective removal of energy from the primary radiation.

It has been observed experimentally that the intensity of the scattered waves depends on the frequency of the primary wave and on the angle of scattering. To calculate this dependence, the extent of the perturbation of the motion of the atomic electrons by the electric field of the primary wave must be determined. This analysis must be accomplished by means of quantum mechanics.

One important feature is that *the scattered waves are more intense when the frequency of the incident radiation is equal to one of the frequencies $\omega_1, \omega_2, \omega_3, \ldots$ of the emission spectrum of the atom (or molecule)*, a result that is known as *resonant fluorescence*.* This physical behavior should be expected since the intensity of the scattered radiation should be greater at those frequencies at which the energy absorption from the wave is greater, and these are the same frequencies as the emission spectrum of

*In the visible region of the electromagnetic spectrum, the luminescence induced in a substance by radiation absorption and subsequent emission is called *fluorescence* when the time delay between absorption and emission is less than 10^{-8} s. When the delay time is longer, the phenomenon is called *phosphorescence*. Fluorescent and phosphorescent radiation are generally of different frequencies.

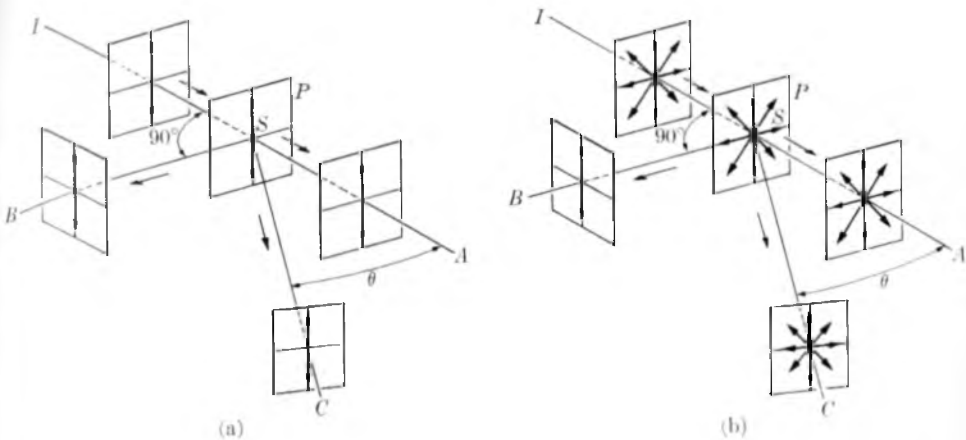


Fig. 12-3. Polarization of scattered radiation. (a) Linearly polarized and (b) unpolarized incident radiation.

the atom (as explained in Section 12.2). However even at frequencies different from those of the emission spectrum, scattering may still be appreciable.

Another interesting property of the scattering phenomenon is that for gases whose molecules have an emission spectrum in the ultraviolet region (see Section 12.9) the scattering of electromagnetic waves falling in the visible region increases with their frequency. This property is easy to understand since the larger the frequency in the visible region, the closer the frequency is to the ultraviolet resonant frequency of the molecule and the larger the amplitude of the forced oscillations; and greater scattering results. As an illustration, the brightness and blue color of the sky are attributed to the scattering of sunlight by air molecules; in particular the blue color of the sky is the result of the more intense scattering of the higher frequencies (or shorter wavelengths). The same process accounts for the bright red colors observed at sunrise and at sunset when the sun's rays traverse a very large thickness of air before reaching the earth's surface; the result is a strong attenuation of the high frequencies (or short wavelengths) by the scattering process.

Scattering can also be produced by small particles (such as smoke or dust) or water droplets (such as clouds) suspended in the air. Liquids carrying a suspension of particles, as in a colloid, show a strong scattering called the *Tyndall effect*.

When the primary radiation is linearly polarized, the atomic oscillations are in the fixed direction of the electric field of the wave, and the scattered radiation has the polarization characteristic of the electric dipole radiation (Fig. 12-3a). However even if the incident radiation is not polarized, the scattered radiation is always partially polarized. Consider, for example, an unpolarized wave incident on an atom S (Fig. 12-3b). The electric dipole oscillations induced in the atom are parallel to the electric field of the wave, and therefore are all in a plane P perpendicular to the direction of propagation IA of the incident wave. The polarization of the scattered radiation in each direction depends on the direction of the dipole oscillation, and therefore is not always fixed when the incident wave is unpolarized. However for any direction SB perpendicular to IS, the scattered radiation is linearly polarized parallel to the plane P, perpendicular to IS, since for these directions the dipoles always oscillate in such a plane. For other directions the degree of polarization of the scattered radiation depends on the angle that the direction of scattering makes with IA. If the incident radiation is unpolarized, the scattered radiation is completely unpolarized along IA.

12.4 Scattering of Electromagnetic Radiation by a Free Electron; Compton Effect

The scattering of electromagnetic radiation by a free electron has certain peculiarities that require a discussion apart from that of scattering by bound electrons or molecules. As we have just seen, scattering is a double process by which an electron in an atom quickly absorbs energy from an electromagnetic wave and quickly reradiates that energy as scattered radiation. Keep in mind that an electromagnetic wave carries energy and momentum; and if some energy E is removed from the wave, a corresponding amount of momentum $p=E/c$ must also be removed from the wave.

A free electron cannot absorb an amount of energy E and at the same time acquire a momentum $p_e=E/c$ because the relation between kinetic energy and momentum for an electron is

$$E_k=E-m_e c^2=\sqrt{(m_e c^2)^2+(p_e c)^2}-m_e c^2$$

at high energies and $E_k=p_e^2/2m_e$ at lower energies. Either of these expressions is incompatible with the relation $p=E/c$ if we have $E=E_k$ as required by the conservation of energy. We should conclude then that a free electron cannot absorb electromagnetic energy without violating the principle of conservation of momentum. The student may wonder why this problem of momentum was not mentioned at all in the previous section when scattering and absorption of electromagnetic waves by bound electrons were discussed. The reason is that although conservation of momentum and energy applies in both cases, in the case of a *bound* electron the energy and momentum absorbed are shared by both the electron and the ion forming the remaining part of the atom. It is always possible to split both energy and momentum in the correct proportions; however, the ion, having a much larger mass, carries (along with some momentum) only a small fraction of the energy available, and usually it is not necessary to consider that fraction at all (see Example 9.13 in Volume I). In the case of a *free* electron, there is no other particle with which the electron can share the energy and the momentum, and no absorption or scattering should be possible.

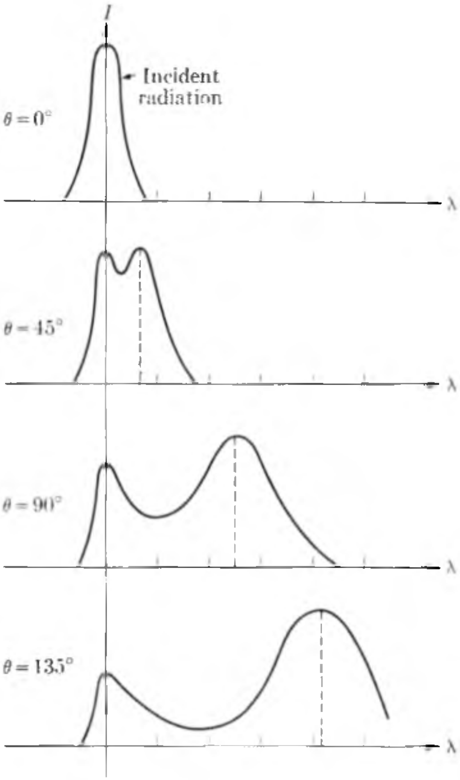


Fig. 12-4. Intensity distribution of the radiation scattered by a free electron at different scattering angles.

Experiment, however, tells a different story. When monochromatic electromagnetic radiation passes through a region in which essentially free electrons are present (for instance a metallic solid contains many electrons not bound to a given atom), we observe that in addition to the incident radiation, there is another sort of radiation of frequency *different* from that of the incident radiation. This new radiation is interpreted as the radiation scattered by the free electrons. The frequency of the scattered radiation is *smaller* than the incident frequency, and accordingly the wavelength of the scattered radiation is *longer* than the incident wavelength (Fig. 12-4). The wavelength of the scattered radiation is also different for each direction of scattering. This interesting phenomenon is called the *Compton effect*, after the American physicist A. H. Compton (1892–1962), who first observed and analyzed it in the early 1920s.

Given that λ is the wavelength of the incident radiation and λ' that of the scattered radiation, Compton found that λ' is determined solely by the direction of the scattering. That is, if θ is the angle between the incident radiation and the direction in which the scattered radiation is observed (Fig. 12-5), the wavelength of the scattered radiation

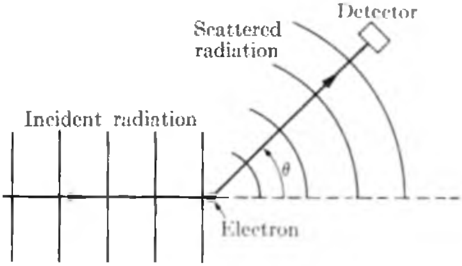


Fig. 12-5. Geometry in Compton scattering.

λ' is determined solely by the angle θ. The experimental relation is

$$\lambda' - \lambda = \lambda_C(1 - \cos\theta) \tag{12.1}$$

where λ_C is a constant whose value if λ and λ' are measured in meters, is

$$\lambda_C = 2.4262 \times 10^{-12}\ \text{m}.$$

This constant is called the *Compton wavelength for electrons.*

Remembering that $\lambda = c/\nu$ where ν is the frequency $\omega/2\pi$ of the wave, we may write Eq. (12.1) in the form

$$\frac{1}{\nu'} - \frac{1}{\nu} = \frac{\lambda_C}{c}(1 - \cos\theta). \tag{12.2}$$

The scattering of an electromagnetic wave by an electron may be visualized as a "collision" between the wave and the electron since the scattering comprises an exchange of energy and momentum. Since the wave propagates with the velocity c, and its energy–momentum relationship as given by Eq. (11.19) is $E = cp$ (which is identical to that for a particle of zero rest mass), this scattering must resemble a collision in which one of the particles has zero rest mass and is moving with velocity c. Such a collision was discussed in Example 11.8 in Volume I and is reproduced at the end of this section as Example 12.1. The result is

$$\frac{1}{E'} - \frac{1}{E} = \frac{1}{m_e c^2}(1 - \cos\theta) \tag{12.3}$$

where E and E' are the energies of the particle of zero rest mass before and after the collision, and m_e is the rest mass of the other particle involved in the collision, in this case an electron. The similarity between Eqs. (12.2) and (12.3) is striking and goes beyond a simple algebraic similarity. Both equations apply to a collision process in its most general sense; and as already stated, the energy–momentum relationship $E = cp$ for an electromagnetic wave is of the same type as that relationship corresponding to a particle of zero rest mass to which Eq. (12.3) applies. The obvious conclusion is to link the frequency ν and the energy E by writing

$$E = h\nu \tag{12.4}$$

where h is a universal constant that describes the proportionality between the frequency of an electromagnetic wave and the energy associated with it in the "collision" process. Then Eq. (12.3) becomes

$$\frac{1}{h\nu'}-\frac{1}{h\nu}=\frac{1}{m_ec^2}(1-\cos\theta)$$

or

$$\frac{1}{\nu'}-\frac{1}{\nu}=\frac{h}{m_ec^2}(1-\cos\theta), \qquad (12.5)$$

which is identical to Eq. (12.2). To obtain the equivalent of Eq. (12.1), multiply Eq. (12.5) by c and use $\lambda=c/\nu$. The result is

$$\lambda'-\lambda=\left(\frac{h}{m_ec}\right)(1-\cos\theta). \qquad (12.6)$$

Then the Compton wavelength for an electron λ_C is related to the mass of the scattering electron by

$$\lambda_C=\frac{h}{m_ec}. \qquad (12.7)$$

From the known values of λ_C, m_e, and c, we obtain the constant h, whose value is

$$h=6.6256\times10^{-34}\ \mathrm{J\ s} \qquad \text{or} \qquad \mathrm{m^2\ kg\ s^{-1}};$$

h is called *Planck's constant*. Planck's constant plays a very important role in physics. Historically the constant first appeared in a different context: the constant was introduced at the end of the nineteenth century by the German physicist Max Planck (1858–1947) as a result of his attempt to explain the intensity of *blackbody radiation*, the electromagnetic radiation in equilibrium with matter. We also referred to Planck's constant in the Section 7.4 discussion of Bohr's theory of the atom. Bohr's work preceded the discovery of the Compton effect. The speed of light c, the fundamental charge e, the electron mass m_e, and Planck's constant h constitute four fundamental constants of physics.

A proton, which has a mass different from that of an electron, has a Compton wavelength (using the value of h above) of

$$\lambda_{C,p}=h/m_pc=1.3214\times10^{-15}\ \mathrm{m}.$$

That this result has been experimentally confirmed ensures the validity of assumption (12.4). However because the Compton wavelength of the proton is 10^{-3} that of the electron, Compton radiation from scattered protons is much less noticeable than that from electrons.

Example 12.1. Discussion of a relativistic collision when the incident particle has zero rest mass and the target particle is an electron at rest in our laboratory system.

▼ The process is shown schematically in Fig. 12-6. The particles are labeled 1 and 2 before the collision, and 3 and 4, respectively, after the collision. Using Eqs. (11.15) and (11.18) of Volume I (see also the appendix), we obtain the value of the momentum and the energy relative to observer O:

$$p_1 = E/c, \qquad p_2 = 0, \qquad p_3 = E'/c, \qquad p_4,$$
$$E_1 = E, \qquad E_2 = m_e c^2, \qquad E_3 = E', \qquad E_4 = c\sqrt{m_e^2 c^2 + p_4^2}.$$

The conservation of momentum allows us to write

$$\boldsymbol{p}_1 = \boldsymbol{p}_3 + \boldsymbol{p}_4,$$

and the conservation of energy for this process yields

$$E + m_e c^2 = E' + c\sqrt{m_e^2 c^2 + p_4^2}.$$

Suppose we are interested in the energy E' of the incident particle after the collision. We must then eliminate $\boldsymbol{p}_4$ from the equations above. From the momentum equation we obtain $\boldsymbol{p}_4 = \boldsymbol{p}_1 - \boldsymbol{p}_3$. Squaring the result gives

$$p_4^2 = p_1^2 + p_3^2 - 2\boldsymbol{p}_1 \cdot \boldsymbol{p}_3.$$

Using the corresponding values for the momenta yields

$$p_4^2 = \frac{E^2}{c^2} + \frac{E'^2}{c^2} - \frac{2EE'}{c^2}\cos\theta.$$

Solving the energy equation above for p_4^2 gives

$$p_4^2 = \frac{1}{c^2}(E + m_e c^2 - E')^2 - m_e^2 c^2$$
$$= \frac{E^2}{c^2} + \frac{E'^2}{c^2} + \frac{2(E - E')m_e c^2}{c^2} - \frac{2EE'}{c^2}.$$

Equating both results for p_4^2, we have

$$\frac{2(E - E')m_e c^2}{c^2} - \frac{2EE'}{c^2} = -\frac{2EE'}{c^2}\cos\theta$$

or

$$E - E' = \frac{EE'}{m_e c^2}(1 - \cos\theta).$$

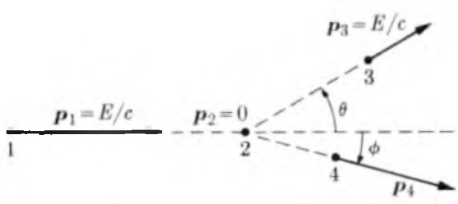


Fig. 12-6. High-energy collison.

Dividing both sides by EE' yields

$$\frac{1}{E'}-\frac{1}{E}=\frac{1}{m_e c^2}(1-\cos\theta).$$

This expression gives E' in terms of E and the scattering angle θ of particle 3. Note that $E \geqslant E'$ always, and the incident particle loses energy as it should since the other particle, initially at rest, is in motion after the collision.

We may then conclude that we can "explain" the scattering of electromagnetic radiation by a free electron by identifying the process with the collision of a free electron and a particle of zero rest mass. ▲

12.5 Photons

Our "explanation" of the Compton effect requires a careful analysis because of its possible, far-reaching consequences. First, let us recapitulate our assumptions.

1. The scattering of electromagnetic radiation by a free electron can be considered as a collision between the electron and a particle of zero rest mass.
2. Electromagnetic radiation plays the role of the particle of zero rest mass, which for brevity will be called a *photon*.
3. The energy and the momentum of the photon (particle of zero rest mass) are related to the frequency and the wavelength of the electromagnetic radiation by

$$E=h\nu, \qquad p=\frac{h}{\lambda}. \tag{12.8}$$

The second relation is due to the fact that $p=E/c=h\nu/c$ and $\nu/c=1/\lambda$. We may visualize the Compton effect as the collision illustrated in Fig. 12-7, in which a photon of frequency ν collides with an electron at rest. The photon transfers a certain amount of energy and momentum to the electron. As a result of the interaction, the energy of the scattered photon is smaller, with a correspondingly smaller frequency ν'.

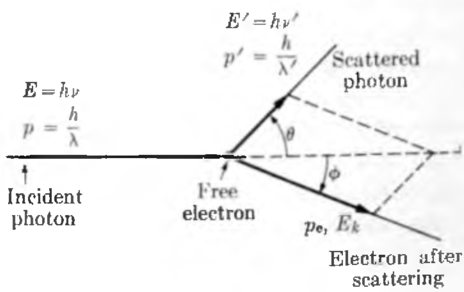


Fig. 12-7. Momentum and energy relations in Compton scattering.

One further test is to see whether the electron after the scattering has a momentum equal to the difference between the momentum of the incident photon and that of the scattered photon. This experiment is difficult, but it has been performed and the results check very well.

What physical meaning can be attributed to the photon concept and to the defining relations (12.8)? It is not a necessary conclusion that electromagnetic radiation is a stream of photons even though that is a possible pictorial explanation. We could interpret the photon energy $E=h\nu$ and the momentum $p=h/\lambda$ as the energy and momentum absorbed by the free electron from the incident electromagnetic wave. The photon of energy $E'=h\nu'$ and the momentum $p'=h/\lambda'$ are then the energy and momentum re-emitted by the electron into the scattered radiation. The electron acquires a kinetic energy $E_k=E-E'$ and a momentum $\boldsymbol{p}_e=\boldsymbol{p}-\boldsymbol{p}'$, related by $E_k=c\sqrt{m_ec^2+p_e^2}-m_ec^2$ as required by high-energy dynamics and by the conservation of energy and momentum. Therefore we may come to the interpretation that the photon is the "quantum" of electromagnetic energy and momentum absorbed or emitted in a single process by a charged particle. The concept of a photon is in fact only applicable to interactions between electromagnetic radiation and charged particles; the concept plays a role in all processes in which electromagnetic radiation interacts with matter, and not just with free electrons. Therefore we state the following principle:

> *When an electromagnetic wave interacts with an electron (or any other charged particle), the amounts of energy and momentum that can be exchanged in the process are those corresponding to the energy and momentum of a photon.*

This principle is one of the fundamental laws of physics and is characteristic of all radiative processes involving charged particles and electromagnetic fields. It does not stem from any law previously stated or discussed, but is a completely new principle to be added to such universal laws as the conservation of energy and momentum. The discovery of this new principle in the first quarter of this century was a milestone in the development of physics. The consequences of this principle have given rise to the branch of physics called *quantum mechanics.*

This important principle is basic to the understanding of emission and absorption of electromagnetic radiation in atoms, molecules, and nuclei. We have mentioned that an atom (or molecule) can emit or absorb electromagnetic radiation of only certain frequencies. Also we indicated in Section 7.4 that the energy of atoms (and molecules) is quantized and can have only certain values corresponding to so-called *stationary states* or *energy levels.* These two important facts are related through the concept of the photon. Suppose that an atom in the stationary state of energy E absorbs electromagnetic radiation of frequency ν and passes to another stationary state of higher energy E'. The change in energy of the atom is $E'-E$. On the other hand the energy of the photon absorbed is $h\nu$. Conservation of energy requires that both quantities be equal. Therefore

$$E'-E=h\nu, \tag{12.9}$$

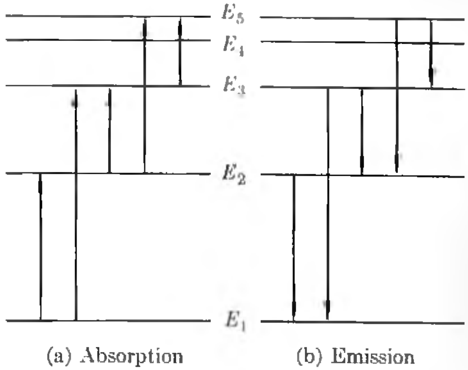


Fig. 12.8. Transitions between stationary states in an atom, molecule, or nucleus. The relative spacing of the energy levels and the possible transitions depend on the nature of the system.

an expression known as *Bohr's formula* because it was first proposed in 1913 by the Danish physicist Niels Bohr (1885–1962). The expression above also applies to the energy of a photon emitted when an atom passes from a stationary state of energy E' to another of lower energy E.

Since the energy of the stationary state is quantized and can have only certain values E_1, E_2, $E_3, \ldots$, Bohr's formula limits the frequency of the radiation emitted or absorbed. Thus a discrete spectrum of frequencies results. Historically Bohr proposed the concept of stationary states to explain the existence of a discrete spectrum of frequencies. Figure 12-8 is a schematic diagram listing some of the possible changes in energy of a system. They correspond to transitions between stationary states or energy levels of an atom. For example the energies of the stationary state of atoms with only one electron (H, He^+, Li^{++}, etc.) is given by (recall Section 7.4)

$$E = -\frac{RZ^2hc}{n^2}$$

where R is Rydberg's constant, whose value in SI is $1.0974 \times 10^7\ m^{-1}$. Therefore in a transition between states with quantum numbers n and $n'(n' > n)$, the frequency of the radiation emitted or absorbed is

$$\nu = \frac{E' - E}{h} = RZ^2c\left(\frac{1}{n^2} - \frac{1}{n'^2}\right)$$

or with numerical values introduced

$$\nu = 3.2899 \times 10^{15} Z^2 \left(\frac{1}{n^2} - \frac{1}{n'^2}\right)\ \text{Hz}.$$

This formula fits the emission and absorption frequencies of the spectra of this kind of atom quite well. The formula is called *Balmer's formula*, after the Swiss mathematician and physicist Johann Balmer (1825–1898), who determined it empirically in attempting to catalog the visible spectral lines of hydrogen.

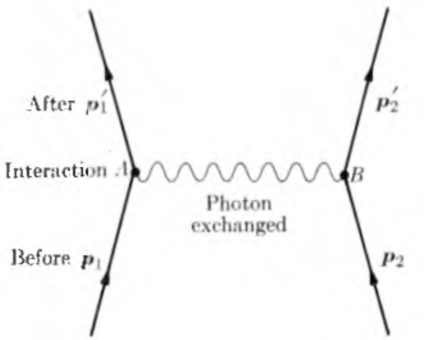


Fig. 12-9. Electromagnetic interaction considered as an exchange of photons. The photons transfer energy and momentum from one charge to the other.

The concept of the photon suggests the simple pictorial representation of the electromagnetic interaction between two charged particles shown in Fig. 12-9. The interaction corresponds to an exchange of momentum and energy. The particles' initial momenta $\boldsymbol{p}_1$ and $\boldsymbol{p}_2$ become $\boldsymbol{p}_1'$ and $\boldsymbol{p}_2'$ after the interaction. Although not localized at a particular instant, the interaction has for simplicity been indicated at a particular time and at positions A and B. Particle 1 interacts with particle 2 via its electromagnetic field, with the result that particle 2 takes from the field certain energy and momentum, equivalent to a photon, with a corresponding change in particle 2's motion. The motion of particle 1 must then be adjusted to correspond to the new field, which is the original field minus one photon. Of course the reverse process is also possible, and particle 1 may absorb a photon from the field of particle 2. We may say then that there has been an exchange of photons between particle 1 and particle 2. In other words

> *electromagnetic interactions can be pictured as being the result of the exchange of photons between the interacting charged particles: and a change in the energy and momentum of both particles results.*

[Diagrams such as Fig. 12-9 are called *Feynman diagrams* to honor their originator, the American scientist Richard P. Feynman (1918–). They are very useful in analyzing complex processes involving different kinds of particle-particle interactions.]

At any instant the total momentum of a system of two charged particles is $\boldsymbol{p}_1 + \boldsymbol{p}_2 + \boldsymbol{p}_{\text{field}}$ where $\boldsymbol{p}_{\text{field}}$ is the momentum associated with the electromagnetic field of the charged particles. This situation corresponds to the picture of an interaction described in Section 4.8 where the principle of conservation of momentum was analyzed when an interaction propagates with a finite velocity. Now that conceptual picture of a field possessing energy and momentum has a firmer theoretical and experimental basis.

At the end of Section 11.3 we indicated that electromagnetic radiation carries intrinsic angular momentum or spin in addition to energy and momentum, and that for circularly polarized waves the spin along the direction of propagation is $\mp E/\omega$. Using the relation $\omega = 2\pi\nu$, we see that the energy of a photon is $E = h\omega/2\pi =$

$\hbar\omega$ where $\hbar = h/2\pi$. Thus circularly polarized photons have a spin along the direction of propagation equal to $\mp\hbar$.

Example 12.2. Energy of a photon expressed in electron volts when its wavelength is given in meters. The wavelength of X-rays in terms of the accelerating voltage applied to an X-ray tube.

▼ From $E = h\nu$ and $\lambda\nu = c$, we may write $E = hc/\lambda$. But

$$hc = (6.6256 \times 10^{-34}\text{ J s})(2.9979 \times 10^{8}\text{ m s}^{-1})$$
$$= 1.9863 \times 10^{-25}\text{ J m.}$$

Remembering that 1 eV $= 1.6021 \times 10^{-19}$ J, we have that $hc = 1.2398 \times 10^{-6}$ eV m. Therefore

$$E = 1.2398 \times 10^{-6}/\lambda$$

where E is expressed in electron volts when λ is expressed in meters.

As explained in connection with Fig. 11-16, X-rays are produced by the impact of fast electrons against the anticathode (anode) of an X-ray tube. The energy of an electron may be radiated as a result of successive collisions, giving rise to several photons, or the energy may all be radiated in just one collision. The most energetic photons coming out of the X-ray tube would be those emitted in the latter process, and they correspond to the shortest wavelength. In other words given that V is the accelerating voltage, the wavelengths of the X-rays produced are equal to or larger than the threshold wavelength, satisfying the relation

$$\lambda_0 = \frac{1.2398 \times 10^{-6}}{V} \approx \frac{1.24 \times 10^{-6}}{V}\text{ m}$$

since in this case the energy E of the photon is equal to the electron's energy, which in turn is equal to V expressed in volts. For example, in a television tube, electrons are accelerated by a potential difference of the order of 18,000 V. When the electrons reach the screen of the tube, they are abruptly stopped and emit X-rays for the same reason as in an X-ray tube. The minimum wavelength of the X-rays produced when the electrons are stopped at the screen is then $\lambda = 6.9 \times 10^{-11}$ m. The intensity, however, is quite low. ▲

12.6 More about Photons: The Photoelectric Effect

Further research has shown that the concept of the photon applies not only to the process of scattering by a free electron, but to *all* processes in which electromagnetic waves interact with matter. Another example that illustrates the use of the photon concept is the *photoelectric effect*. In 1887 the German physicist Heinrich Hertz (1857–1894) observed that the intensity of the electric discharge between two charged electrodes could be increased if the electrodes were illuminated with ultraviolet radiation. This observation suggested the availability of more charged particles or electrons in the presence of light. A year later another German physicist, Wilhelm Hallwachs (1859–1922), observed an electronic emission from the freshly cleaned surfaces of certain metals such as Zn, Rb, K, Na, etc., when they were illuminated. These electrons are called *photoelectrons* because of the method of their production. The electronic emission increases with the intensity of the radiation falling on the metal surface since more energy is available to release electrons. However for each substance there is a minimum frequency ν_0 of electromagnetic radiation such that for radiation of frequency less than ν_0, no photoelectrons are produced, no matter how intense the radiation may be.

We have explained before that in a metal there are electrons that are more or less free to move throughout the crystal lattice. These electrons do not escape from the metal at normal temperature because if one escapes, the electrical balance of the metal is destroyed and the metal becomes positively charged, attracting the electron. Unless the electron has enough energy to overcome this attraction, the electron will fall back on the metal. One way to increase the energy of the electrons is by heating the metal. The "evaporated" electrons are then called *thermoelectrons.* This kind of electronic emission exists in electron tubes. Another way to release electrons from a metal is by the absorption of energy from electromagnetic radiation and the consequent production of photoelectrons.

The *photoelectric effect* is a process by which conduction electrons in metals and in other substances absorb energy from the electromagnetic field and escape from the substance, in contrast to the absorption process discussed in Section 12.2, which corresponds to absorption by an electron bound to an atom or molecule. Designate the energy required by an electron to escape from a given metal by ϕ. Then, if the electron absorbs the energy E, the difference $E-\phi$ will appear as the kinetic energy of the electron; and we may write in the low-energy or nonrelativistic limit

$$\tfrac{1}{2}mv^2 = E-\phi. \tag{12.10}$$

If E is less than ϕ, no electronic emission will result. From the photon concept if E is the energy absorbed by an electron from the electromagnetic radiation and ν is the frequency of the radiation, then $E=h\nu$, according to Eq. (12.8). Thus we may write Eq. (12.10) as

$$\tfrac{1}{2}mv^2 = h\nu-\phi. \tag{12.11}$$

This equation was first proposed by Albert Einstein in 1905, prior to the discovery of the Compton effect, as a means for explaining the observed relation between the kinetic energy of the photoelectrons and the frequency of the incident radiation. Not all electrons require the same energy ϕ to escape from a metal. The minimum value ϕ_0 is called the *work function* of the metal. Then the maximum kinetic energy of the electrons is

$$\tfrac{1}{2}mv^2_{\max} = h\nu-\phi_0. \tag{12.12}$$

The maximum kinetic energy $\frac{1}{2}mv^2_{\max}$ can be measured by the method indicated in

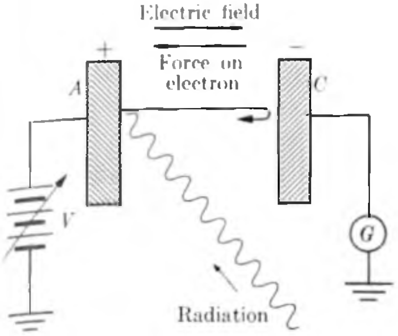


Fig. 12-10. Experimental arrangement for observing the photoelectric effect.

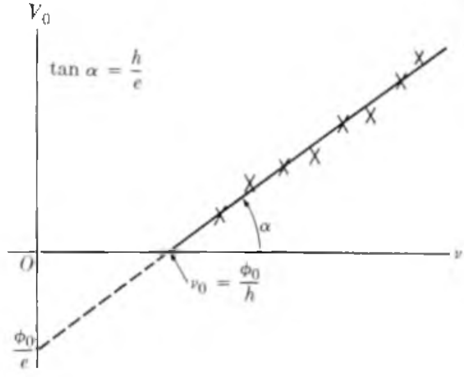


Fig. 12-11. Relation between stopping potential and frequency in the photoelectric effect.

Fig. 12-10. By applying a potential difference V between the plates A and C, we can retard the motion of the photoelectrons. At a particular voltage V_0 the current detected by the galvanometer becomes zero, indicating that no electrons, not even the fastest, are reaching plate C. Then by Eq. (1.28), $\frac{1}{2}mv_{max}^2 = eV_0$, and Eq. (12.12) becomes

$$eV_0 = h\nu - \phi_0. \tag{12.13}$$

Changing the frequency ν of the illuminating radiation yields a series of values for the stopping potential V_0. The result of plotting the values of V_0 against ν is a straight line as shown in Fig. 12-11. The slope of the straight line is $\tan\alpha = h/e$. Measuring α, and using the known value of e, we may calculate Planck's constant h and obtain the same result found for the Compton effect. This agreement can be considered as a further justification of the photon concept.

From Eq. (12.12), we see that for the frequency $\nu_0 = \phi_0/h$, the kinetic energy of the electrons is zero. Therefore ν_0 is the minimum frequency at which there is photoelectric emission, and is called the *threshold frequency*. For frequencies smaller than ν_0, there is no emission at all.

When electromagnetic radiation has sufficiently high frequency (or photons of sufficient energy), electrons may be ejected from atoms (or molecules) in what is called the *atomic photoelectric effect*. This process is responsible for most of the absorption of X- and γ-rays by any material. The atomic photoelectric effect results in a corresponding ionization of the material (including air) through which the X- and γ-rays pass, and is one of the mechanisms by which radiation affects matter. A similar process is the *photonuclear effect*, by which a particle, usually a proton, after it absorbs electromagnetic radiation, is ejected from a nucleus. These photons must have much more energy and a corresponding frequency much higher than the photons involved in the atomic photoelectric effect so that the former fall within the range of the higher-energy γ-rays.

Note that the conservation of momentum was not mentioned in the discussion of the photoelectric effect. The reason is again that the electron absorbing the electromagnetic radiation is bound to the crystal lattice of the solid, or to an atom or a molecule; the momentum of the absorbed photon is shared by the electron and the lattice, atom, or molecule. However because of the relatively large mass of the lattice, atom, or molecule, its kinetic energy is negligible; and one may assume without noticeable error that all the energy of the photon goes to the electron. The same analysis applies to the protons in the photonuclear effect.

12.7 Propagation of Electromagnetic Waves in Matter: Dispersion

So far only the propagation of electromagnetic waves in vacuum has been considered. Experiments reveal that the velocity of propagation of an electromagnetic wave through matter is different from the velocity of propagation in vacuum. To understand the reason for these different velocities of propagation in matter and in vacuum, recall that the discussion in Section 11.2 was based on the absence of charges and currents. However when an electromagnetic wave propagates through matter, even if there are no free charges and currents, the time-varying electromagnetic field induces certain charges and currents in the substance as a result of the polarization and magnetization of matter. If the substance is homogeneous and isotropic, the net effect of the polarization and magnetization of the medium by the electromagnetic wave is to replace the constants ϵ_0 and μ_0 in the Maxwell equations by the electric permittivity ϵ and the magnetic permeability μ characteristic of the material. Everything in the calculations of Section 11.2 remains the same except that the velocity of the wave now becomes

$$v=\frac{1}{\sqrt{\epsilon\mu}}. \tag{12.14}$$

The ratio between the velocity of electromagnetic waves in vacuum, c, and in matter, v, is called the *index of refraction* of the substance, and is designated by n. This concept is useful for describing the properties of materials in relation to electromagnetic waves. Thus

$$n=\frac{c}{v}=\frac{1}{\sqrt{\epsilon_0\mu_0}}\sqrt{\epsilon\mu}=\sqrt{\frac{\epsilon\mu}{\epsilon_0\mu_0}}.$$

However $\epsilon/\epsilon_0=\epsilon_r$ and $\mu/\mu_0=\mu_r$ where ϵ_r and μ_r are the relative permittivity and permeability of the medium. Then

$$n=\frac{c}{v}=\sqrt{\epsilon_r\mu_r}. \tag{12.15}$$

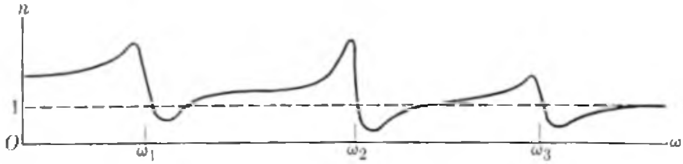


Fig. 12-12. Variation of the index of refraction with frequency and wavelength.

In general μ_r differs very little from 1 for the majority of substances that transmit electromagnetic waves in the visible region (see Table 6-1); a satisfactory approximation is therefore

$$n=\sqrt{\epsilon_r}. \tag{12.16}$$

This relation affords a simple experimental method for determining the relative permittivity of the substance if the index of refraction is obtained independently (as may be done). The consistency of the values of ϵ_r obtained by this method with those from other kinds of measurement gives a satisfactory foundation to the theory. In Section 2.7 we calculated ϵ_r, given by Eq. (2.24). Then using N for the number of electrons per unit volume to avoid confusion with the index of refraction, one may write

$$n^2=\epsilon_r=1+\frac{Ne^2}{m_e\epsilon_0}\left(\sum_i \frac{f_i}{\omega_i^2-\omega^2}\right) \tag{12.17}$$

where f_i is the fraction of the oscillations with characteristic frequency ω_i. Therefore the index of refraction depends on the wave frequency and hence also on the wavelength in a manner similar to that illustrated in Fig. 2-2 for ϵ_r and shown in Fig. 12-12, in which ω_1, $\omega_2, \ldots$ are the characteristic frequencies of the emission spectrum of the substance. Consequently the phase velocity $v=c/n$ of the electromagnetic wave in matter also depends on the frequency of the radiation. Therefore electromagnetic waves suffer dispersion when propagating in matter. That is, a pulse containing several frequencies will be distorted because each component will travel with a different velocity as discussed in Chapter 10.

The group velocity v_g from Eq. (10.56) is given by

$$v_g=v+k\frac{dv}{dk}.$$

Now, since $v_g=d\omega/dk$,

$$\frac{dv}{dk}=\frac{dv}{d\omega}\cdot\frac{d\omega}{dk}=v_g\frac{dv}{d\omega}.$$

Using $v=c/n$ gives

$$\frac{dv}{d\omega}=-\frac{c}{n^2}\frac{dn}{d\omega}$$

where Fig. 12-12 shows the functional relationship between n and ω. Therefore

$$v_g = v - \frac{v_g ck}{n^2}\frac{dn}{d\omega}.$$

Solving for v_g gives

$$v_g = \frac{v}{1+(ck/n^2)(dn/d\omega)} = \frac{c}{n+\omega(dn/d\omega)} \tag{12.18}$$

where the last expression is obtained by using $k = \omega/v = \omega n/c$.

When $dn/d\omega$ is positive, the group velocity is smaller than the phase velocity. Such a situation is called *normal dispersion*: but if $dn/d\omega$ is negative, then the group velocity is larger than the phase velocity and *anomalous dispersion* results. The possibility exists in this case that the group velocity as here defined is larger than c and that therefore an electromagnetic pulse can apparently be transmitted at a velocity larger than c. This situation is in contradiction to the results derived from the Lorentz transformation and the principle of relativity. However a careful analysis of the transmission of an electromagnetic signal, made by Brillouin, Sommerfeld, and others (an analysis that is very complex mathematically), revealed that even in a dispersive medium it is impossible to transmit a signal with a velocity larger than c. Figure 12-13 shows the variation of the phase velocity v, the group velocity v_g, and the signal velocity v_σ near a characteristic frequency ω_i. The signal velocity practically coincides with the group velocity except near the characteristic frequency, and is never larger than c, even in the region of anomalous dispersion.

When n is larger than one so that v is smaller than c, there is the possibility that a charged particle q, emitting electromagnetic waves, moves in the medium with a velocity v_q larger than the phase velocity v of the electromagnetic waves. This situation corresponds to that depicted in Fig. 10-28 for Mach waves in a fluid. Then the electromagnetic waves propagate along conical surfaces that make an angle α with the direction of propagation given by

$$\sin\alpha = \frac{v_q}{v}$$

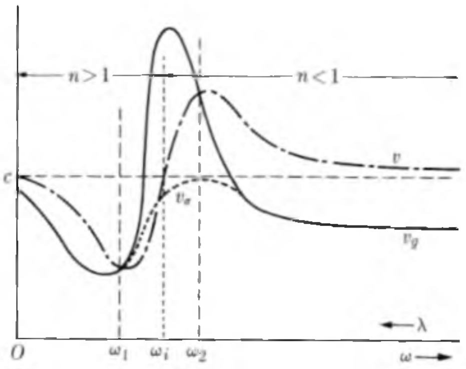


Fig. 12-13. Phase, group, and signal velocities of an electromagnetic pulse in a dispersive medium. There is anomalous dispersion in the region $\omega_1 < \omega_i < \omega_2$.

in accordance with Eq. (10.61). These waves are called *Cerenkov radiation* to honor the Russian scientist Pavel Cerenkov (1904–). Because the effective direction of propagation of the wave front is related to the velocity of the charged particle, measurement of the angle of Cerenkov radiation may be used to measure the particle speed. Devices used for this purpose are called *Cerenkov detectors*, which are widely used in experiments with fundamental particles because the detectors provide direct information about the velocity of the particles.

We have said that electromagnetic waves appear to propagate in matter with a phase velocity different from their propagation velocity in vacuum; but that difference seems to stem from the fact that the permittivity and permeability of matter are different from those of vacuum. This difference (in the permittivity and permeability) is in turn a consequence of the electric and magnetic polarization of matter under the action of the incoming electromagnetic wave. Thus when an electromagnetic wave falls on a piece of matter, the wave induces oscillations in the charged particles of the atoms or molecules, which then emit secondary or "scattered" waves (see Section 12.3). These scattered waves are superposed on the original wave to give a resultant wave. The phases of the secondary waves are generally different from those of the original wave since a forced oscillator is not always in phase with the driving force (remember Section 12.13 of Volume I). A detailed analysis, here omitted, indicates that this phase difference affects the resultant wave in such a way that the wave appears to have a velocity different from that in vacuum. This result is particularly satisfying since from the atomic point of view, all charges, both free and bound, are equivalent, and the electromagnetic waves they emit must all propagate with the velocity c. It is the wave resulting from the superposition of their individual waves, each with a different phase, that consequently appears to have a different velocity of propagation.

Example 12.3. Group velocity for very-high-frequency electromagnetic radiation, such as X-rays.

▼ If ω is much greater than the characteristic frequencies ω_i in Eq. (12.17), we may neglect the ω_i's and write Eq. (12.17) in the form

$$n^2 = 1 - \frac{Ne^2}{m_e \epsilon_0 \omega^2} \sum_i f_i = 1 - \frac{Ne^2}{m_e \epsilon_0 \omega^2}$$

since $\Sigma_i f_i = 1$ as indicated in Section 2.7. With the approximation given in (M.28) $(1-x)^{1/2} = 1 - \frac{1}{2}x + \cdots$, which applies when $x \ll 1$, the index of refraction is

$$n = 1 - \frac{Ne^2}{2\epsilon_0 m_e \omega^2},$$

which is smaller than one and gives $v > c$. From this expression we get $dn/d\omega = Ne^2/\epsilon_0 m_e \omega^3$, and substitution in Eq. (12.18) yields

$$v_g = \frac{c}{n + \omega(dn/d\omega)} = \frac{c}{1 - (Ne^2/2\epsilon_0 m_e \omega^2) + \omega(Ne^2/\epsilon_0 m_e \omega^3)}$$

$$= \frac{c}{1 + (Ne^2/2\epsilon_0 m_e \omega^2)}.$$

Therefore although the phase velocity v is larger than c because n is smaller than one, the group velocity v_g is smaller than c. ▲

12.8 Doppler Effect in Electromagnetic Waves

In Section 10.11 we discussed the Dopler effect for elastic waves and other kinds of mechanical waves that consist of matter in motion. The Doppler effect for electromagnetic waves must be discussed separately because in the first place electromagnetic waves do not consist of matter in motion, and therefore the velocity of the source relative to the medium does not enter into the discussion. Second, their velocity of propagation is c and is the same for all observers irrespective of their relative motion. The Doppler effect for electromagnetic waves must necessarily be computed by means of the principle of relativity.

To an observer in one inertial frame of reference, a plane, harmonic, electromagnetic wave can be described by a function of the form sin $(kx-\omega t)$ multiplied by an appropriate amplitude factor. To an observer attached to a different inertial frame of reference, the coordinates x and t must be replaced by x' and t' as given by the Lorentz transformation; and therefore the observer will write sin $(k'x'-\omega' t')$ where k' and ω' are not necessarily the same as for the other observer. On the other hand the principle of relativity requires that the expression $kx-\omega t$ must remain invariant when we pass from one inertial observer to another. Therefore we must have

$$kx-\omega t=k'x'-\omega' t'.$$

Using the first and fourth equations of the reciprocal Lorentz transformation (see the appendix) gives

$$k\frac{x'+vt'}{\sqrt{1-v^2/c^2}}-\omega\frac{t'+vx'/c^2}{\sqrt{1-v^2/c^2}}=k'x'-\omega' t'$$

or

$$\frac{k-\omega v/c^2}{\sqrt{1-v^2/c^2}}x'-\frac{\omega-kv}{\sqrt{1-v^2/c^2}}t'=k'x'-\omega' t'.$$

Therefore

$$k'=\frac{k-\omega v/c^2}{\sqrt{1-v^2/c^2}}, \qquad \omega'=\frac{\omega-kv}{\sqrt{1-v^2/c^2}}. \tag{12.19}$$

Remembering that $\omega=ck$, we can write these equations in the form

$$k'=k\frac{1-v/c}{\sqrt{1-v^2/c^2}} \qquad \omega'=\omega\frac{1-v/c}{\sqrt{1-v^2/c^2}}. \tag{12.20}$$

For small velocities, that is, $v \ll c$, the denominator may be set to unity, resulting in

$$\omega' = \omega\left(1 - \frac{v}{c}\right),$$

which is the same as Eq. (10.59) for motion of the observer relative to the source along the line of propagation. [Note that v_0 in Eq. (10.59) is now denoted by v, and v by c.]

Equation (12.20) relates the frequencies ω and ω' as measured by two observers O and O' when O' is moving along the X-axis with a velocity v relative to O. When the relative motion of the two observers is not along the line of propagation, a more elaborate calculation (see Example 12.4) indicates that

$$\omega' = \omega \frac{1 - (v/c)\cos\theta}{\sqrt{1 - v^2/c^2}} \tag{12.21}$$

where θ is the angle between the direction of propagation and the direction of the relative motion. Assuming that O is at rest relative to the source of the electromagnetic wave, we see that if the source O and the observer O' are receding from each other, O' observes a lower frequency, or correspondingly longer wavelength. This situation is observed in the spectrum of stars and is called the *red shift* since the visible spectrum of the light from receding stars is shifted toward the red (or longer) wavelengths. This factor allows estimating the velocity with which these stars are receding.

Figure 12-14 shows the red shift of the calcium lines H and K observed in the spectra of several nebulas. The shift is indicated by the horizontal arrow. The data given in Fig. 12-14 has been plotted in Fig. 12-15, which shows that the greater the distance of the nebula, the larger the shift and therefore the larger the relative velocity. This information supports the theory of an expanding universe. Recently certain nebulas have been observed to have such a large red shift that they appear to be receding at velocities as large as half the velocity of light; such high velocity has provoked thoughts that the Doppler effect may not be solely responsible for the red shift. It is interesting to note that the light from Andromeda (Fig. 1-6 in Volume I) shows a shift toward shorter wavelengths, or a blue shift. This seems to indicate that the present motion of the solar system within our rotating galaxy is in a direction toward this nebula.

Figure 12-16 shows the shift of the spectrum from the star Arcturus, which is about 36 light years from the sun. The two spectra were recorded six months apart; the shift of one is toward the red and the shift of the other is toward the blue. This shift is due to the reversal in the direction of the orbital motion of the earth relative to Arcturus.

We have derived expression (12.20), applying the principle of relativity to the phase $kx - \omega t$ of the wave. On the other hand we have attached certain momentum and energy to the radiation and used them to develop the concept of the photon. We must then see if our logic is consistent by checking the Lorentz transformation

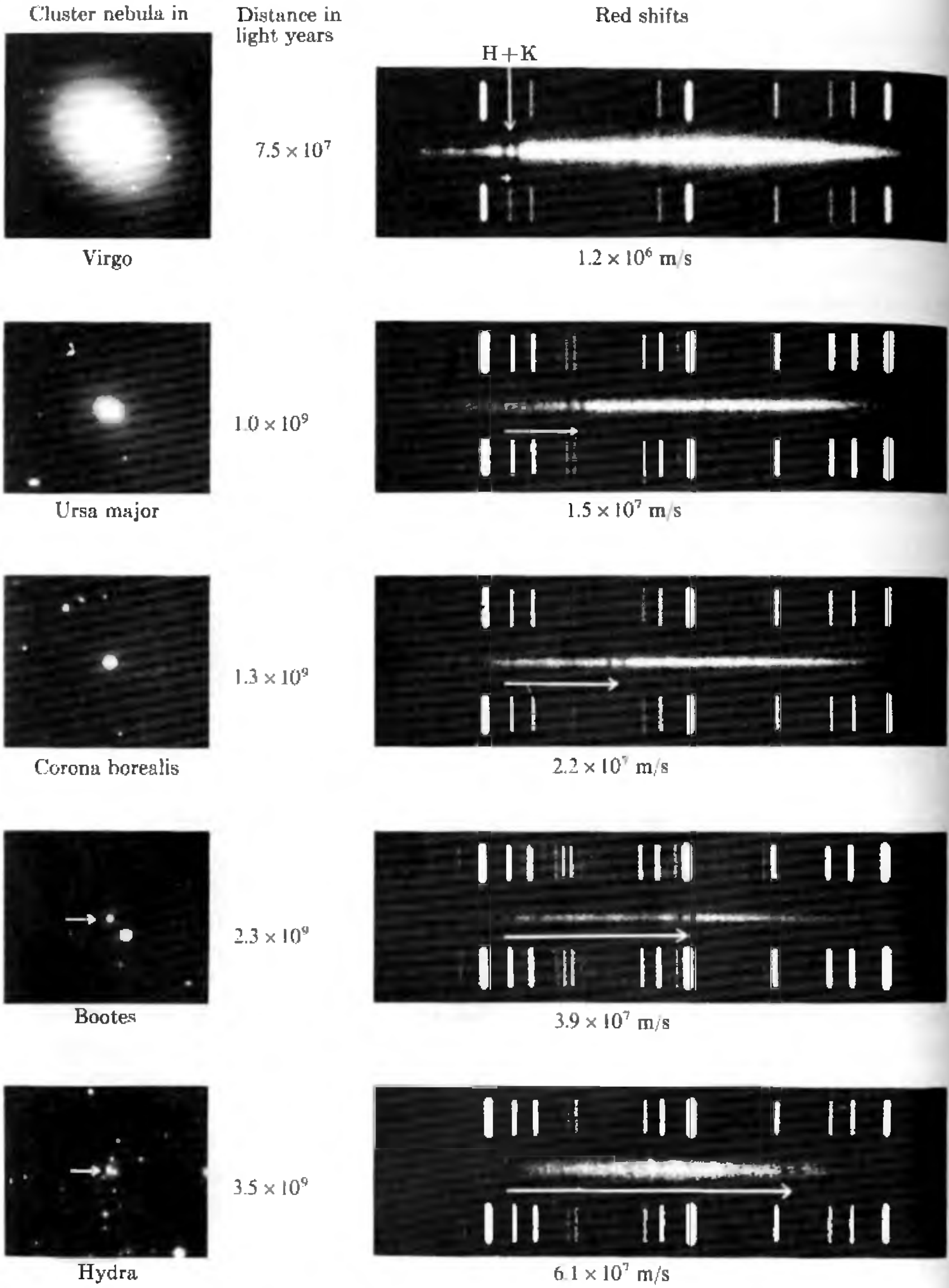


Fig. 12-14. Doppler effect in extragalactic nebulas. The red shift of the spectral H and K calcium lines (indicated by the arrow) increases with the distance of the nebula; this increase suggests greater recessional velocities. (Photograph courtesy of Mt. Wilson and Palomar Observatories.)

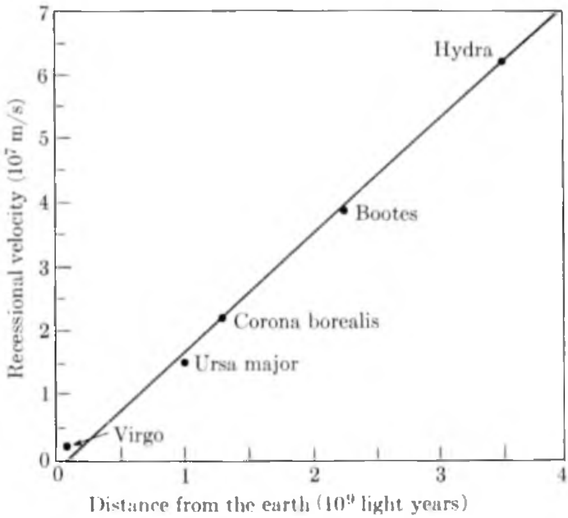


Figure 12-15

of the energy and momentum for a photon to see if they are compatible with Eq. (12.19). From Eq. (12.8) we have, for a photon, that

$$E = h\nu, \qquad p = \frac{h}{\lambda}. \tag{12.22}$$

Applying the Lorentz transformation for energy and momentum from one inertial frame to another gives

$$p' = \frac{p - vE/c^2}{\sqrt{1 - v^2/c^2}} \quad \text{and} \quad E' = \frac{E - vp}{\sqrt{1 - v^2/c^2}} \tag{12.23}$$

Fig. 12-16. Spectra (λ 4200 Å to λ 4300 Å) of the constant-velocity star Arcturus taken about six months apart. (a) July 1, 1939; measured velocity $+18$ km s^{-1} relative to earth. (b) January 19, 1940; measured velocity -32 km s^{-1}. The velocity difference of 50 km s^{-1} is entirely due to the change in orbital velocity of the earth. One can see the shift in the spectral lines clearly when one compares them with the two reference spectra. (Photograph courtesy of Mt. Wilson and Palomar Observatories.)

Then using the relations $E'=h\nu'$, $p'=h/\lambda'=h\nu'/c$, and similarly for E and p, and canceling the common factor h in all terms both equations reduce to

$$\nu'=\nu\frac{1-v/c}{\sqrt{1-v^2/c^2}} \tag{12.24}$$

which is equivalent to Eq. (12.20). We may then conclude that the concept of the photon is certainly consistent with the principle of relativity. If we assume that the relation $E=cp$ holds along with $\omega=ck$. we see that the mere comparison of Eqs. (12.19) and (12.23) would have tempted the physicist to look for a correspondence of the type $E\rightarrow\omega$ or ν and $p\rightarrow k$ or $1/\lambda$. If this had been the guiding principle, the concept of a photon would most probably have evolved earlier as a theoretical requirement of the model.

Example 12.4. Proof of relation (12.21) for the Doppler effect.

▼ When the direction of propagation of a plane electromagnetic wave makes an angle θ with respect to the direction of relative motion of two observers O and O', instead of the second equation of (12.23) we must write the equation

$$E'=\frac{E-vp_x}{\sqrt{1-v^2/c^2}}.$$

However $p_x=p\cos\theta$. Then remembering that $E=cp$ for a photon, we may write

$$E'=\frac{E-vp\cos\theta}{\sqrt{1-v^2/c^2}}=E\frac{1-(v/c)\cos\theta}{\sqrt{1-v^2/c^2}} \tag{12.25}$$

Using $E=h\nu=h\omega/2\pi$ and canceling common factors finally yields

$$\omega'=\omega\frac{1-(v/c)\cos\theta}{\sqrt{1-v^2/c^2}},$$

which is Eq. (12.21). ▲

Example 12.5. Relation between the directions of propagation of a plane electromagnetic wave as determined by two observers in relative motion. This effect is called *aberration*.

▼ Consider a source at rest relative to observer O where the observer sees an electromagnetic wave propagating in a direction making an angle θ with the X-axis. The X-axis coincides with the direction of the relative motion of the two observers. Then according to Eq. (11.23) of Volume I,

$$p'_x=\frac{p_x-vE/c^2}{\sqrt{1-v^2/c^2}};$$

but $p_x=p\cos\theta$; and similarly for observer O', $p'_x=p'\cos\theta'$. Therefore setting $E=cp$ gives

$$p'\cos\theta'=p\frac{\cos\theta-v/c}{\sqrt{1-v^2/c^2}}.$$

If we make $p=h/\lambda=h\omega/2\pi c$, and similarly for p', and cancel common factors, we obtain

$$\omega' \cos \theta' = \omega \frac{\cos \theta - v/c}{\sqrt{1-v^2/c^2}}.$$

Combining this with Eq. (12.21) to eliminate the frequencies yields

$$\cos \theta' = \frac{\cos \theta - v/c}{1-(v/c)\cos \theta}, \tag{12.26}$$

which relates the directions of propagation of the electromagnetic wave as determined by the two inertial observers. ▲

12.9 The Spectrum of Electromagnetic Radiation

Electromagnetic waves cover a wide range of frequencies or wavelengths, and may be classified according to their main source. The classification does not have very sharp boundaries since different sources may produce waves in overlapping ranges of frequencies.

The usual classification of the electromagnetic spectrum has seven specific categories. They are as follows.

1. *Radiofrequency waves.* These have wavelengths ranging from a few kilometers down to 0.3 m. The frequency range is from a few Hz up to 10^9 Hz. The energy of the photons goes from almost zero up to about 10^{-5} eV. These waves, which are used in television and radio broadcasting systems, are generated by electronic devices, mainly oscillating circuits.

2. *Microwaves.* The wavelengths of microwaves range from 0.3 m down to 10^{-3} m. The frequency range is from 10^9 Hz up to 3×10^{11} Hz. The energy of the photons goes from about 10^{-5} eV up to 10^{-3} eV. These waves are used in radar and other communication systems, as well as in the analysis of very fine details of atomic and molecular structure, and are also generated by electronic devices. The microwave region is also designated as UHF (ultrahigh frequency relative to radio frequency).

3. *Infrared spectrum.* This region covers wavelengths from 10^{-3} m down to 7.8×10^{-7} m. The frequency range is from 3×10^{11} Hz up to 4×10^{14} Hz and the energy of the photons goes from 10^{-3} eV up to about 1.6 eV. The region is further subdivided into three: the *far infrared*, from 10^{-3} m to 3×10^{-5} m; the *middle infrared*, from 3×10^{-5} m to 3×10^{-6} m; and the *near infrared*, extending up to about 7.8×10^{-7} m. These waves are produced by molecules and hot bodies, and have many applications in industry, medicine, astronomy, etc.

4. *Light or visible spectrum.* This is a narrow band formed by the wavelengths to which the human retina is sensitive. The band extends from a wavelength of 7.8×10^{-7} m down to 3.8×10^{-7} m and frequencies from 4×10^{14} Hz up to 8×10^{14} Hz. The energy of the photons goes from 1.6 eV up to about 3.2 eV. Light is produced by atoms and molecules as a result of internal adjustment in the motion of its components, mainly the electrons. It is not necessary to emphasize the importance of light in our world.

Light is so important that a special branch of applied physics, called *optics*, has evolved. Optics deals with light phenomena as well as vision, and includes the design of optical instruments. Because of the similarity in the behavior of the infrared and the ultraviolet regions of the spectrum, the field of optics now includes both, in addition to the visible spectrum. The different sensations, called *colors*, that light produces on the eye depend on the frequency, or the wavelength, of the electromagnetic wave, and correspond to the following ranges for the average person:

Color	λ, m	ν, Hz
Violet	$3.90–4.55 \times 10^{-7}$	$7.69–6.59 \times 10^{14}$
Blue	4.55–4.92	6.59–6.10
Green	4.92–5.77	6.10–5.20
Yellow	5.77–5.97	5.20–5.03
Orange	5.97–6.22	5.03–4.82
Red	6.22–7.80	4.82–3.84

The sensitivity of the eye depends also on the wavelength of light; this sensitivity is a maximum for wavelengths of approximately 5.6×10^{-7} m. Because of the relation between color and wavelength or frequency, an electromagnetic wave of well-defined wavelength or frequency is also called a *monochromatic wave* (from *monos*, one, and *chromos*, color).

Vision is the result of the signals transmitted to the brain by two elements present in a membrane called the *retina*, lying in the back of the eye. These elements are the *cones* and the *rods*. Cones are active in the presence of intense light, such as that which exists during the daylight hours. Cones are very sensitive to the frequency or color of the wave. Rods, on the other hand, are able to act under very dim illumination, such as that in a darkened room; they are quite insensitive to frequency or color. The vision associated with cones is called *photopic*; that associated with rods is called *scotopic*. The sensitivity of the eye for different wavelengths for both kinds of vision is illustrated in Fig. 12-17.

5. *Ultraviolet rays.* These wavelengths cover from 3.8×10^{-7} m down to about 6×10^{-10} m, with frequencies from 8×10^{14} Hz to about 3×10^{17} Hz. The energy of the photons goes from about 3 eV up to 2×10^{3} eV. These waves are produced by

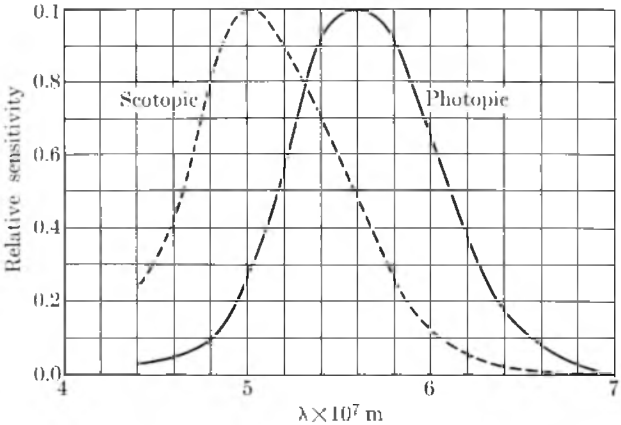


Fig. 12-17. Sensitivity of the eye for scotopic and photopic vision.

atoms and molecules and in electric discharges. Their energy is of the order of magnitude of the energy involved in many chemical reactions, accounting for many of their chemical effects. The sun is a very powerful source of ultraviolet radiation, and it is this factor that is mainly responsible for suntans. The sun's ultraviolet radiation also interacts with the atoms in the upper atmosphere and produces a large number of ions, thus explaining why the upper atmosphere at a height greater than about 80 km is highly ionized and is called the *ionosphere*. When some living cells absorb ultraviolet radiation, they can be destroyed as a result of the chemical reactions produced by the ionization and dissociation of their molecules. For that reason ultraviolet rays are used in some medical applications and also in sterilization processes in which microorganisms, such as bacteria, may be killed.

6. *X-rays.* This part of the electromagnetic spectrum extends from wavelengths of about 10^{-9} m down to wavelengths of about 6×10^{-12} m, or frequencies between 3×10^{17} Hz and 5×10^{19} Hz. The energy of the photons goes from 1.2×10^{3} eV up to about 2.4×10^{5} eV. This part of the electromagnetic spectrum was discovered in 1895 by the German physicist W. Roentgen (1845–1923) when he was studying cathode rays. X-rays are produced by the inner, or more tightly bound, electrons in atoms. Another source of X-rays is the bremsstrahlung or decelerating radiation mentioned in Section 11.7. In fact this is the most common way of producing X-rays in commercial X-ray tubes. A beam of electrons, accelerated by a potential of several thousand volts, impinges on a metallic target called the anode (Fig. 11-16) (actually this method was the way in which the X-rays were produced in the original experiment of Roentgen). X-rays, because of the greater energy of their photons, produce more profound effects on the atoms and molecules of the substances through which the X-rays propagate, dissociating or ionizing the molecules. X-rays are used in

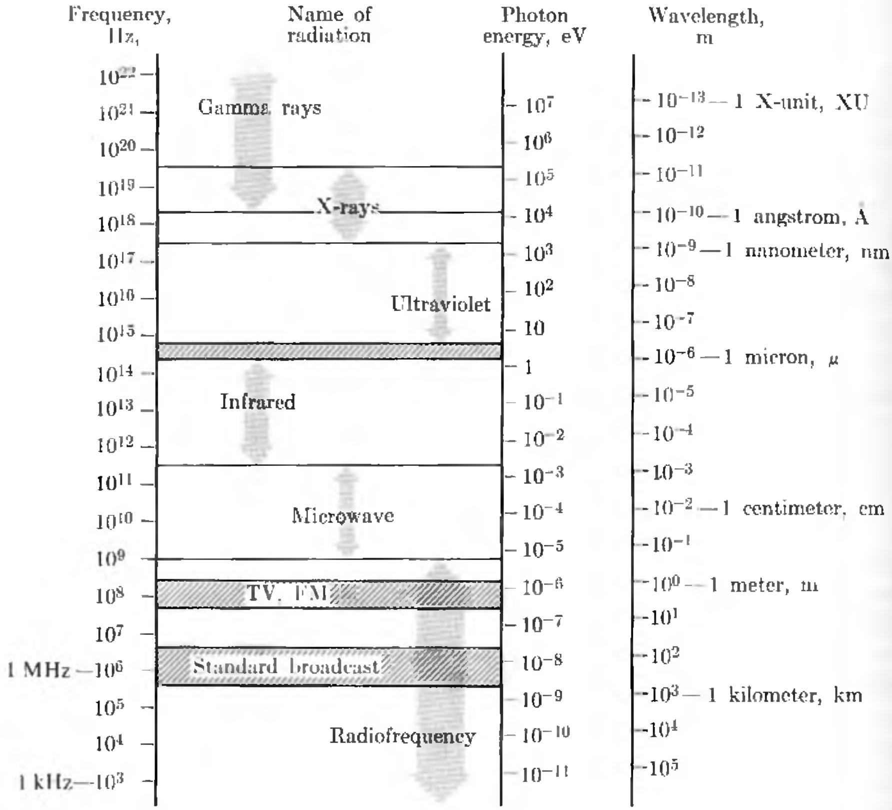


Fig. 12-18. The electromagnetic spectrum.

medical diagnosis because the relatively greater absorption of X-radiation by bone as compared with tissue allows for a fairly well-defined "photograph" of bone structure. As a result of the chemical processes they induce, X-rays also cause serious damage to living tissues and organisms; X-rays are used for treatment of cancer since they seem to have a tendency to destroy diseased tissue more readily than healthy tissue. It should be emphasized that *any* amount of X-radiation does destroy some good tissue; an exposure to a large dose may cause enough destruction to produce sickness or death.

7. *Gamma rays.* These electromagnetic waves are of nuclear origin. They overlap the upper limit of the X-ray spectrum. Their wavelength runs from about 10^{-10} m down to well below 10^{-14} m, with a corresponding frequency range from 3×10^{18} Hz up to more than 3×10^{22} Hz. The energies of the photons go from 10^4 eV to about 10^7 eV. These energies are of the same order of magnitude as those involved in nuclear processes, and therefore the absorption of γ-rays may produce some nuclear changes. These rays are produced by many radioactive substances, and are present in large quantities in nuclear reactors. Although not easily absorbed by most substances, gamma rays when they are absorbed by living organisms produce serious effects. Their handling requires a heavy protective shielding.

In cosmic radiation there are electromagnetic waves of even shorter wavelengths or larger frequencies, and with photons that are correspondingly more energetic. These waves are of special interest in astronomical research.

From the breadth of the spectrum of electromagnetic radiation, we can easily understand why the different parts of the spectrum behave differently when propagating through matter. For example those waves having photons with an energy comparable to the characteristic energies of electrons in atoms or of atoms in molecules will interact more strongly with atoms and molecules. This is the case for infrared, visible, and ultraviolet radiation. Radiation having longer wavelength and carrying photons with less energy generally interacts weakly with matter. Radio-frequency waves are examples of such radiation. Waves such as X- and γ-rays, that have high energy or very short wavelength, are also absorbed very little in matter; however, when they are absorbed, the effects of these waves are more profound and produce not only atomic and molecular ionization, but also nuclear break-up in many cases.

Figure 12-18 relates the various sections of the electromagnetic spectrum in terms of energy, frequency, and wavelength.

Problems

12.1 Radiation having a wavelength of 10^{-10} m (or 1 Å) undergoes Compton scattering in a carbon sample. The scattered radiation is observed at a direction perpendicular to that of incidence. Find (a) the wavelength of the scattered radiation, and (b) the kinetic energy and direction of motion of the recoil electrons. (*Hint:* See Problem 12.32.)

12.2 Refer to the preceding problem; if the electrons recoil at an angle of 60° relative to the incident radiation, find (a) the wavelength and direction of the scattered radiation, and (b) the kinetic energy of the electron. (*Hint:* See Problems 12.31 and 12.32.)

12.3 (a) Prove that the kinetic energy of the recoiling electron in the Compton effect is given by

$$E_k = \frac{h\nu\alpha(1-\cos\theta)}{[1+\alpha(1-\cos\theta)]}$$

where $\alpha = h\nu/m_e c^2$. (b) Show that the maximum energy of the recoil electron is

$$\frac{(h\nu)^2}{(h\nu+\frac{1}{2}m_e c^2)} \approx h\nu - \tfrac{1}{2}m_e c^2$$

if $h\nu \gg \frac{1}{2}m_e c^2$.

12.4 A photon having an energy of 10^4 eV collides with a free electron at rest and is scattered through an angle of 60°. Find (a) the changes in energy, frequency, and wavelength of the photon; and (b) the kinetic energy, momentum, and direction of the recoiling electron.

12.5 Determine the frequency and the wavelength of the photons absorbed by the following systems: (a) a nucleus absorbing energy in the amount of 10^3 eV, (b) an atom absorbing 1 eV, and (c) a molecule absorbing 10^{-2} eV.

12.6 Sodium atoms absorb or emit electromagnetic radiation of 5.9×10^{-7} m, corresponding to the yellow region of the visible spectrum. Determine the energy of the photons that are absorbed or emitted.

12.7 To separate the carbon and oxygen atoms that form the carbon monoxide molecule, a minimum energy of 11 eV is required. Find (a) the minimum frequency and (b) maximum wavelength of the electromagnetic radiation required to dissociate the molecule.

12.8 A photon having an energy of 10^4 eV is absorbed by a hydrogen atom at rest. As a result, the electron is ejected in the same direction as the incident radiation. Neglecting the energy required to separate the electron (about 13.6 eV), find (a) the momentum and (b) the kinetic energy of the electron and of the proton.

12.9 Find the energy of a photon having the same momentum as a 40-MeV (a) proton, (b) electron. (*Hint:* Nonrelativistic equations may be used.) (*Hint:* Note that the proton can be treated nonrelativistically, but for the electron relativistic mechanics is required.)

12.10 The binding energy of an electron in lead is 9×10^4 eV. When lead is irradiated with a certain electromagnetic radiation and the photoelectrons enter a magnetic field of 10^{-2} T, they describe a circle of radius 0.25 m. Compute (a) the momentum and energy of the electrons, and (b) the energy of the photons absorbed. (c) Can you neglect the effect of the recoiling lead ion?

12.11 When a certain metal surface is illuminated with light of different wavelengths, the stopping potentials of the photoelectrons are measured as shown in the table:

$\lambda(\times 10^{-7}$ m)	V(V)
3.66	1.48
4.05	1.15
4.36	0.93
4.92	0.62
5.46	0.36
5.79	0.24

(a) Plot the stopping potential as ordinate against the frequency of the light as abscissa. From the graph determine (b) the threshold frequency, (c) the photoelectric work function of the metal, and (d) the ratio h/e.

12.12 The photoelectric work function of potassium is 2.0 eV. Supposing that light having a wavelength of 3.6×10^{-7} m falls on potassium, find (a) the stopping potential of the photoelectrons, and (b) the kinetic energy and the velocity of the fastest electrons ejected.

12.13 A uniform monochromatic beam of wavelength 4.0×10^{-7} m falls on a material having a work function of 2.0 eV. If the beam has an intensity of 3.0×10^{-9} W m^{-2}, find (a) the number of electrons emitted per m^2 and per s, and (b) the energy absorbed per m^2 and per s.

12.14 Consider a gas whose molecules behave as dipole oscillators, with a restoring constant of $k = 3 \times 10^2$ kg s^{-2}. The oscillating particles are electrons. (a) Compute their characteristic frequency. (b) Write the index of refraction of the gas as a function of the frequency if the gas is at STP. (c) Obtain the values of the index for $\lambda = 5 \times 10^{-7}$ m and $\lambda = 1.02 \times 10^{-7}$ m.

12.15 Verify that the quantity $Ne^2/m\epsilon_0\omega^2$ in Example 12.3 is small compared with unity in the X-ray region.

12.16 Consider a substance moving with a velocity v parallel to the X-axis. Let $V' = c/n'$ be the velocity of light in the substance as measured by an observer O' at rest relative to the substance. Show that the velocity V of a wave propagating along the X-axis through the substance, as measured by an observer O relative to which the substance moves with velocity v, is

$$V \approx \frac{c}{n'} + v\left(1 - \frac{1}{n'^2}\right).$$

(*Hint:* Use the Lorentz transformation of velocities.)

12.17 Show that when light propagates through a medium moving with velocity v parallel to the X-axis, the Doppler effect is

$$\nu' = \nu(1 - nv/c)$$

if $v \ll c$.

12.18 (a) Using the result of Problem 12.17, show that

$$n' = n - \left(\frac{n\nu\nu}{c}\right)\frac{dn}{d\nu}$$

where n' is evaluated at ν' and n at ν. (b) Since (referring to Problem 12.16) we must note that n' has to be evaluated at ν', show that the result of Problem 12.16 can also be written in the form

$$V \approx \frac{c}{n} + v\left[1 - \frac{1}{n^2} - \left(\frac{\lambda}{n}\right)\frac{dn}{d\lambda}\right].$$

(*Hint:* Note that $\lambda\nu = c/n$, and that in the last term n' can be replaced by n.)

CHALLENGING PROBLEMS

12.19 The minimum energy required to eject a photoelectron from a given metal surface is 4.0 eV. Light of what frequency would be needed to give photoelectrons ejected from the metal surface a maximum kinetic energy of 3.0 eV? [AP-B: 1969]

12.20 The graph in Fig. 12-19 shows the maximum kinetic energy of electrons ejected from a photoelectric surface as a function of the frequency of light striking the surface. What is the minimum energy photon that will eject a photoelectron? [AP-B; 1971]

12.21 The energy level diagram of a particular atom is shown in Fig. 12-20. It is known that absorption of a photon of frequency 4.8×10^{14} Hz will cause a transition from the ground state A to state B. (a) Determine the frequency of the photon that causes transition from state A to state C when the photon is absorbed by the atom. (b) Determine the momentum of the photon in part (a) above. (c) What is the minimum amount of energy required to ionize the atom when it is in its ground state? [AP-B: 1972]

12.22 A monochromatic light source emits light of wavelength 5,600 Å at the rate of 2.00 W. (a) Show that the energy per photon of the light emitted is 2.22 eV. (b) Determine the momentum per photon of the light emitted. (c) Determine the number of photons emitted per second. (d) The source is placed at the center of a hollow sphere of radius R, the surface of which absorbs all light. If the source radiates photons equally in all directions, find an expression for the pressure of the light on the surface of the sphere. [AP-B; 1972]

12.23 (a) Describe and interpret an experiment in which electromagnetic radiation exhibits particlelike behavior. (b) Describe and interpret an experiment in which electrons exhibit wave-

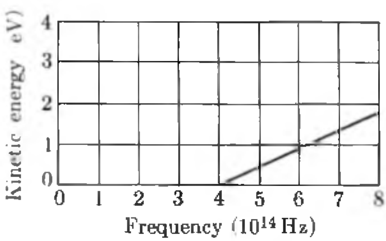


Figure 12-19

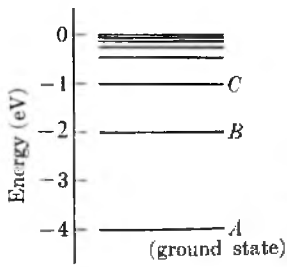


Figure 12-20

like behavior. [AP-B; 1973]

12.24 A parallel beam of monochromatic visible light enters your laboratory through a hole in the wall. You cannot investigate the source of light, though you have all the apparatus you wish for investigating the properties of the light beam. Describe how you could determine experimentally the number of photons per second entering the laboratory. [AP-B; 1974]

12.25 Figure 12-21 shows part of an energy-level diagram for a certain atom. The wavelength of the radiation associated with transition A is 6,000 Å and that associated with transition B is 3,000 Å. (a) Determine the energy of a photon associated with transition A. (b) Determine the wavelength of the radiation associated with transition C. (c) Describe qualitatively what will happen to an intense beam of white light (4,000–8,000 Å) that is sent through this gaseous element. [AP-B; 1975]

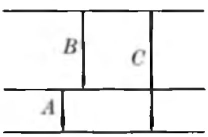


Figure 12-21

12.26 Light of wavelength λ_1 incident on a clean metal surface ejects photoelectrons of maximum kinetic energy KE_{max}. (a) Discuss the effect on the photoelectrons as the wavelength of the incident radiation is made longer and longer. (b) Discuss the effect on the photoelectrons if the intensity of the radiation is gradually increased while the wavelength remains constant at λ_1. (c) State an experimental observation in a photoelectric experiment that is not satisfactorily explained by a wave model of light. [AP-B; 1976]

12.27 An object of mass 0.5 kg (500 g) is at rest on a frictionless surface. A burst of 10^{20} photons strikes the object and all the photons are completely absorbed. Each photon has an energy of 100 eV $= 1.6 \times 10^{-17}$ J. As a result of absorbing the photons, the object's temperature rises 5 K. (a) Calculate the momentum of a single photon. (b) Calculate the speed of the object after it has absorbed the photons. (c) If, instead, all the photons had been reflected, how would the answer to part (b) be changed? (d) Calculate the specific heat of the material of which the object is made. [AP-B; 1978]

12.28 In a photoelectric experiment, radiation of several different frequencies was made to shine on a metal surface and the maximum kinetic energy of the ejected electrons was measured at each frequency. Selected results of the experiment are presented in the table below:

Frequency (Hz)	Maximum kinetic energy of electrons (eV)
0.5×10^{15}	No electrons ejected
1.0×10^{15}	1.0
1.5×10^{15}	3.0
2.0×10^{15}	5.0

(a) Plot the data from this photoelectric experiment. (b) Determine the threshold frequency of the metal surface. (c) Determine the work function of the metal surface. (d) When light of frequency 2.0×10^{15} Hz strikes the metal surface, electrons of assorted speeds are ejected from the surface. What minimum retarding potential would be required to stop all of the electrons ejected from the surface by light of frequency 2.0×10^{15} Hz? (e) Investigation reveals that some electrons ejected from the metal surface move in circular paths. Suggest a reasonable explanation for this electron behavior. [AP-B; 1980]

12.29 A helium atom of mass m moving with speed v zigzags back and forth between two parallel walls of length L separated by distance a as shown in Fig. 12-22. (a) In terms of a, v, and θ, calculate the time interval Δt between two successive collisions with the right-hand wall. (b) In terms of m, v, and θ, calculate the magnitude of the momentum Δp imparted to the right-hand wall each time the atom collides with it. (c) Calculate the average force that the atom exerts on the right-hand wall, and express the resulting pressure P on the wall in terms of

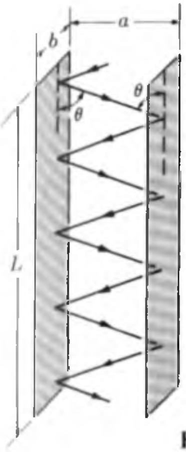


Figure 12-22

θ, the volume V of the region bounded by the walls, and the kinetic energy E of the atom. (d) Suppose, instead, that a photon of energy E is following the zigzag path. Calculate the magnitude of the momentum Δp it imparts to the right-hand wall in each collision, and express the resulting pressure P in terms of E, θ, and V. [AP-B; 1980]

12.30 The sun is 1.5×10^{11} m from the earth. Energy from the sun is received at the earth's surface at the rate of 1.4 kW m^{-2}. (a) This energy flux falls on a pond of water 100 m^2 in area and 0.1 m in depth. Assume all of this energy heats the water. Find the average temperature rise of the pond after 10^3 seconds. (b) Determine the rate in kilograms per second at which the sun's mass is being converted to energy. The surface area of a sphere is $4\pi r^2$. [AP-B; 1981].

12.31 Show that if the electron is scattered in a direction making an angle ϕ with the incident photon in a Compton scattering, the kinetic energy of the electron is

$$E_k = \frac{h\nu(2\alpha \cos^2 \phi)}{[(1+\alpha)^2 - \alpha^2 \cos^2 \phi]},$$

where $\alpha = h\nu/m_e c^6$.

12.32 Show that, in a Compton scattering, the relation between the angles defining the directions of the scattered photon and the recoil electron is $\cot \phi = (1+\alpha) \tan \frac{1}{2}\theta$.

12.33 Electromagnetic radiation of wavelength equal to 10^{-5} m falls normally on a metal sample of mass 10^{-1} kg, and an electron is ejected in a direction opposite to the incident radiation. Using the laws of conservation of energy and momentum, obtain the energy of the electron and the recoil energy of the metal sample. Assume that the work function is zero. Does the result justify not considering momentum conservation in our calculation of the photoelectric effect?

12.34 Show that for gases the second term in Eq. (12.17) is small, and that we may write

$$n \approx 1 + \frac{Ne^2}{2m_e\epsilon_0}\left(\sum_i \frac{f_i}{\omega_i^2 - \omega^2}\right).$$

For only one resonant frequency, the expression becomes

$$n \approx 1 + \frac{Ne^2}{2m_e\epsilon_0(\omega_i^2 - \omega^2)}.$$

12.35 The index of refraction of hydrogen gas at STP is $n = 1 + 1.400 \times 10^{-4}$ at $\lambda = 5.46 \times 10^{-7}$ m and $n = 1 + 1.547 \times 10^{-4}$ at $\lambda = 2.54 \times 10^{-7}$ m. Assuming a single resonant frequency, compute this frequency and the number of electronic oscillators per unit volume. Compare with the number of molecules per unit volume. (*Hint:* Use the result of Problem 12.34.)

12.36 Referring to the preceding problem, compute the index of refraction of hydrogen for $\lambda = 4 \times 10^{-7}$ m, a pressure of 10^6 Pa, and a temperature of 300 K.

12.37 Consider a glass plate of index of refraction n and thickness Δx interposed between a monochromatic source S and an observer O,

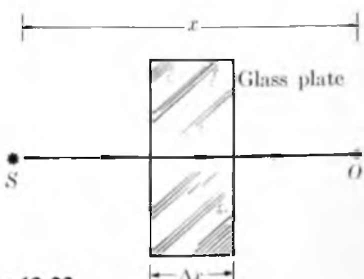


Figure 12-23

as shown in Fig. 12-23. (a) Show that if absorption by the glass plate is neglected, the effect of the glass plate on the wave received by O is to add a phase difference equal to

$$\delta = -\omega(n-1)\frac{\Delta x}{c}$$

without changing the amplitude $\mathscr{E}_0$ of the wave. (b) If the phase difference is small, either because Δx is very small or because n is very close to one, show that the wave received at O can be considered as a superposition of the original wave of amplitude $\mathscr{E}_0$, with no plate present, with a wave of amplitude $\mathscr{E}_0\omega(n-1)\Delta x/c$ having a phase shift of $-\pi/2$. (This problem shows the effect of a material medium on an electromagnetic wave.)

CHAPTER THIRTEEN

REFLECTION AND REFRACTION

13.1 Introduction

In all the kinds of waves discussed in Chapters 10, 11, and 12, the velocity of propagation depends on some physical properties of the medium through which the waves propagate. For example the velocity of elastic waves depends on an elastic modulus and the density of the medium. The velocity of electromagnetic waves depends on the permittivity and permeability of the substance through which the waves propagate.

This dependence of the velocity of propagation of a wave on the properties of the medium gives rise to the phenomena of *reflection* and *refraction*, which occur when a wave crosses a surface separating two media in which the wave propagates with different velocity. The *reflected wave* is a new wave that propagates back into the medium through which the initial wave was propagating. The *refracted wave* is the wave transmitted into the second medium. The energy of the incident wave splits and is divided between the reflected and the refracted waves. In many instances the reflected wave receives more of the energy as is the case for mirrors. In other instances the refracted wave carries most of the energy. When a transverse wave is polarized, the polarization is usually affected both in reflection and refraction; we shall ignore this effect for the present and discuss polarization in the next chapter.

13.2 Huygen's Principle

The propagation of a wave is described by the equations of the field to which the wave corresponds. Therefore if the source of a wave is known, its propagation from one region to another can in principle be traced if the changes in the properties of the medium are taken into account. It is also possible, however, to compute the amplitude of a wave at a particular point of space without making direct reference to the sources. Around 1680 the Dutch physicist Christiaan Huygens (1629–1695) proposed a simple mechanism for tracing the propagation of waves. His construction is applicable to either elastic or mechanical waves in a material medium.

Recall that a *wave surface* or a *wave front* is a surface having constant phase passing through those points of the medium that are reached by the wave motion at the same time. For example for a plane wave the disturbance is expressed by $f(\boldsymbol{u}\cdot\boldsymbol{r}-vt)$, and a wave surface is composed of all points at which the phase $\boldsymbol{u}\cdot\boldsymbol{r}-vt$ has the same value at a given time. Therefore the wave surface is given by the equation

$$\boldsymbol{u}\cdot\boldsymbol{r}-vt=\text{const},$$

which for a given time t corresponds to a plane perpendicular to the unit vector $\boldsymbol{u}$. Similarly for spherical waves the wave surfaces are given by $r-vt=\text{const}$, which for a given t correspond to spheres.

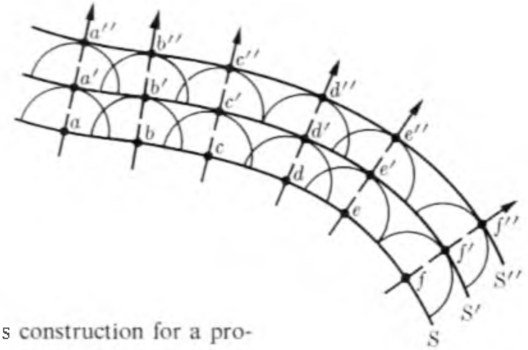


Fig. 13-1. Huygens's construction for a progressing wave.

Huygens visualized a method for passing from one wave surface to another when the wave is assumed to be the result of the motion of the particles comprising the medium. Consider a wave surface S (Fig. 13-1). When the wave motion reaches this surface, each particle $a, b, c, \ldots$ on the surface becomes a secondary source of waves, and each emits *secondary waves* (indicated by the small semicircles), which reach the next layer of particles in the medium. These particles are then set in motion and form the next wave surface S'. The surface S' is tangent to all secondary waves. The process keeps repeating itself and results in the propagation of a wave through the medium. This pictorial representation of the propagation of a wave looks very reasonable when the wave is an elastic wave resulting from mechanical vibrations of atoms or molecules in a body.

However, this representation has no physical meaning in cases like an electromagnetic wave's propagation in a vacuum, in which there are no vibrating particles. Huygen's construction, although quite plausible when applied to mechanical waves in matter, therefore required a revision after it was recognized that other waves of a different kind also exist in nature. This revision was accomplished at the end of the nineteenth century by the German Gustav Kirchhoff (1824–1887), who replaced Huygens's intuitive construction by a more mathematical treatment. Kirchhoff's calculations are too complicated to be reproduced here. The final result of his calculations, however, is relatively simple as will be seen in the succeeding paragraphs.

Wave motion is regulated by the general wave equation (10.65). That is,

$$\frac{\partial^2 \xi}{\partial t^2} = v^2 \left(\frac{\partial^2 \xi}{\partial x^2} + \frac{\partial^2 \xi}{\partial y^2} + \frac{\partial^2 \xi}{\partial z^2} \right) \tag{13.1}$$

where ξ may be the displacement of the atoms of a substance for an elastic wave, the electric or magnetic field for an electromagnetic wave, and so forth. Describing wave propagation in any given medium consists primarily in obtaining a solution $\xi(\mathbf{r}, t)$ to this differential equation. The solution of Eq. (13.1) depends on ξ satisfying the physical conditions of the problem; that is, the position and the nature of the sources, the physical surfaces of discontinuity, etc. Mathematicians call these con-

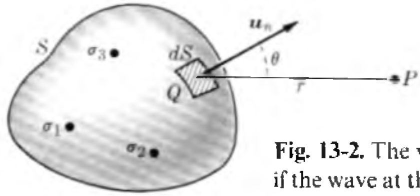


Fig. 13-2. The wave at P can be computed if the wave at the closed surface S is known.

ditions the *boundary conditions*. The theory of differential equations states that under special conditions we can find the solution of an equation such as Eq. (13.1) if we know the values of the function $\xi(\boldsymbol{r}, t)$ over a closed surface S (Fig. 13-2).

To be more concrete, suppose that we want to evaluate the wave motion at some point P. If we know the sources $\sigma_1, \sigma_2, \sigma_3, \ldots$, we may add all their contributions at P to obtain the resultant wave motion. Now suppose that, instead, we know only the value of ξ at all points of the arbitrary but closed surface S. In this case we may also obtain the wave at P even if the distribution of the sources is ignored. Mathematically this statement can be expressed in the following way. Let $f(Q, t)$ represent the wave at each point Q on the surface S at time t. Let r represent the distance from the surface element dS around Q to the point P. The perturbation at P at time t can be expressed by an integral of the form*

$$\xi_P(t) = \oint_S g(\theta) \frac{f(r - vt)}{r} \, dS \tag{13.2}$$

where the integral extends over the entire surface S. This integral has a rather simple physical interpretation: the factor $(1/r)f(r-vt)$ represents a spherical wave that is emitted by a surface element dS at time $t-r/v$ and that reaches P at time t where r/v is the time for propagation from dS to P. The factor $g(\theta)$ is a directional factor, meaning that the waves emitted by dS do not have the same amplitude in all directions. If θ is defined as the angle between the normal to the surface at dS and the direction from dS to P (Fig. 13-2), the form of $g(\theta)$ is

$$g(\theta) = \tfrac{1}{2}(1 + \cos\theta).$$

The maximum amplitude ($g = 1$) then corresponds to $\theta = 0$, or forward propagation; and the minimum amplitude ($g = 0$), to $\theta = \pi$, or backward propagation. We conclude then that *we can obtain the perturbation at a point P at time t if we assume that each surface element dS of the closed surface S acts as a secondary source of waves.* The computational process of Kirchhoff just described is essentially equivalent to Huygens's statement, but without reference to a mechanical model.

We shall have occasion to use Huygens's principle, as reformulated by Kirchhoff, in many of the forthcoming discussions of wave propagation, especially those dealing with diffraction and scattering in later chapters.

---

*The actual expression is somewhat more complicated; but Eq. (13.2) suits our purpose here and provides an adequate approximation, applicable to the kind of problems to be discussed in this book.

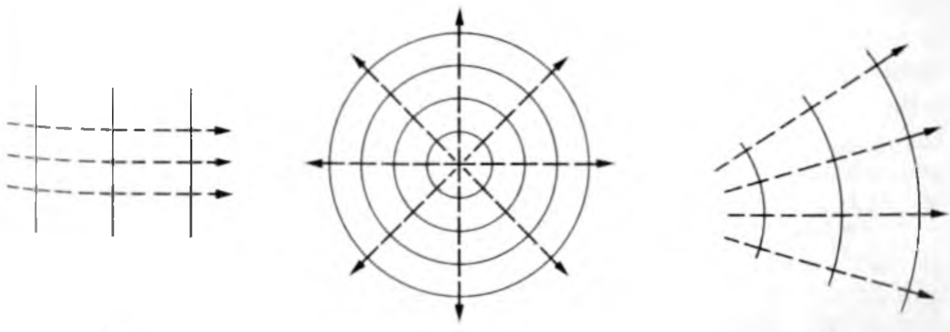
13.3 Malus's Theorem

Another important tool for tracing the propagation of a wave through a medium is *Malus's theorem*, named after the French physicist E. Malus (1775–1812). Referring back to Fig. 13-1, note that a series of lines may be drawn perpendicular to the successive wave surfaces (indicated by the broken lines with arrows). These lines, called *rays*, correspond to the lines of propagation of the wave. The relation between rays and wave surfaces is similar to the relation between lines of force and equipotential surfaces. Points that are on different wave surfaces and are joined by a given ray are called *corresponding points* (see points a, a', a'' or points b, b', b'' in Fig. 13-1). The time required by the wave to go from S to S'' must be the same measured along any ray since all points on S'' must have the same phase. We may thus state that

> *the time separation between corresponding points of two wave surfaces is the same for all pairs of corresponding points.*

We may then conclude that the distances aa'', bb'', cc'', etc., must depend on the velocity of the wave motion at each point along a given ray. In a homogeneous isotropic medium, in which the velocity is the same at all points and in all directions, the separation between two wave surfaces must be the same for all corresponding points; and therefore the rays are straight lines because symmetry suggests that there is no reason for their bending to one side or another. This situation has already been seen with plane and spherical waves as illustrated in parts (a) and (b) of Fig. 13-3. Therefore in the general case the family of wave surfaces must have a common set of normals as shown in Fig. 13-3(c), and must be equally spaced along these normals.

Consider now the case where a wave propagates through a succession of homogeneous isotropic media. At the crossing of each interface separating two adjacent media, the direction of propagation may change (that is, the rays may change direction); but while they are going through a given medium, the rays will still be



(a) (b) (c)

Fig. 13-3. Plane waves, spherical waves, and waves of arbitrary shape.

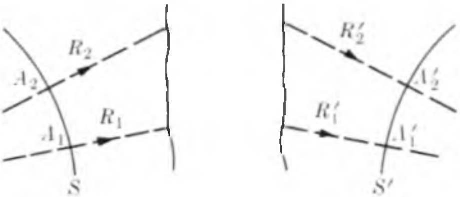


Fig. 13-4. Corresponding rays in incident and outgoing waves.

straight lines perpendicular to the wave surfaces. Let S (Fig. 13-4) be one wave surface in the first medium. Then two rays, R_1 and R_2, may be traced perpendicular to S. Successive wave surfaces in that medium must be perpendicular to R_1 and R_2. If after the wave motion passes through a series of different media, we observe another wave surface S', we find that the rays R_1 and R_2 have been transformed into rays R_1' and R_2', which are also perpendicular to S'. In other words *the relation of orthogonality between rays and wave surfaces is conserved throughout all the process of wave propagation*. Therefore A_1 and A_1' are corresponding points as are A_2 and A_2'. Malus's theorem further assumes that the time required by the wave to propagate from A_1 to A_1' on wave surfaces S and S' must be the same as the time required to go from A_2 to A_2' on the same pair of wave surfaces.

13.4 Reflection and Refraction of Plane Waves

Consider a plane wave propagating in medium (1) in the direction of the unit vector $\boldsymbol{u}_i$ as the wave approaches the boundary with another medium (Fig. 13-5). When the wave reaches the plane surface AB that separates medium (1) from medium (2), a wave is transmitted into the second medium and another wave is reflected back into medium (1). These are the *refracted* and *reflected* waves, respectively. When the angle of incidence is oblique, the refracted waves propagate in a direction indicated by unit vector $\boldsymbol{u}_r$, which is different from $\boldsymbol{u}_i$; and the reflected waves propagate in a direction, indicated by unit vector $\boldsymbol{u}_r'$, which is symmetric to $\boldsymbol{u}_i$ with respect to the surface

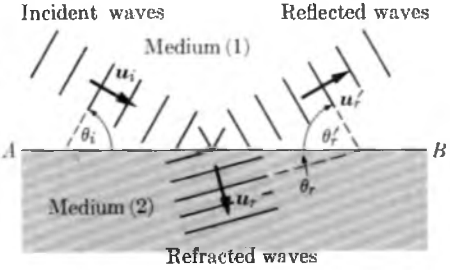
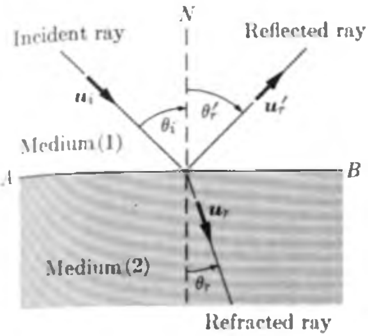
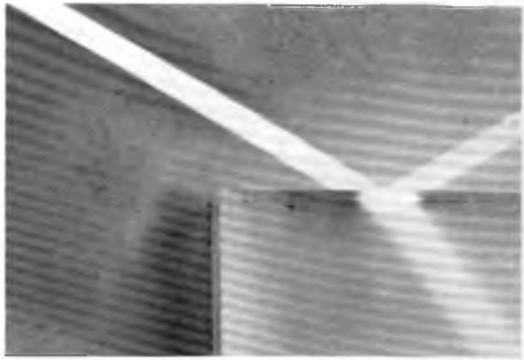


Fig. 13-5. Incident, reflected, and refracted plane waves.





(b)

Fig. 13-6. (a) Incident, reflected, and refracted rays. (b) A pencil of light reflected from and refracted into a glass block. (From *Physics*: Boston: D. C. Heath, 1960.)

normal. Figure 13-6 indicates the corresponding situation for rays. The angles θ_i, θ_r, and θ_r' are called the angles of *incidence*, *refraction*, and *reflection*, respectively. The directions of the three vectors $\boldsymbol{u}_i$, $\boldsymbol{u}_r$ and $\boldsymbol{u}_r'$ are related by the following experimentally verified laws.

(1) *The directions of incidence, refraction, and reflection are all in one plane*, which is normal to the surface of separation and therefore contains the normal N to the surface.

(2) *The angle of reflection is equal to the angle of incidence.* That is,

$$\theta_r' = \theta_i. \tag{13.3}$$

(3) *The ratio between the sine of the angle of incidence and the sine of the angle of refraction is constant.* This is called *Snell's law*, named after the Dutch mathematician Willebrod Snell von Royen (1591–1626), and is expressed by

$$\frac{\sin \theta_i}{\sin \theta_r} = n_{21}. \tag{13.4}$$

The constant n_{21} is called the *index of refraction* of medium (2) relative to medium (1). The numerical value of the constant depends on the nature of the wave and on the properties of the two media.

These laws remain valid when neither the wave surface nor the interface is plane because at each point a limited section of either surface can be considered as plane and the rays at that point behave according to Eqs. (13.3) and (13.4).

The three laws can be verified experimentally without great difficulty. They can also be proved theoretically if the basic concepts of wave propagation, and in particular the theorem of Malus, are used. For example the first law can be justified on the basis of symmetry considerations alone since the incident ray and the normal N determine a plane, and there is no *a priori* reason for the refracted and the reflected

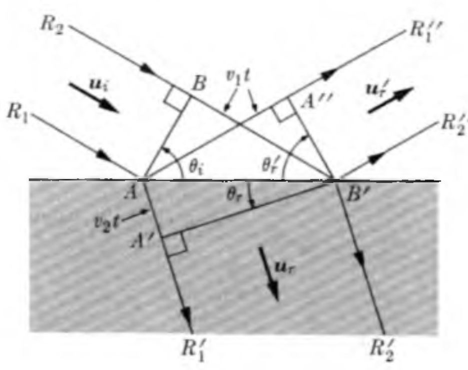


Figure 13-7

rays to be deflected from this plane. To prove the second and third laws, consider two incident rays R_1 and R_2 (Fig. 13-7) that are parallel since the incident waves are plane. Ray R_1 hits the interface at A, and R_2 hits the surface at B'. Because the geometrical situation is the same at A and at B', we conclude that the refracted rays R_1' and R_2' are also parallel as are the reflected rays R_1'' and R_2''. Since the rays R_1 and R_2 were chosen arbitrarily, the refracted and reflected waves are also plane because they must be perpendicular to a corresponding set of parallel rays as required by Malus's theorem.

Now consider the following wave surfaces: AB in the incident wave, $A'B'$ in the refracted wave. and $A''B'$ in the reflected wave so that A, A' and B, B' are two sets of corresponding points for the refracted wave. and A, A'' and B, B'' are two sets of corresponding points for the reflected wave. As implied in Malus's theorem, the rays between corresponding points in the wave surfaces must have been traversed in equal times. Call t the time taken by the incident wave to go from B to B' along ray R_2 with velocity v_1. In the same time the refracted wave moved along ray R_1' from A to A' with velocity v_2, and the reflected wave moved along ray R_1'' from A to A'' with velocity v_1. Then

$$BB' = v_1 t, \qquad AA' = v_2 t, \qquad AA'' = v_1 t;$$

and from the geometry of the figure,

$$\sin\theta_i = \frac{BB'}{AB'} = \frac{v_1 t}{AB'},$$

$$\sin\theta_r = \frac{AA'}{AB'} = \frac{v_2 t}{AB'},$$

$$\sin\theta_r' = \frac{AA''}{AB'} = \frac{v_1 t}{AB'}.$$

Comparing the first and third relations, $\sin\theta_r' = \sin\theta_i$ or $\theta_r' = \theta_i$, which is the law for reflection, Eq. (13.3). Dividing the first relation into the second yields

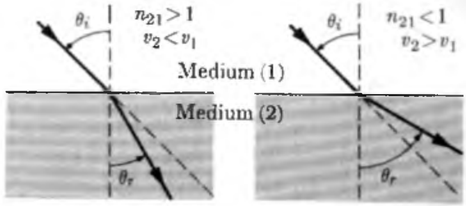


Fig. 13-8. Refracted rays for (a) $n_{21}>1$ and (b) $n_{21}<1$.

$$\frac{\sin\theta_i}{\sin\theta_r}=\frac{v_1}{v_2},$$

which expresses Snell's law, Eq. (13.4), since the ratio v_1/v_2 between the two velocities of propagation is constant. Comparing this equation with Eq. (13.4) shows that the relative index of refraction of two substances is equal to the ratio of the velocities of propagation of the wave in the substances, or

$$n_{21}=\frac{v_1}{v_2}. \tag{13.5}$$

For each kind of wave motion a particular medium is chosen as the reference or standard, and the velocity of propagation of the wave in that medium is designated by c. The *index of refraction* of any other medium for that wave motion is defined by

$$n=\frac{c}{v}. \tag{13.6}$$

For electromagnetic waves the reference medium is vacuum, and thus $c\approx3\times10^8$ m s^{-1}. Now for two substances

$$\frac{n_2}{n_1}=\frac{c}{v_2}\times\frac{v_1}{c}=\frac{v_1}{v_2}=n_{21} \tag{13.7}$$

so that the relative index of refraction of two substances is equal to the ratio of their respective indexes of refraction. Using relation (13.7), we can write Snell's law, Eq. (13.4), in more symmetric form;

$$n_1\sin\theta_i=n_2\sin\theta_r. \tag{13.8}$$

Depending on whether $v_2 \lessgtr n_1$, then $n_2 \gtrless n_1$ and $n_{21} \gtrless 1$, resulting in $\theta_i \gtrless \theta_r$ as indicated in Fig. 13-8.

In the second case, $n_{21}<1$, a special situation arises when

$$\sin\theta_i=n_{21}; \tag{13.9}$$

then $\sin\theta_r=1$ or $\theta_r=\pi/2$, indicating that the refracted ray is parallel to the surface. The angle θ_i given by Eq. (13.9) is called the *critical angle* and is designated by θ_c. The geometrical situation is illustrated in Fig. 13-9. When $n_{21}<1$, and $\theta_i>\theta_c$ or

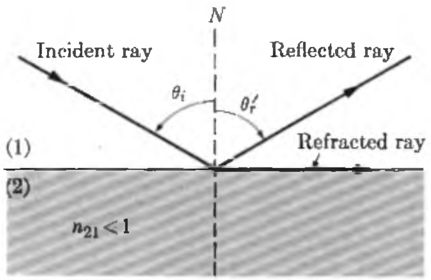


Fig. 13-9. Total reflection occurs when $n_{21}<1$ and θ_i is larger than the critical angle λ.

$\sin\theta_i > n_{21}$, it then follows that $\sin\theta_r > 1$, which is impossible. Therefore in this case there is no refracted ray; there is *total reflection*. This situation may come about for example when light passes from glass into air. Strictly speaking, as shown in Fig. 13-9, there is a wave propagating in the second medium parallel to the surface; but the amplitude of the wave decreases very rapidly with depth, and the wave is confined to a *very* thin layer along the surface.

Example 13.1. Passage of a wave through a medium limited by plane parallel sides.

▼ Consider a plate of thickness a and a ray AB (Fig. 13-10) whose angle of incidence is θ_i. Ignore the reflected ray. The angle of refraction is θ_r, corresponding to the refracted ray BC. From relation (13.8)

$$n_1 \sin\theta_i = n_2 \sin\theta_r.$$

At C the refraction is from medium (2) into medium (1) so that Eq. (13.8) gives

$$n_2 \sin\theta_i' = n_1 \sin\theta_r'.$$

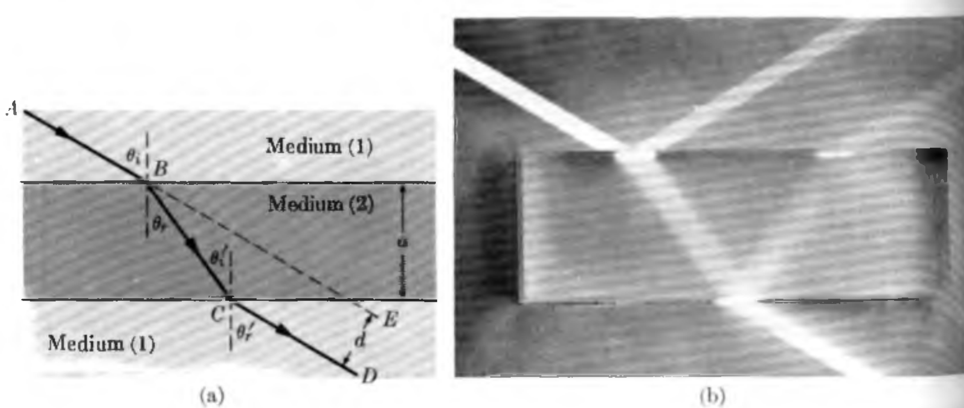


Fig. 13-10. Propagation of a ray through a parallel plate. In (b) the photograph shows the displacement of a pencil of light. (From *Physics*; Boston: D. C. Heath, 1960.)

Also from the geometry of Fig. 13-10, $\theta_i' = \theta_r$. Therefore, combining the two relations gives $\sin \theta_i = \sin \theta_r'$ or $\theta_i = \theta_r'$, which says that ray CD is parallel to ray AB, but is laterally displaced. We leave it to the student to verify that the lateral displacement of the ray is

$$d = a \frac{\sin(\theta_i - \theta_r)}{\cos \theta_r}.$$

It may also be easily verified that if instead of one plate, there are several parallel plates of different materials, the emergent and the incident rays are still parallel. ▲

13.5 Reflection and Refraction of Spherical Waves

The reflection and the refraction of *spherical* waves at a plane surface will now be examined. Consider spherical waves generated at a point source O and incident on a plane surface S. Two new sets of waves are then produced: the reflected and the refracted or transmitted as shown in Fig. 13-11. In order to trace the shape of the reflected and refracted wave fronts, it would be necessary to draw many reflected and refracted rays. The corresponding reflected and refracted wave surfaces are perpendicular to the rays. In Fig. 13-11, one set of these rays has been drawn at B on the assumption that $n_{21} > 1$. According to laws (2) and (3) for reflection and refraction, we have

$$\theta_r' = \theta_i, \qquad \frac{\sin \theta_i}{\sin \theta_r} = n_{21}.$$

When extended back into medium (2), the reflected ray BD intersects the extended normal AO at point I'. Because triangles OAB and $I'AB$ are right triangles and the angles at O and I' are the same, we see that $AO = AI'$. Since B is an arbitrary point, we conclude that *all reflected rays pass through a point I', symmetric from O relative to the plane surface.* This point is called the *image* of O produced by reflection.

Therefore *when spherical waves fall on a plane surface, the reflected waves are spherical and symmetrical with respect to the incident waves.* This symmetry was to be expected because the reflected waves propagate backward with the same velocity as the incident waves in such a way that the reflected waves remain symmetric relative to the reflecting surface.

When the direction of the refracted ray BC is extended back into medium (1) the ray intersects the normal OA at a point I such that $\tan \theta_r = AB/AI$. Since $\tan \theta_i = AB/AO$, the ratio of the two tangents is

$$\frac{\tan \theta_i}{\tan \theta_r} = \frac{AI}{AO}$$

or

$$AI = AO \frac{\tan \theta_i}{\tan \theta_r}. \tag{13.10}$$

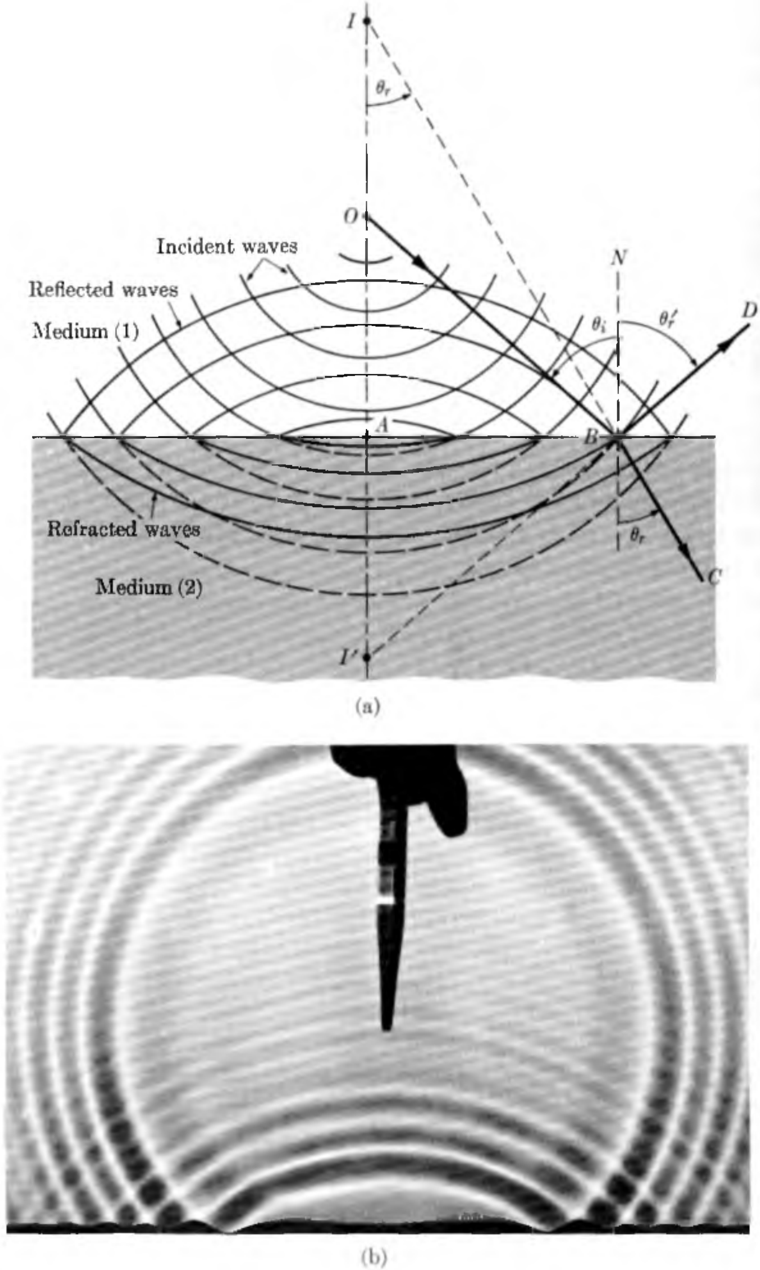


Fig. 13-11. (a) Incident, reflected, and refracted spherical waves. In (b), the photograph shows incident and reflected surface waves in a liquid medium. (From *Physics*; Boston: D. C. Heath, 1960.)

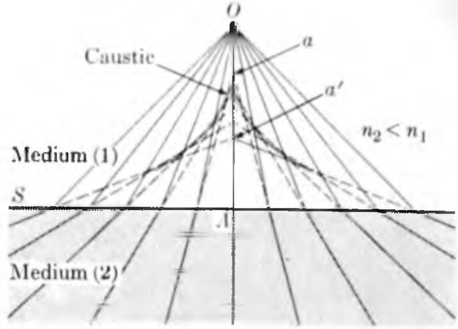


Fig. 13-12. Refraction of rays that proceed from a point source. When extended back, the refracted rays do not intersect at a single point.

Recall that Snell's law of refraction requires that $\sin\theta_i/\sin\theta_r$ be constant and equal to n_{21}; but then $\tan\theta_i/\tan\theta_r$ cannot be constant. Therefore the refracted rays do not all pass through the same point. We conclude then that *when spherical waves fall on a plane surface, the refracted waves are not spherical.*

Since the refracted rays do not pass through a single point, they do not form a point image of O as the reflected rays do. The refracted rays intersect at several points along the normal OA as well as on a conical surface called the *caustic*, shown in Fig. 13-12. The intersection of reflected rays can be observed without difficulty in the case of the refraction of light waves. The point a, formed by the intersection of the least-inclined rays, can be calculated very easily because then the angles θ_i and θ_r of Fig. 13-11 are very small and the tangents may be replaced by sines in Eq. (13.10); the result is

$$AI \approx AO\frac{\sin\theta_i}{\sin\theta_r} = n_{21}AO \qquad \text{(for small angles).} \tag{13.11}$$

13.6 More about the Laws of Reflection and Refraction

We have established the laws of reflection and refraction by using a geometrical reasoning based on Malus's theorem. We shall now discuss these laws in a more analytical form. Suppose that an incident wave is described by an equation of type (10.63). That is,

$$\xi_i = \xi_{0i}\sin(\boldsymbol{k}_i\cdot\boldsymbol{r} - \omega t). \tag{13.12}$$

The refracted and reflected waves will be respectively

$$\xi_r = \xi_{0r}\sin(\boldsymbol{k}_r\cdot\boldsymbol{r} - \omega t) \tag{13.13}$$

and

$$\xi_r' = \xi_{0r}'\sin(\boldsymbol{k}_r'\cdot\boldsymbol{r} - \omega t). \tag{13.14}$$

Note that the same ω is used in both the reflected and the refracted waves as in the incident wave because it is an experimental fact that the frequency of wave motion does not change on reflection or refraction.

The physical property (a displacement, or a pressure, or an electric or magnetic field) ascribed to ξ is such that the value at the surface separating the two media must be the same from whichever side it is observed. (In the case of an electromagnetic wave, the relation among the components of the electric field and the magnetic field may be of a somewhat different nature; but the relation is still linear and involves the fields on both sides of the surface.) In medium (1) we have the incident and reflected waves, which yield $\xi_i + \xi'_r$ for the resulting disturbance at the surface; in medium (2) we have only the refracted wave, which yields ξ_r. Then at the surface of separation,

$$\xi_i + \xi'_r = \xi_r. \tag{13.15}$$

In order that this be satisfied at all points and times at the surface of separation, it is necessary that the phases in Eqs. (13.12), (13.13), and (13.14) be identical. That is,

$$\boldsymbol{k}_i \cdot \boldsymbol{r} - \omega t = \boldsymbol{k}_r \cdot \boldsymbol{r} - \omega t = \boldsymbol{k}'_r \cdot \boldsymbol{r} - \omega t \tag{13.16}$$

for all points $\boldsymbol{r}$ on the surface (Fig. 13-13). After the common term ωt is cancelled, Eq. (13.16) reduces to

$$\boldsymbol{k}_i \cdot \boldsymbol{r} = \boldsymbol{k}_r \cdot \boldsymbol{r} = \boldsymbol{k}'_r \cdot \boldsymbol{r}. \tag{13.17}$$

If the XYZ axes are chosen as indicated in Fig. 13-13 so that the surface of separation coincides with the XZ-plane and the direction of incidence lies in the XY-plane, then $\boldsymbol{r}$ must be in the XZ-plane and given by $\boldsymbol{r} = \boldsymbol{u}_x x + \boldsymbol{u}_z z$. Similarly $\boldsymbol{k}_i = \boldsymbol{u}_x k_{ix} + \boldsymbol{u}_y k_{iy}$; and since we do not know *a priori* whether $\boldsymbol{k}_r$ and $\boldsymbol{k}'_r$ are also in the same plane, we

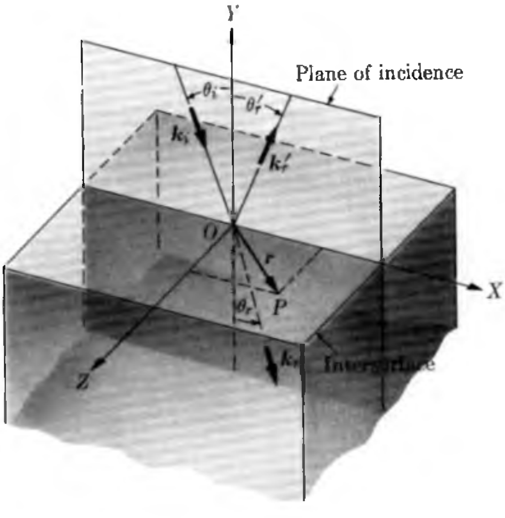


Figure 13-13

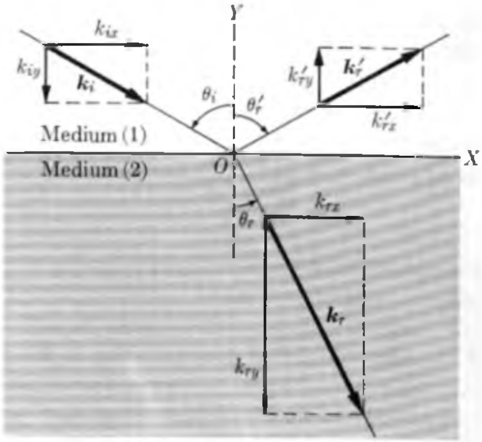


Fig. 13-14. Propagation vectors in incident, reflected, and refracted waves.

must write $\boldsymbol{k}_r = \boldsymbol{u}_x k_{rx} + \boldsymbol{u}_y k_{ry} + \boldsymbol{u}_z k_{rz}$ and $\boldsymbol{k}_r' = \boldsymbol{u}_x k_{rx}' + \boldsymbol{u}_y k_{ry}' + \boldsymbol{u}_z k_{rz}'$. Substituting these expressions into Eq. (13.17) and recalling the expression for the scalar product gives

$$k_{ix}x = k_{rx}x + k_{rz}z = k_{rx}'x + k_{rz}'z.$$

Because this relation must hold for all points on the plane XZ, it follows that

$$k_{ix} = k_{rx} = k_{rx}' \qquad \text{and} \qquad k_{rz} = k_{rz}' = 0. \tag{13.18}$$

The second group of equations indicates that the vectors $\boldsymbol{k}_r$ and $\boldsymbol{k}_r'$ have no components along the Z-axis so that the vectors are also in the XY-plane; and the incident, the reflected, and the refracted rays are in the same plane; this is the first law (1) mentioned in Section 13.4.

Next we see, from Fig. 13-14 that $k_{ix} = k_i \sin\theta_i$, $k_{rx} = k_r \sin\theta_r$, and $k_{rx}' = k_r' \sin\theta_r'$; and from Eq. (10.6), $k_i = k_r' = \omega/v_1$ and $k_r = \omega/v_2$. Using all these relations in the first group of equations of (13.18) and factoring out the common factor ω gives

$$\frac{1}{v_1}\sin\theta_i = \frac{1}{v_2}\sin\theta_r = \frac{1}{v_1}\sin\theta_r'.$$

From these relations we see that $\sin\theta_i = \sin\theta_r'$ or $\theta_i = \theta_r'$ and $\sin\theta_i/\sin\theta_r = v_1/v_2 = n_{21}$ immediately follow. Thus we recover the remaining two laws for reflection and refraction originally developed in Section 13.4.

When Eq. (13.16) is satisfied, Eq. (13.15) reduces to

$$\xi_{0i} + \xi_{0r}' = \xi_{0r}, \tag{13.19}$$

which is a relation among the amplitudes of the three waves. If Eq. (13.15), or its equivalent, Eq. (13.19), is the only requirement that must be satisfied, there is not enough information to determine the amplitude of the reflected and refracted waves. However because of the physics of the problem, another boundary condition is usually required, such as the continuity of the stresses or of the pressure across the interface in the case of elastic waves, or the continuity of certain components of the electric and magnetic fields in the case of electromagnetic waves. Therefore a second relation (or boundary condition) involving the amplitudes ξ_{0i}, ξ_{0r}, and ξ'_{0r} must exist. If the two boundary conditions are used, the amplitudes ξ_{0r} and ξ'_{0r} in terms of ξ_{0i} can be determined as illustrated in the following example.

Example 13.2. Reflection and transmission of transverse waves at a point where two strings of different materials are joined.

▼ Consider two strings (1) and (2) (Fig. 13-15), attached at one point: this point will be chosen as the origin of coordinates. The strings are subject to the same tension T. For mathematical convenience use the alternative form for waves given in Eq. (10.10). There is an incident wave coming from the left, propagating with the velocity $v_1 = \omega/k_1$, and having the form

$$\xi_i = \xi_{0i} \sin(\omega t - k_1 x).$$

At the point of discontinuity, a refracted or transmitted wave propagates with the velocity $v_2 = \omega/k_2$ along string (2), and has the form

$$\xi_r = \xi_{0r} \sin(\omega t - k_2 x).$$

A reflected wave, propagating back along string (1) with velocity ω/k_1 with the form

$$\xi'_r = \xi'_{0r} \sin(\omega t + k_1 x),$$

is also produced. Note that k_1 is used for the reflected and the incident waves because they propagate in the same medium: string (1). The vertical displacement at any point on string (1) is $\xi = \xi_i + \xi'_r$. On string (2) the vertical displacement is $\xi = \xi_r$. Point O, where the strings are joined, corresponds to $x = 0$. At this point, in conformity with Eq. (13.15), we must have $\xi_i + \xi'_r = \xi_r$, which becomes

$$\xi_{0i} \sin \omega t + \xi'_{0r} \sin \omega t = \xi_{0r} \sin \omega t$$

or

$$\xi_{0i} + \xi'_{0r} = \xi_{0r}. \tag{13.20}$$

This equation is a condition relating the amplitudes that is similar to Eq. (13.19). To obtain a second relation between the amplitudes, we follow the discussion in Section 10.7, and note that the vertical force at any point in string (1) is

$$F_y = T \sin \alpha \approx T \tan \alpha = T \frac{\partial \xi}{\partial x} = T\left(\frac{\partial \xi_i}{\partial x} + \frac{\partial \xi'_r}{\partial x}\right)$$

since α is small and $\sin \alpha$ is practically equal to $\tan \alpha$. Then, taking derivatives of ξ_i and ξ'_r, we can write

$$F_y = Tk_1[-\xi_{0i} \cos(\omega t - k_1 x) + \xi'_{0r} \cos(\omega t + k_1 x)].$$

Similarly the vertical force at any point on string (2) is found to be

$$F_y = T \frac{\partial \xi_r}{\partial x} = -Tk_2 \xi_{0r} \cos(\omega t - k_2 x).$$

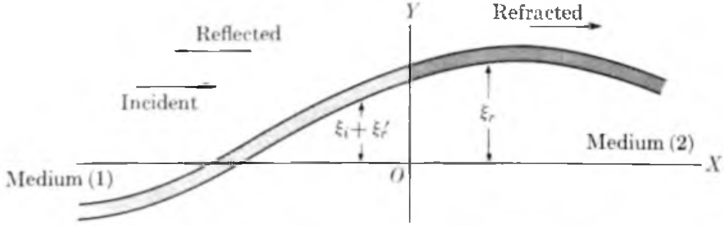


Fig. 13-15. Transverse waves in two attached strings of different linear densities.

Now at the junction the vertical force must be the same, whether the computation uses T_y for string (1) or for string (2). Then setting $x=0$ in the two expressions above for T_y, equating them, and canceling the common factor cos ωt gives

$$k_1(\xi_{0i}-\xi'_{0r})=k_2\xi_{0r}. \tag{13.21}$$

This equation is a second condition to be satisfied by the three amplitudes and is imposed by the physical nature of the wave. Solving the system of equations (13.20) and (13.21) yields

$$\xi_{0r}=\left[\frac{2k_1}{k_1+k_2}\right]\xi_{0i}, \qquad \xi'_{0r}=\left[\frac{k_1-k_2}{k_1+k_2}\right]\xi_{0i}, \tag{13.22}$$

which determine the amplitudes of the refracted and reflected waves relative to the amplitude of the incident wave. Noting that $k=\omega/v$, we may write instead

$$\xi_{0r}=\left[\frac{2v_2}{v_1+v_2}\right]\xi_{0i}, \qquad \xi'_{0r}=\left[\frac{v_2-v_1}{v_1+v_2}\right]\xi_{0i}. \tag{13.23}$$

Since in the case of transverse waves in a string $v=\sqrt{T/\sigma}$, according to Eq. (10.37) where σ is the mass per unit length, we may also write

$$\xi_{0r}=\left[\frac{2\sqrt{\sigma_1}}{\sqrt{\sigma_1}+\sqrt{\sigma_2}}\right]\xi_{0i}, \qquad \xi'_{0r}=\left[\frac{\sqrt{\sigma_1}-\sqrt{\sigma_2}}{\sqrt{\sigma_1}+\sqrt{\sigma_2}}\right]\xi_{0i}. \tag{13.24}$$

The ratios ξ_{0r}/ξ_{0i} and ξ'_{0r}/ξ_{0i} are called the *coefficients of refraction* (or transmission) *and of reflection*, respectively, and are designated by $\mathfrak{t}$ and $\mathfrak{r}$, respectively. Thus

$$\mathfrak{t}=\frac{2\sqrt{\sigma_1}}{\sqrt{\sigma_1}+\sqrt{\sigma_2}}, \qquad \mathfrak{r}=\frac{\sqrt{\sigma_1}-\sqrt{\sigma_2}}{\sqrt{\sigma_1}+\sqrt{\sigma_2}}. \tag{13.25}$$

Note that $\mathfrak{t}$ is always positive so that ξ_{0r} always has the same sign as ξ_{0i} and the transmitted wave is always in phase with the incident wave. But $\mathfrak{r}$ is positive or negative depending on whether $\sigma_1 \gtrless \sigma_2$ so that the reflected wave may be in phase with or in opposition to the incident wave. In the second case the reflected wave has suffered a phase shift of π relative to the incident wave. The two situations are illustrated in Fig. 13-16.

The student may check the flow of energy across the junction by using the rate of energy flow on strings (1) and (2). The energy transmitted is proportional to $\mathfrak{t}^2$ and the energy reflected is proportional to $\mathfrak{r}^2$. ▲

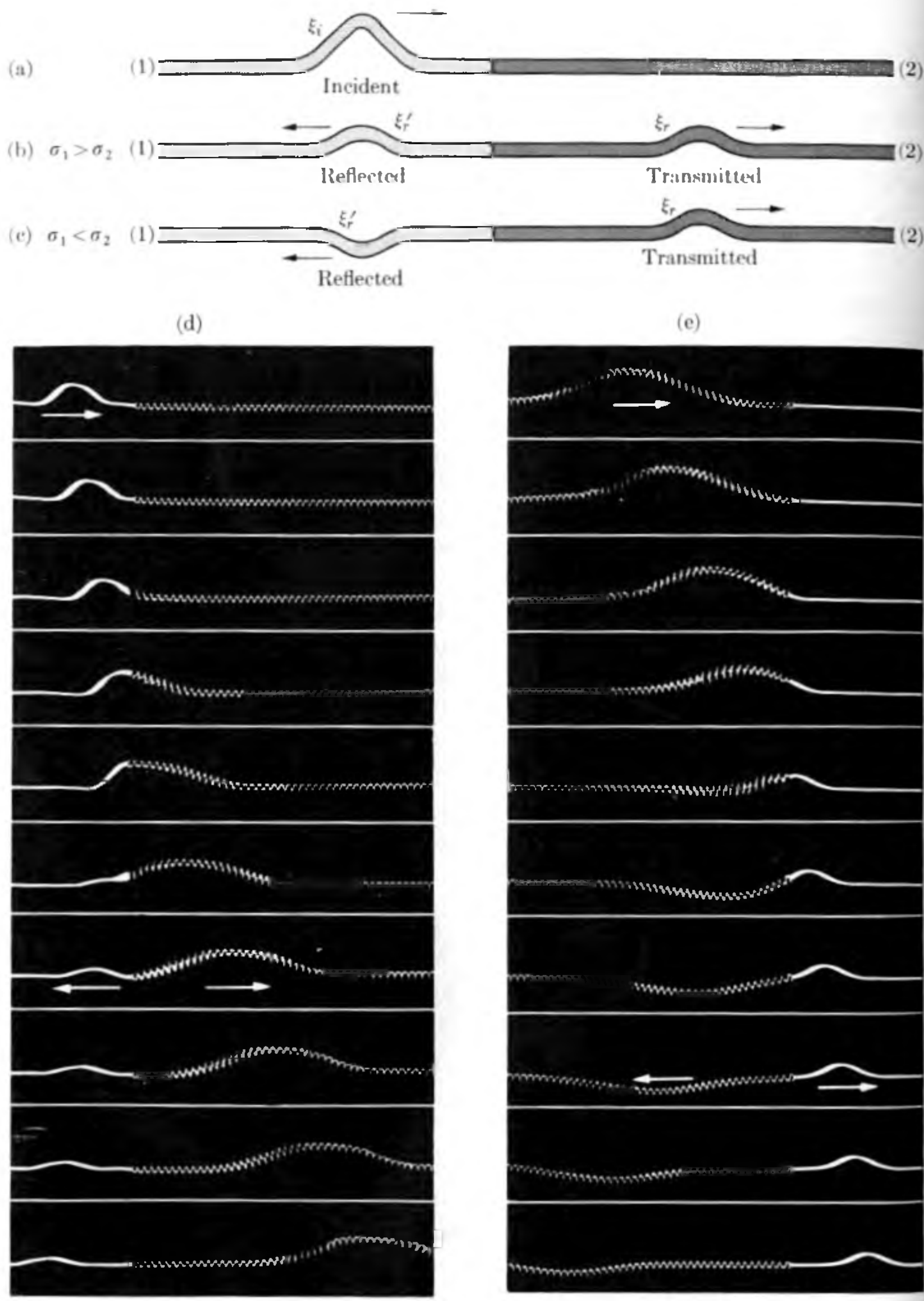


Fig. 13-16. Incident, reflected, and transmitted waves in two attached strings of different linear densities. In (b) and (d) the string carrying the incident wave is heavier; in (c) and (e), the string on the left is lighter. (Photographs from *Physics*; Boston: D. C. Heath, 1960.)

13.7 Reflection and Refraction at Metallic Surfaces

It was demonstrated in Section 2.5 that inside a conductor a static electric field is zero. The situation is not entirely the same when the electric field is time dependent. However even when the electric field is time dependent, an electromagnetic wave is greatly attenuated when it propagates in a conductor, such as a metal or an ionized gas. We shall not present the detailed theory here, but we shall indicate one of the fundamental changes that take place in the equations regulating the propagation of an electromagnetic wave in a conductor.

Equations (11.1) through (11.5) remain the same, but Eq. (11.6) must be modified to take into account the currents induced in the conductor by the electric field of the wave. From Eq. (3.41) we saw that the current density is $\boldsymbol{j}=\sigma\boldsymbol{\mathscr{E}}$ where σ is the electrical conductivity of the metal. When this current is incorporated into Eq. (11.6), a straightforward manipulation gives, for the equation satisfied by the electric field.

$$\frac{\partial^2 \mathscr{E}}{\partial x^2}=\epsilon\mu\frac{\partial^2 \mathscr{E}}{\partial t^2}+\mu\sigma\frac{\partial \mathscr{E}}{\partial t}, \tag{13.26}$$

instead of the simpler wave equation (11.7). The new term, $\mu\sigma\,\partial\mathscr{E}/\partial t$, since it is a first-order time derivative, is similar to the damping term $-\lambda\,dx/dt$ in a damped oscillator. Therefore the addition of this term indicates that the wave is damped while it progresses through the metal. Thus the intensity of the wave decreases rapidly as it penetrates the conductor. The solution of Eq. (13.26) can be expressed in the form

$$\mathscr{E}=\mathscr{E}_0 e^{-\alpha x}\sin(kx-\omega t) \tag{13.27}$$

where the velocity of propagation $v=\omega/k$ and the damping coefficient α are expressed by some complicated algebraic relations among μ, ϵ, and σ. The exponential in Eq. (13.27) indicates that the wave is damped while it is progressing into the conducting medium. When the frequency is small so that ω^2 can be neglected, and the material is a very good conductor so that $\sigma \gg \epsilon\omega$, the student may verify by direct substitution of Eq. (13.27) in Eq. (13.26) that

$$k=\alpha\approx\sqrt{\tfrac{1}{2}\mu\sigma\omega}. \tag{13.28}$$

The velocity of propagation is then

$$v=\frac{\omega}{k}=\sqrt{\frac{2\omega}{\mu\sigma}}. \tag{13.29}$$

This analysis explains two important characteristics concerning conductors. One characteristic is their opacity, resulting from the strong absorption of the waves so that no wave is transmitted through the conductor unless it is a very thin sheet. Conductors are therefore excellent for shielding a region from electromagnetic waves.

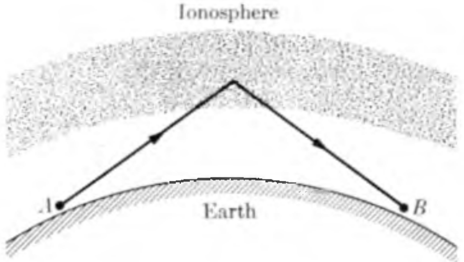


Fig. 13-17. Reflection of radio waves by the ionosphere.

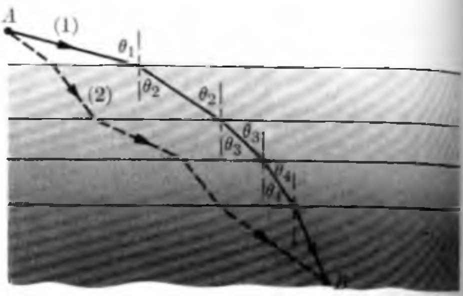


Fig. 13-18. Path of a ray in a stratified medium.

(This shielding is done by surrounding the region by a metal grid for example.) The other characteristic of conductors is their great reflectivity, which results because only a small fraction of the energy of the incident wave penetrates the conductor, and most of the energy goes into the reflected wave. This high reflectivity is typical of metals.

An ionized layer of gas can also act as a conductor, reflecting electromagnetic waves falling on the layer. This phenomenon is used in radio communication to transmit a radio signal around the earth. The signal is reflected back to the earth when the signal reaches a highly ionized layer in the atmosphere, called the *ionosphere*, which is 80 to 100 km above the earth's surface. In this way communication between two widely separated points A and B is possible even if a wave cannot propagate in a straight line between the points (Fig. 13-17).

13.8 Propagation in a Nonhomogeneous Medium; Fermat's Principle

The reflection and the refraction phenomena described in the previous sections correspond to a situation in which a wave passes from one homogeneous medium to another. However in many instances a wave propagates in a medium whose properties vary from point to point. For example on a hot day the lower layers of air are much warmer than the upper ones, and sound waves as well as light waves suffer a continuous refraction.

Consider the propagation of a wave through a stratified medium [ray (1) in Fig. 13-18]; that is, through a medium composed of several layers in which the velocity of propagation is different. If a wave reaches the first surface with an angle of incidence θ_1, the successive refractions satisfy the conditions

$$n_1 \sin \theta_1 = n_2 \sin \theta_2,$$

$$n_2 \sin \theta_2 = n_3 \sin \theta_3,$$

or

$$n \sin \theta = \text{const.} \tag{13.30}$$

Next consider a medium in which the index of refraction depends on one coordinate, say y. This medium can be considered as a stratified medium in which successive layers are very thin. Then Eq. (13.30) still holds, and we may write

$$n(y) \sin \theta = C, \tag{13.31}$$

where C is a constant. This expression gives the angle θ at every point on the ray's path, and the path of the ray can be traced through the inhomogeneous medium by applying Eq. (13.31).

Another, rather elegant method that yields the path of a ray in an inhomogeneous medium is the method suggested by the French mathematician Pierre de Fermat (1601–1665). *Fermat's principle* can be stated in the following way:

> *in traveling from one point to another, a ray chooses the path for which the propagation time has a minimum value.*

The actual path for the case given in Fig. 13-18 is computed by using Eq. (13.31) and is shown by line (1). Another arbitrary path is indicated by line (2). This path is not a physically possible path since Snell's law is not satisfied at each boundary between two surfaces. We may compute the times required for the ray to follow each of the two paths if we know the length of each segment of the path and the velocity of propagation in each medium. Fermat's principle stipulates that the time for the actual physical path will be smaller than the time needed for any arbitrary and nonphysical neighboring path; that is, $t_1 < t_2$.

It should be clear that the time required by the ray to go from any point A to any other point B along a path is a function of the path; that is, $t_{AB} = f(\text{path})$. This is a new type of functional dependence in that the variables in the function f are not the coordinates of a point, but the parameters defining a path joining A and B. The requirement that t_{AB} be a minimum can be stated by saying that $d(t_{AB}) = 0$ for a small change in the values of the parameters corresponding to the physical path. A special mathematical technique known as the *calculus of variations* permits finding the values of the parameters of the path satisfying $d(t_{AB}) = 0$; and in this way the path of the ray can be determined.

We shall not elaborate further on how Fermat's principle can be used to trace the path of a ray in an inhomogeneous medium, but only recognize that Fermat's principle and Snell's law are compatible as is shown in the following example.

Example 13.3. Verification that Fermat's principle is compatible with Snell's law.

▼ Consider Fig. 13-19, in which a surface S separates two media of indexes of refraction n_1 and n_2. A ray that travels from A to C follows the path ABC. Then, recalling that $v = c/n$, the time

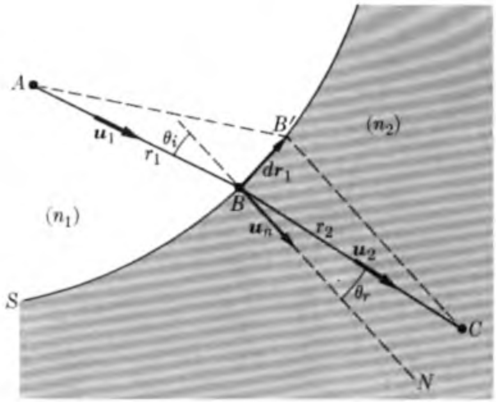


Figure 13-19

needed for light to traverse this path is

$$t=t_{AB}+t_{BC}=\frac{r_1}{v_1}+\frac{r_2}{v_2}=\frac{1}{c}(n_1r_1+n_2r_2).$$

Fermat's principle requires that

$$d(t)=\frac{1}{c}(n_1\ dr_1+n_2\ dr_2)=0 \tag{13.32}$$

where $d(t)$ is the change in t for neighboring paths such as $AB'C$ produced by the corresponding changes dr_1 and dr_2 in r_1 and r_2, respectively. Now remembering that $r^2=\boldsymbol{r}\cdot\boldsymbol{r}$, we have

$$r\ dr=\boldsymbol{r}\cdot d\boldsymbol{r} \qquad \text{or} \qquad dr=\left(\frac{\boldsymbol{r}}{r}\right)\cdot d\boldsymbol{r}=\boldsymbol{u}\cdot d\boldsymbol{r}$$

where $\boldsymbol{u}=\boldsymbol{r}/r$ is the unit vector in the direction of $\boldsymbol{r}$. Therefore $dr_1=\boldsymbol{u}_1\cdot d\boldsymbol{r}_1$ and $dr_2=\boldsymbol{u}_2\cdot d\boldsymbol{r}_2$; but since $\boldsymbol{r}_1+\boldsymbol{r}_2=\overrightarrow{AB}+\overrightarrow{BC}=\overrightarrow{AC}=$const., therefore $d\boldsymbol{r}_1+d\boldsymbol{r}_2=0$ or $d\boldsymbol{r}_2=-d\boldsymbol{r}_1$ so that $dr_2=-\boldsymbol{u}_2\cdot d\boldsymbol{r}_1$. Then with the constant factor $1/c$ eliminated, Eq. (13.32) gives

$$(n_1\boldsymbol{u}_1-n_2\boldsymbol{u}_2)\cdot d\boldsymbol{r}_1=0. \tag{13.33}$$

As indicated in Fig. 13-19, the vector $d\boldsymbol{r}_1$ is in the plane tangent to S at B. According to Eq. (13.33), the vector $n_1\boldsymbol{u}_1-n_2\boldsymbol{u}_2$ is perpendicular to $d\boldsymbol{r}_1$ and thus parallel to the normal $\boldsymbol{u}_N$ to the surface S at the point B. The implication of this result is that the incident ray, the refracted ray, and the normal to the surface are all in one plane, which is the first law as given in Section 13.4. Recall that if two vectors are parallel their vector product is zero; therefore $(n_1\boldsymbol{u}_1-n_2\boldsymbol{u}_2)\times\boldsymbol{u}_N=0$ or

$$n_1\boldsymbol{u}_1\times\boldsymbol{u}_N=n_2\boldsymbol{u}_2\times\boldsymbol{u}_N. \tag{13.34}$$

Since all the vectors in the equation above are unit vectors, Eq. (13.34) implies (in magnitude) that $n_1\sin\theta_i=n_2\sin\theta_r$, which is Snell's law, Eq. (13.4). ▲

Example 13.4. Radius of curvature of a ray when the wave propagates in a medium with variable index of refraction.

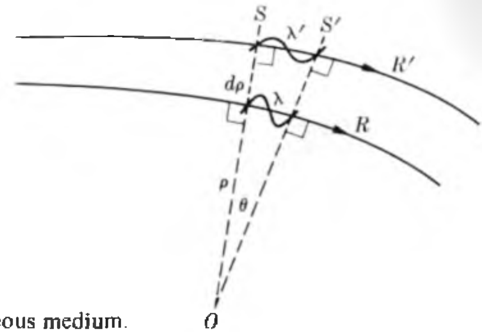


Fig. 13-20. Curvature of rays in a nonhomogeneous medium.

▼ Consider two very close rays R and R' (Fig. 13-20) and the two wave surfaces S and S', separated in time by one period. Then their separation in distance along any ray is one wavelength. Since the velocity of propagation varies from point to point, the wavelength is also variable because

$$\lambda = \frac{v}{\nu} = \frac{c}{\nu n}.$$

Let λ and λ' be the wavelengths along the corresponding rays. Then from Fig. 13-20 we see that $\rho\theta = \lambda$ and $(\rho + d\rho)\theta = \lambda'$. Then $\theta\, d\rho = \lambda' - \lambda = d\lambda$; but $\theta = \lambda/\rho$ so that $(\lambda/\rho)\, d\rho = d\lambda$ or

$$\begin{aligned} \frac{1}{\rho} &= \frac{1}{\lambda}\frac{d\lambda}{d\rho} = \frac{d}{d\rho}(\ln \lambda) \\ &= -\frac{d}{d\rho}(\ln n) \end{aligned} \tag{13.35}$$

since $\ln \lambda = \ln(c/\nu n) = \ln c - \ln \nu - \ln n$, and both c and ν are constant. Equation (13.35) indicates that the ray's path is curved so that its concavity is toward the direction in which the index of refraction increases. If the index of refraction is constant, the ray follows a straight-line path because then $\rho = \infty$. ▲

Problems

13.1 The following rule has been proposed to construct the refracted ray (Fig. 13-21). At the point of incidence, two circles of radii 1 and n are drawn (using arbitrary units). The incident ray is extended until it intersects the circle of radius 1. A perpendicular to the surface is drawn through that point, and the intersection of the perpendicular with the circle of radius n is found. The refracted ray passes through this

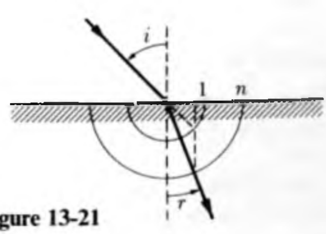


Figure 13-21

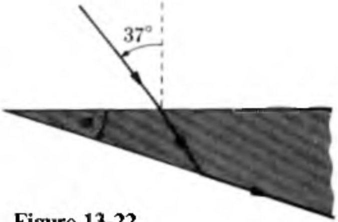


Figure 13-22

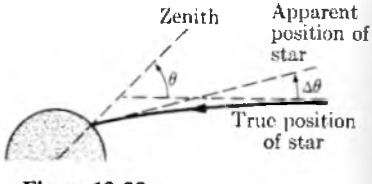


Figure 13-23

point. (a) Justify this rule. (b) Apply it to the case in which $n = 1.5$ and the angle of incidence i is 60°. (c) Repeat for $n = 0.80$ and an angle of incidence that is 30°, and another that is 60°. Verify your results by the use of Snell's law.

13.2 A plate of glass ($n = 1.6$) with parallel sides is 8×10^{-2} m thick. (a) Calculate the lateral displacement of a ray of light whose angle of incidence is 45°. (b) Using the method of Problem 13.1, plot the path of the ray.

13.3 A ray of light makes an angle of incidence of 35° with a plate of glass ($n = 1.3$) that is 6×10^{-2} m thick. Directly below the first plate is another plate whose index of refraction is 1.5. (a) What is the angle of incidence and angle of refraction at the boundary between the plates? (b) If the second plate is 5×10^{-2} m thick, determine the amount of lateral displacement of the ray.

13.4 A ray of light falls on a piece of glass ($n = 1.6$). The angle of incidence of the ray is 37°. What must be the minimum angle α between the first side of the glass and its second surface for total reflection? (See Fig. 13-22.)

13.5 A copper wire of radius 1×10^{-3} m is attached to a copper wire of radius 8×10^{-4} m. Find t and r at the junction for waves propagated along the system from the first to the second wire.

13.6 (a) For the situation discussed in Example 13.2, show that the intensity of the transmitted wave plus the intensity of the reflected wave add to the intensity of the incident wave. (b) What is the physical meaning of this result?

13.7 Estimate how deeply an electromagnetic wave penetrates copper when the amplitude of the wave decreases to $1/e$ of its value at the surface, if the frequency is (a) in the microwave region, 6×10^9 Hz, (b) in the visible region, 6×10^{14} Hz, and (c) in the X-ray region, 3×10^{18} Hz. Assume that $\mu \approx \mu_0$.

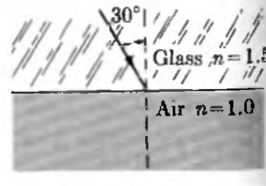
13.8 The index of refraction of air is $n = 1 + 0.00024\rho$ where ρ is the density of air (in kg m^{-3}). Let θ be the true zenith angle of a star and $\theta - \Delta\theta$ the apparent zenith angle with respect to an observer looking at the star through the atmosphere (Fig. 13-23). (a) Write the equation giving $\Delta\theta$ as a function of the true zenith angle θ, the density ρ, the atmospheric pressure p, and the absolute temperature T. (b) Compute $\Delta\theta$ at sea level for a star with $\theta = 45°$; assume a temperature $T = 298$ K.

CHALLENGING PROBLEMS

13.9 A light ray is incident on a glass-air interface as shown in Fig. 13-24. Find the sines of the angles of reflection and refraction. [AP-B; 1969]

13.10 A hollow box with transparent parallel walls is held under water as shown in Fig. 13-25.

Figure 13-24



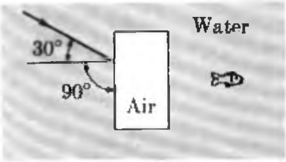


Figure 13-25

A light ray is incident upon the box from the water at an angle of 30° with respect to the normal. If the index of refraction of water is 1.33, determine the direction (or trigonometric function of the angle) of this ray in the box and again in the water as it leaves the box [AP-B: 1971]

13.11 If its index of refraction is large enough, a glass prism in air will totally reflect light that is normally incident on its base. For the prism shown in Fig. 13-26, determine the lowest index of refraction that will give total reflection. [AP-B; 1972]

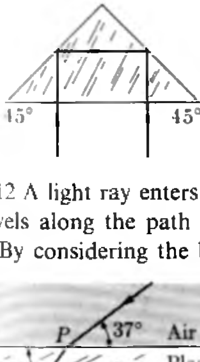


Figure 13-26

13.12 A light ray enters a block of plastic and travels along the path shown in Fig. 13-27a. (a) By considering the behavior of the ray at point P, determine the speed of light in the plastic. (b) Determine what will happen to the light ray when it reaches point Q; use a diagram like Fig. 13-27a to illustrate your conclusion. (c) There is an air bubble in the plastic block that happens to be shaped like a plano-convex lens as shown in Fig. 13-27b. Sketch what happens to parallel rays of light that strike this air bubble. Explain your reasoning. [AP-B; 1979]

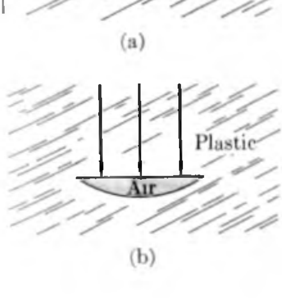


Figure 13-27

13.13 Copper and steel wires of the same radius are joined to make a long string. (a) Find t and r at the junction for waves propagated along the string. Let the common radius be 1×10^{-3} m. (b) Assuming that the incident wave has a frequency of 10 Hz, that the amplitude is 2×10^{-2} m, and that the tension is 50 N, write the equations for the incident, the reflected, and the transmitted waves. (Density of copper is 8.89×10^3 kg m^{-3}; of steel, 7.80×10^3 kg m^{-3}.)

13.14 An inhomogeneous stratified medium has an index of refraction that varies in the Y-direction, that is, $n(y)$. Show that the equation of the path of a ray satisfying Eq. (13.31) is

$$x = x_0 + \int_{y_0}^{y} \frac{dy}{\sqrt{n^2(y)/C^2 - 1}}.$$

13.15 The trajectory of a ray in a nonhomogeneous medium is represented by

$$x = A \sin\left(\frac{y}{B}\right).$$

Compute the index of refraction n in the space between the planes $x = A$ and $x = -A$; assume that n depends only on x and has the value n_0 at $x = 0$.

13.16 The index of refraction of a certain medium is given by $n = h + kx$. Compute the trajectory of the ray passing through the origin of the coordinate axes and forming at this point an angle ϕ_0 with the X-axis. Plot the trajectory of the ray; assume that $h = 1$, $k = 1$, and $\phi_0 = 45°$.

CHAPTER FOURTEEN

REFLECTION AND REFRACTION OF ELECTROMAGNETIC WAVES. POLARIZATION

14.1 Introduction

The case of reflection and refraction of electromagnetic waves requires special attention because it involves two fields: the electric and the magnetic components of the wave. Both electric and magnetic fields are perpendicular to the direction of propagation of each wave, but otherwise they may have any orientation around it. Thus when we discuss the reflection and refraction of electromagnetic waves, the analysis requires a consideration of the state of polarization of the wave. Such an analysis is easier if we think of each field as having a component parallel to the plane of incidence and a component perpendicular to the plane of incidence. This description then allows us to look at the effects of reflection and refraction on the polarization of a transverse wave.

14.2 Reflection and Refraction of Electromagnetic Waves

Consider an unpolarized beam of light falling on the plane interface between two media. The angle of incidence is θ_i and the angle of refraction is $\theta_{r'}$ ($\theta_{r'}=\theta_i$); the angle of refraction is θ_r and is related to θ_i by Snell's Law. At a given point in time the incident wave may be broken up into two perpendicular components; one with the electric field in the plane of incidence is designated $\mathscr{E}_\parallel$; the other with the electric field normal to the plane of incidence is designated $\mathscr{E}_\perp$. Because of the perpendicularity of

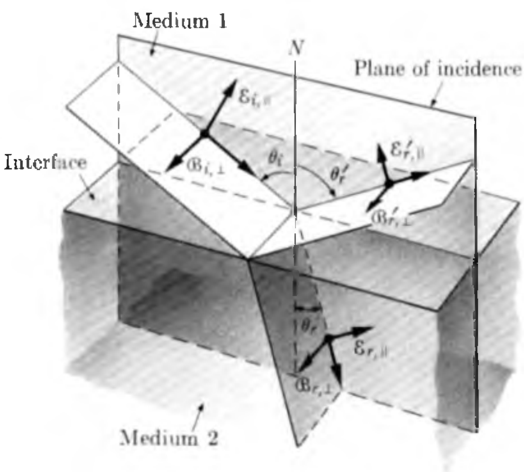


Fig. 14-1. Electric and magnetic fields in the incident, reflected, and refracted waves for polarization parallel to the plane of incidence.

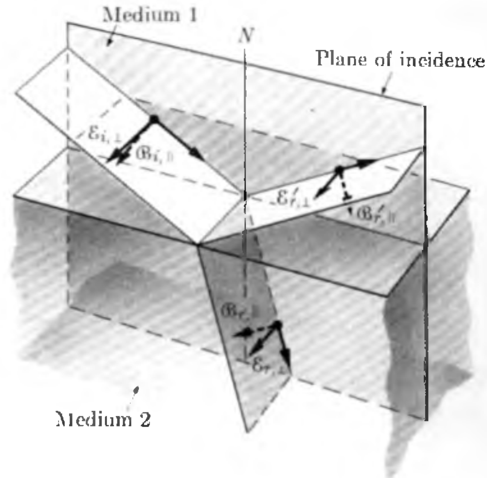


Fig. 14-2. Electric and magnetic fields in the incident, reflected, and refracted waves for polarization perpendicular to the plane of incidence.

$\mathscr{E}$ and $\mathscr{B}$, a component $\mathscr{B}_\perp$ is associated with $\mathscr{E}_\parallel$, and a component $\mathscr{B}_\parallel$ is associated with $\mathscr{E}_\perp$. As defined in Chapter 11, the polarization of an electromagnetic wave is conventionally determined by the direction of the electric field: Fig. 14-1 indicates the $\mathscr{E}$ and $\mathscr{B}$ components for polarization in the plane of incidence; and Fig. 14-2, those components for polarization perpendicular to the plane of incidence. The arrows in each case indicate the directions considered positive for the $\mathscr{E}$-components. The general case is a combination of both polarizations since as indicated before, the fields $\mathscr{E}$ and $\mathscr{B}$ can always be separated into parallel (||) and perpendicular ($\perp$) components.

Maxwell's equations provide certain relations among the parallel and perpendicular components of the electric and the magnetic fields on both sides of the surface separating two media; these relations allow us to establish the relations between the components of the electric field in the incident, refracted, and reflected waves. From these relations we can calculate the coefficients of reflection and refraction or transmission as we did in Example 13.2. When $\mu_1 = \mu_2 \approx \mu_0$, which applies in a wide range of cases, the results obtained are the following:

$$
\begin{aligned}
r_\parallel &= \left(\frac{\mathscr{E}'_r}{\mathscr{E}_i}\right)_\parallel = \frac{n_2 \cos\theta_i - n_1 \cos\theta_r}{n_1 \cos\theta_r + n_2 \cos\theta_i}, \\
r_\perp &= \left(\frac{\mathscr{E}'_r}{\mathscr{E}_i}\right)_\perp = \frac{n_1 \cos\theta_i - n_2 \cos\theta_r}{n_1 \cos\theta_i + n_2 \cos\theta_r}, \\
t_\parallel &= \left(\frac{\mathscr{E}_r}{\mathscr{E}_i}\right)_\parallel = \frac{2n_1 \cos\theta_i}{n_1 \cos\theta_r + n_2 \cos\theta_i}, \\
t_\perp &= \left(\frac{\mathscr{E}_r}{\mathscr{E}_i}\right)_\perp = \frac{2n_1 \cos\theta_i}{n_1 \cos\theta_i + n_2 \cos\theta_r}.
\end{aligned}
\tag{14.1}
$$

If Snell's law is applied to each of Eqs. (14.1), they may be rewritten as

$$
\begin{aligned}
r_{\parallel} &= \frac{\tan(\theta_i-\theta_r)}{\tan(\theta_i+\theta_r)} \\
r_{\perp} &= -\frac{\sin(\theta_i-\theta_r)}{\sin(\theta_i+\theta_r)} \\
t_{\parallel} &= \frac{2\cos\theta_i \sin\theta_r}{\sin(\theta_i+\theta_r)\cos(\theta_i-\theta_r)} \\
t_{\perp} &= \frac{2\cos\theta_i \sin\theta_r}{\sin(\theta_i+\theta_r)} .
\end{aligned}
\tag{14.2}
$$

For normal incidence, $\theta_i=0$, giving $\theta_r=0$; therefore the general equations (14.1) reduce to

$$r_{\parallel} = \frac{n_{21}-1}{n_{21}+1}, \qquad r_{\perp} = -r_{\parallel}, \tag{14.3}$$

and

$$t_{\parallel} = \frac{2}{n_{21}+1}, \quad t_{\perp} = t_{\parallel} \tag{14.4}$$

where $n_{21}=n_2/n_1$.

From Eq. (14.3) we see that on normal incidence one of the reflected components always suffers a phase change of π; which component is out of phase with the incoming wave depends on whether n_{21} is greater or less than 1. The refracted wave is always transmitted without a phase change.

There are a number of other interesting results. For example it may be seen from the first of Eqs. (14.2) that when $\theta_i+\theta_r=\pi/2$ so that the reflected and the refracted rays are perpendicular, the denominator becomes infinitely large and $r_{\parallel}=0$. That is, the reflected wave is totally polarized in a plane perpendicular to the plane of incidence. This case is shown in Fig. 14-3, in which for clarity, only the electric components of the field are drawn. Thus *when the reflected and the refracted rays are perpendicular, the reflected ray is totally polarized with the electric field perpendicular to the plane of incidence.* The corresponding incident angle θ_i is said to be the *polarizing angle.* When $\theta_i+\theta_r=\pi/2$, $\sin\theta_r=\sin(\pi/2-\theta_i)=\cos\theta_i$ and Snell's law gives

$$\tan\theta_i=n_{21} \tag{14.5}$$

for the *polarizing angle.* Thus total linear polarization of the reflected wave occurs when the angle of incidence is set such that its tangent equals the relative index of refraction. This result is called *Brewster's law*, and θ_i is often called the Brewster angle. This phenomenon was first studied by the Scottish physicist Sir David Brewster (1781–1868).

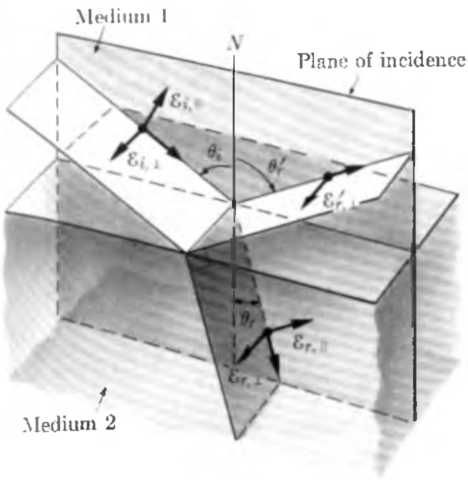


Fig. 14-3. Polarization of an electromagnetic wave by reflection (Brewster's Law).

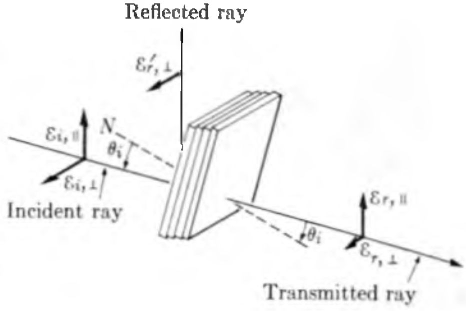


Fig. 14-4. Polarization of an electromagnetic wave by successive refractions from a stack of plates.

Note from Eqs. (14.1) or (14.2) that the refraction or transmission coefficients $t_{\parallel}$ and $t_{\perp}$ can never be zero, and therefore the refracted wave is never completely polarized. However if an electromagnetic wave is transmitted through a stack of thin parallel plates (Fig. 14-4), and the angle of incidence equals the polarizing angle, the final transmitted wave has a much smaller component $\mathscr{E}_{r,\perp}$ since this component tends to go with each reflected wave every time the wave is reflected as it passes from one plate to the next. Therefore if enough plates are in the stack, the transmitted wave is almost totally polarized, and the transmitted electric field oscillates in the plane of incidence.

Example 14.1. Reflection of an electromagnetic wave incident normal to a conductor.

▼ When an electromagnetic wave falls normal to a boundary between vacuum and a conductor, the electric vector is tangent (or parallel) to the surface of the boundary (Fig. 14-5a). Generally then the amplitudes of the incident, reflected, and refracted waves are related by

$$\mathscr{E}_{0,i}+\mathscr{E}_{0,r'}=\mathscr{E}_{0,r}; \tag{14.6}$$

but since the electric field inside a conductor is zero, one may write $\varepsilon_{0,r'}=-\varepsilon_{0,i}$; therefore

$$r=\frac{\mathscr{E}_{0,r'}}{\mathscr{E}_{0,i}}=-1. \tag{14.7}$$

Thus not only is the entire wave reflected, but also there is a phase shift of π radians.

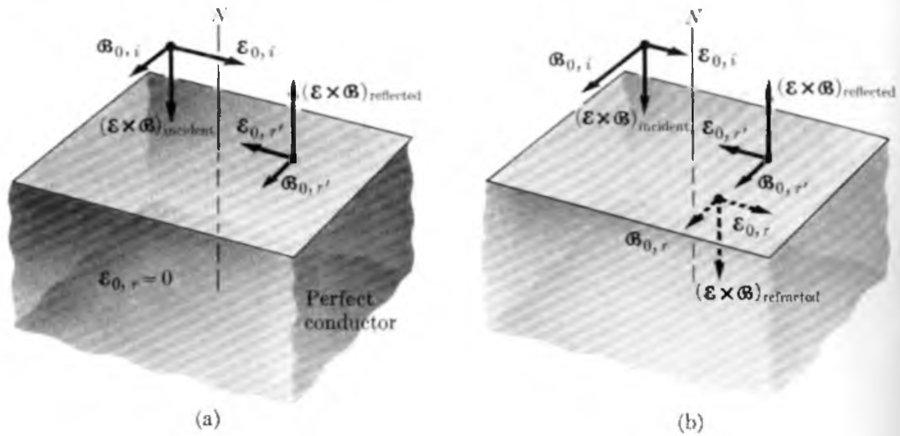


Fig. 14-5. Reflection and refraction at normal incidence. (a) Reflection at a perfect conductor. Since the electric vector suffers a 180° phase change, the magnetic vector does not have a phase change. (b) Reflection at a dielectric. Whether the reflected magnetic vector has a phase change depends on the reflected electric vector.

Furthermore since the vector $\mathscr{E}\times\mathscr{B}$ gives the direction of propagation, the value above for the reflected electric field requires that the reflected magnetic field be in phase with the incident magnetic field. ▲

Example 14.2. Reflection and refraction of an electromagnetic wave at the boundary of vacuum and a dielectric.

▼ Consider the case in which the wave is incident normal to the surface (Fig. 14-5b), and assume that the dielectric constant of the material is a real number ϵ. As in Example 14.1 the electric vector at the boundary is written

$$\mathscr{E}_{0,i}+\mathscr{E}_{0,r'}=\mathscr{E}_{0,r}.$$

Now however, the electric field within the dielectric is not zero and a wave is refracted or transmitted into the dielectric. From Eqs. (11.13) and (12.15) we recall that

$$\mathscr{B}_{0,i}=\frac{1}{c}\mathscr{E}_{0,i}, \qquad \mathscr{B}_{0,r'}=-\frac{1}{c}\mathscr{E}_{0,r'}$$

and

$$\mathscr{B}_{0,r}=\frac{1}{v}\mathscr{E}_{0,r}=\frac{n}{c}\mathscr{E}_{0,r}.$$

Since the magnetic field vector is continuous across the boundary,

$$\mathscr{B}_{0,i}+\mathscr{B}_{0,r'}=\mathscr{B}_{0,r}$$

or from the equations above,

$$\frac{1}{c}\mathscr{E}_{0,i}-\frac{1}{c}\mathscr{E}_{0,r'}=\frac{n}{c}\mathscr{E}_{0,r}. \tag{14.8}$$

Table 14-1. Indexes of Refraction of Several Substances for Electromagnetic Waves*

Substance	n	Substance	n
Air	1.00029	Flint glass	1.65
Alcohol (293 K)	1.36	Ice	1.31
Carbon bisulfide	1.63	Quartz	1.51
Crown glass	1.52	Sodium (liquid)	4.22
Diamond	2.417	Water (298 K)	1.33

*Average values in the visible region of the spectrum.

Note that the minus sign in the second term is necessary in order to have the correct direction for $\mathscr{E} \times \mathscr{B}$ in the reflected wave. From the two equations (14.6) and (14.8), we find that

$$\mathscr{E}_{0,r'} = -\left(\frac{n-1}{n+1}\right)\mathscr{E}_{0,i}, \qquad \mathscr{E}_{0,r} = \left(\frac{2}{n+1}\right)\mathscr{E}_{0,i}$$

and (14.9)

$$\mathscr{B}_{0,r'} = \left(\frac{n-1}{n+1}\right)\mathscr{B}_{0,i}, \qquad \mathscr{B}_{0,r} = \left(\frac{2}{n+1}\right)\mathscr{B}_{0,i}.$$

The first of these two equations should be compared with Eqs. (14.1). We also see then that $r = (n-1)/(n+1)$ and $t = 2/(n+1)$, which agree with Eq. (14.3). ▲

Example 14.3. The coefficients of reflection and transmission for electromagnetic waves in the visible region for crown glass, at an angle of incidence equal to 30°.

▼ Table 14-1 lists the index of refraction in the visible region for a number of materials. From the table, $n_{\text{crown}} = 1.52$; and with $n_{\text{air}} = 1$, Snell's law gives $\sin\theta_i = 1.52 \sin\theta_r$. Setting $\theta_i = 30°$ gives $\theta_r = 19° 12'$. Therefore applying relations (14.1) or (14.2) yields

$$r_{\parallel} = 0.164, \quad r_{\perp} = -0.247, \quad t_{\parallel} = 0.766, \quad t_{\perp} = 0.753.$$

Note that the perpendicular component of the reflected wave has suffered a phase change of π radians. The Brewster angle for crown glass corresponds to $\tan\theta_i = 1.52$ or $\theta_i = 56° 40'$. ▲

14.3 Propagation of Electromagnetic Waves in an Anisotropic Medium

When a transverse wave propagates through an anisotropic medium, the velocity of propagation of the wave may depend both on the direction of polarization and on the direction of propagation of the wave. This double dependence is particularly true in the case of electromagnetic waves (the only ones considered in this section). The polarizability of most molecules is not the same in all directions. Since the molecules in gases and liquids are oriented at random, this directional dependence of the polarizability does not give rise to any particular effect; and the medium effectively behaves macroscopically as an isotropic substance. However in a crystalline solid the molecules are more or less oriented and their orientation is "frozen"; that is, they are not free to rotate around their equilibrium positions within the crystal lattice. Thus in general the properties of the crystal depend on the direction along which they are measured. Depending on their molecular structure and arrangement, crystalline solids may behave optically as either isotropic or anisotropic media.

That the polarizability of the medium is not the same in all directions means that the polarization $\mathscr{P}$ does not have the same direction as the electric field $\mathscr{E}$, and their relative directions are different for different orientations in the crystal. As a result, the displacement vector $\mathscr{D} = \epsilon_0\mathscr{E} + \mathscr{P}$ is also generally not parallel to $\mathscr{E}$. We find, however, that $\mathscr{E}$ and $\mathscr{D}$ are parallel along at least three perpendicular directions, called principal axes*, characteristic of each substance. With the coordinate axes XYZ oriented parallel to these principal axes, and designating the three principal values of the permittivity of the substance that correspond to each of the principal axes by ϵ_1, ϵ_2, ϵ_3, the components of $\mathscr{D}$ for an arbitrary orientation of $\mathscr{E}$, by extension of Eq. (2.14), are

$$\mathscr{D}_x = \epsilon_1\mathscr{E}_x, \qquad \mathscr{D}_y = \epsilon_2\mathscr{E}_y, \qquad \mathscr{D}_z = \epsilon_3\mathscr{E}_z. \tag{14.10}$$

We may also speak of three principal indexes of refraction n_1, n_2, and n_3, each one associated with the corresponding permittivity as indicated by Eq. (12.16).

Both experiment and theory (based on Maxwell's equations and the preceding discussion) show that

> *in an anisotropic medium, for each direction of propagation of a plane electromagnetic wave there correspond two possible, mutually perpendicular states of polarization, each of which propagates with a different velocity.*

Thus no matter what the initial state of polarization, when an electromagnetic wave penetrates an anisotropic substance, the wave splits into two waves, polarized at right angles to each other and propagating with different phase velocities. This situation gives rise to the phenomenon of double refraction, which will be discussed in Section 14.5.

If the wave's direction of propagation is given, we can determine the phase velocity and state of polarization using a geometrical method suggested by the French physicist Augustin Fresnel (1788–1827). Construct an ellipsoid with axes n_1, n_2, and n_3, oriented according to the three principal axes of the substance; this figure is

---

*The situation here is mathematically very similar to that encountered with rigid body rotation. Recall that L and ω are not parallel except in the case of rotation along a principal axis of the body.

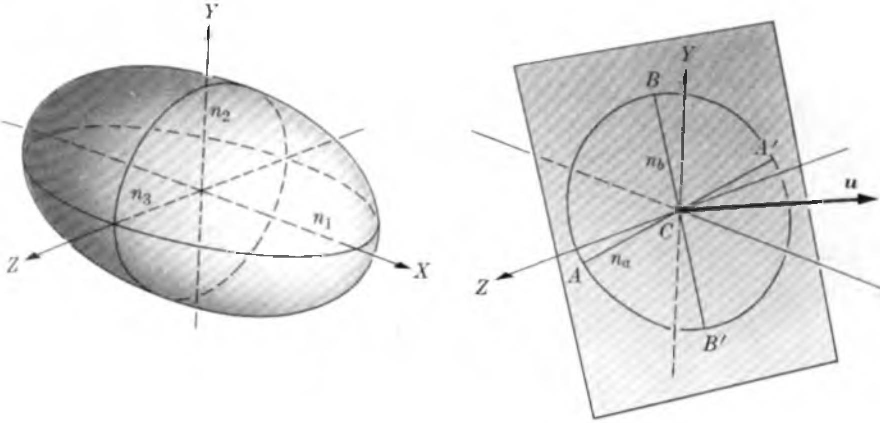


Fig. 14-6. (a) The Fresnel ellipsoid. (b) Ellipse $ABA'B'$ is the intersection with the ellipsoid of a plane perpendicular to $\boldsymbol{u}$, passing through C.

called the *Fresnel ellipsoid* (Fig. 14-6a). Then given the direction of propagation of the wave, determined by the unit vector $\boldsymbol{u}$, draw a plane through the center C of the ellipsoid and perpendicular to $\boldsymbol{u}$. The intersection of the plane and the ellipsoid is an ellipse (Fig. 14-6b). The directions of the two axes AA' and BB' of this ellipse determine the planes of polarization of the wave for that direction of propagation. The lengths CA and CB of the two axes of the ellipse give the indexes of refraction n_a and n_b for each polarization, and therefore the corresponding phase velocity.

Isotropic media are characterized by the fact that all three principal indexes of refraction are equal ($n_1 = n_2 = n_3$). Fresnel's ellipsoid is a sphere and the index of refraction is the same in all directions. Hence no special polarization direction exists since all intersections are circles. *Cubic* crystals, as well as most noncrystalline media, behave this way.

Another special case is that in which two principal indexes of refraction are the same, say $n_2 = n_3$. The direction corresponding to the unequal index n_1 is called the *optical axis*; it is an axis of symmetry of the crystal. For that reason these substances are called *uniaxial crystals*. To this class belong the *trigonal*, *hexagonal*, and *tetragonal* crystal systems. When $n_2 < n_1$, the crystal is called *positive*; when $n_2 > n_1$, the crystal is *negative*. The Fresnel ellipsoid of a uniaxial crystal is an ellipsoid of revolution around the optical axis (Fig. 14-7). From the geometrical properties of an ellipsoid of revolution, the intersection of a plane through the center C and perpendicular to the direction of propagation $\boldsymbol{u}$ (of an electromagnetic wave) is an ellipse, one of whose axes (CO) is always equal to n_2 and is directed perpendicular both to the direction of propagation and to the optical axis; the other axis (CE) of the ellipse has a variable length n_e between n_2 and n_1, and is in the plane determined by the direction of propagation and the optical axis. In this case two waves may be defined: the ordinary and the extraordinary.

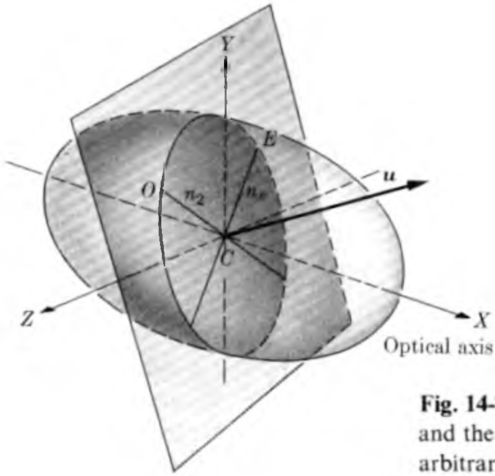


Fig. 14-7. Directions of polarization of the ordinary and the extraordinary rays in a uniaxial crystal for an arbitrary direction of propagation.

The *ordinary wave* is linearly polarized in the plane determined by CO and $\boldsymbol{u}$, and is thus perpendicular to the plane determined by the direction of propagation and the optical axis. The ordinary wave propagates in all directions with the same velocity $v_o = v_2 = c/n_2$. The ordinary wave therefore behaves as a wave in an isotropic medium, and for this reason is called ordinary.

The *extraordinary wave* is linearly polarized in the plane determined by CE and $\boldsymbol{u}$ or (what is the same thing) by the direction of propagation and the optical axis; but the extraordinary wave's velocity v_e depends on the direction of propagation and varies from v_2 to v_1 (corresponding to an index of refraction between n_2 and n_1).

When waves are propagating along the optical axis, the ellipse of intersection is a circle of radius n_2; and the two waves propagate with the same velocity v_2. This may be considered as another definition of the optical axis (Fig. 14-8a): the optical axis is that direction along which there is only one velocity of propagation. When waves are propagating perpendicular to the optical axis, the ellipse of intersection has semi-axes n_1 and n_2, and the extraordinary wave has the velocity v_1 (Fig. 14-8b).

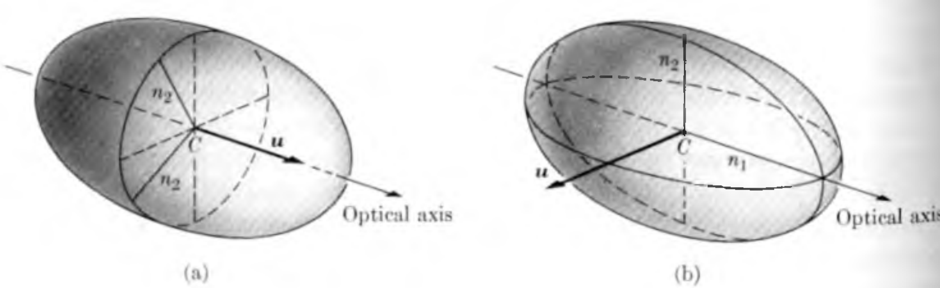


Fig. 14-8. Directions of polarization of the ordinary and the extraordinary rays in a uniaxial crystal for propagation (a) parallel or (b) perpendicular to the optical axis.

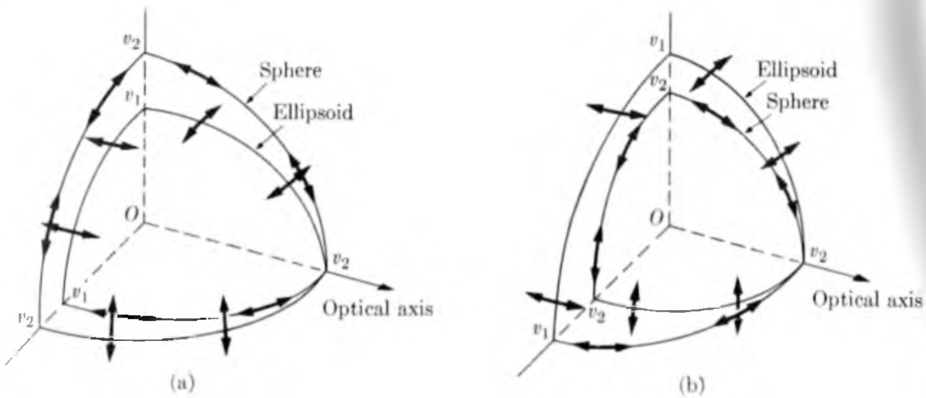


Fig. 14-9. Fresnel velocity surface for uniaxial crystals. (a) A positive crystal in which $n_2 < n_1$ (or $v_2 > v_1$). (b) A negative crystal in which $n_2 > n_1$ (or $v_2 < v_1$).

Another useful geometrical construction is obtained by plotting for each direction of propagation, vectors having lengths equal to v_o and v_e, the phase velocities of the ordinary and the extraordinary waves, respectively; the result is a double surface (Fig. 14-9) called a *Fresnel velocity surface.* One surface is a sphere of radius $v_o = v_2$; this surface corresponds to the velocity of the ordinary wave. The other surface is an ellipsoid of revolution with axes v_1 and v_2; that surface corresponds to the extraordinary wave. The two surfaces are tangent at their intersection with the optical axis. The state of polarization for several directions of propagation is indicated in Fig. 14-9. The ordinary ray is polarized along a meridian; the extraordinary ray is polarized in a longitudinal direction.

In the general case of three different indexes of refraction, it can be proved that there are two directions for which the velocities of propagation of the two polarized waves are equal. These directions, also called optical axes, are perpendicular to the planes whose intersections with the Fresnel ellipsoid are circles. Substances in which these axes exist are called *biaxial* and belong to the *orthorhombic, monoclinic,* and *triclinic* crystal systems. The Fresnel velocity surface for biaxial crystals is more complicated, and we shall not enter into a discussion of the geometrical details. Table 14-2 lists the indexes of refraction for selected uniaxial and biaxial materials.

Many normally isotropic substances become anisotropic when subject to mechanical stresses or to strong static electric or magnetic fields perpendicular to the direction of propagation, giving rise to the *Kerr electro-optic effect* and the *Cotton-Mouton magneto-optic effect*. In all cases the anisotropy of the substance is due to the partial orientation of the molecules that results from the stresses or the applied fields.

Example 14.4. The phase difference between the ordinary and the extraordinary waves, and the state of polarization of the emergent wave when a linearly polarized wave falls on a thin plate of a uniaxial material.

Table 14-2. Principal Indexes of Refraction of Several Crystals*

Substance	n_1	n_2	n_3
Uniaxial:			
Apatite	1.6417	1.6461	
Calcite	1.4864	1.6583	
Quartz	1.5533	1.5442	
Zircon	1.9682	1.9239	
Biaxial:			
Aragonite	1.5301	1.6816	1.6859
Gypsum	1.5206	1.5227	1.5297
Mica	1.5692	1.6049	1.6117
Topaz	1.6155	1.6181	1.6250

*For sodium light, $\lambda = 5.893 \times 10^{-7}$ m.

▼ Figure 14-10 shows the experimental arrangement in which the uniaxial crystal has been cut with its faces parallel to the optical axis and placed with its optical axis (index n_1) perpendicular to the direction of propagation of the incident electromagnetic wave. The direction of the optical axis has been designated by Y. The Z-direction corresponds to the direction of polarization of the ordinary ray (index n_2). Suppose that a linearly polarized wave making an angle α with the Y-axis falls on the plate. For convenience write $\mathscr{E} = \mathscr{E}_0 \sin(\omega t - kx)$ for the electric field in the incident wave. The components of the electric field in the incident wave along the Y- and Z-axes are

$$\mathscr{E}_y = \mathscr{E}_{0y} \sin(\omega t - kx), \qquad \mathscr{E}_z = \mathscr{E}_{0z} \sin(\omega t - kx)$$

where

$$\mathscr{E}_{0y} = \mathscr{E}_0 \cos\alpha \qquad \text{and} \qquad \mathscr{E}_{0z} = \mathscr{E}_0 \sin\alpha.$$

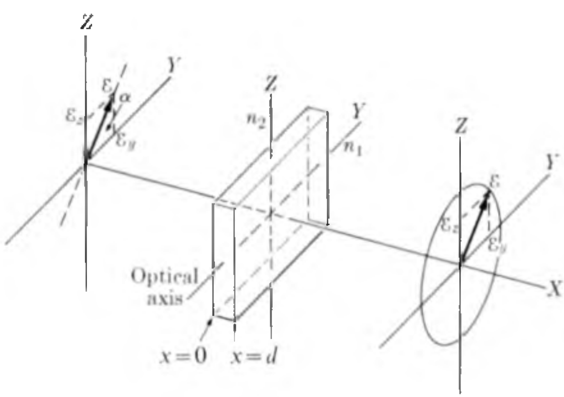


Fig. 14-10. Change of polarization of an electromagnetic wave after traversing a parallel plate cut from a uniaxial crystal.

When it propagates through the crystal, the linearly polarized wave is separated into two waves with their electric fields along the Y- and Z-axes, respectively. These component waves correspond to the extraordinary and the ordinary waves. Since the velocities of propagation of the waves are $v_1=c/n_1$ and $v_2=c/n_2$, the corresponding propagation vectors are

$$k_1=\frac{\omega}{v_1}=\frac{\omega n_1}{c}=kn_1, \qquad k_2=kn_2$$

where $k=\omega/c$. Therefore after the waves have traversed the thickness d, the respective electric fields are represented by the expressions

$$\mathscr{E}_y=\mathscr{E}_{0y}\sin(\omega t-k_1 d), \qquad \mathscr{E}_z=\mathscr{E}_{0z}\sin(\omega t-k_2 d);$$

the result is a phase difference between the two waves of

$$\delta=(k_1-k_2)d=k(n_1-n_2)d=2\pi(n_1-n_2)d/\lambda.$$

After traversing the plate, the two waves recombine into a single wave. Because of the phase difference, the transmitted wave will generally be elliptically polarized. The axes of the ellipse will be parallel to the Y- and Z-axes if δ is an odd multiple of $\pi/2$; that is, if

$$(n_1-n_2)d=\text{odd integer}\times\frac{\lambda}{4}. \tag{14.11}$$

If δ is a multiple of π, that is, if

$$(n_1-n_2)d=\text{integer}\times\frac{\lambda}{2}, \tag{14.12}$$

the transmitted wave will be linearly polarized. In this case if the integer is even, the transmitted wave is linearly polarized in the same plane as the incident wave; but if the integer is odd, the wave is polarized in a plane symmetric with the XZ-plane. If the initial angle α is 45°, these two planes will be perpendicular.

The plates corresponding to the two conditions given above are called a *quarter-wave plate* and a *half-wave plate*. These types of plates are widely used in the analysis of polarized light.

The situation also works in the opposite direction: elliptically polarized light that passes through a quarter-wave plate becomes plane polarized. ▲

14.4 Dichroism

Some anisotropic substances absorb the ordinary and the extraordinary waves in very different proportions. Under such conditions an electromagnetic wave propagating through a sufficiently thick piece of the substance becomes gradually polarized in one plane since either the ordinary or the extraordinary wave is almost completely absorbed. The phenomenon is called *dichroism* and is illustrated in Fig. 14-11, in which $\mathscr{E}_0$ is the amplitude of the electric field in the incident wave and the wave is

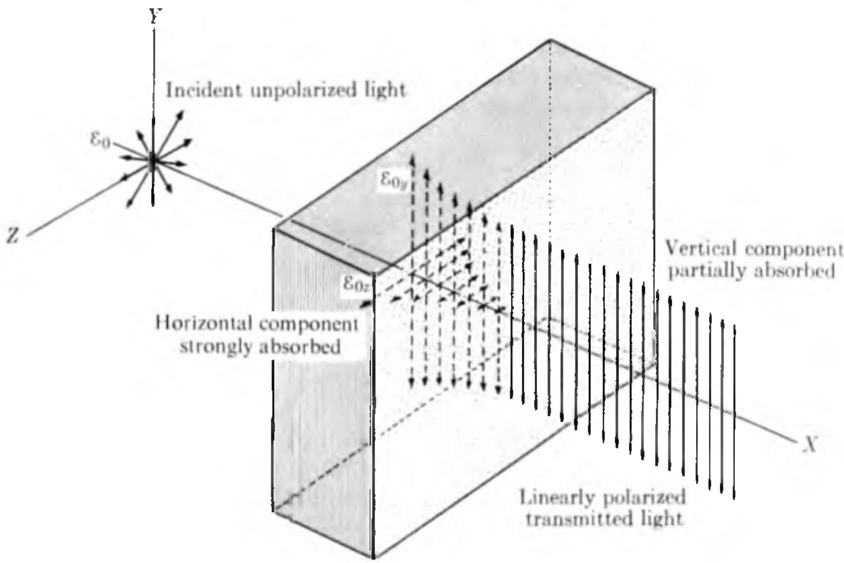


Fig. 14-11. Dichroism.

traveling in the X-direction. As it penetrates the substance, the incident wave is separated into ordinary and extraordinary waves, polarized parallel to the Y- and Z-axes. The waves' amplitudes are $\mathcal{E}_{0y}$ and $\mathcal{E}_{0z}$. If $\mathcal{E}_{0z}$ is absorbed more than $\mathcal{E}_{0y}$, then after the waves have traversed a certain length, we almost have only $\mathcal{E}_{0y}$, resulting in linearly polarized light.

Since dichroism is the result of a difference in absorption coefficients and is dependent on the frequency of the electromagnetic wave, a substance will exhibit the phenomenon at certain frequencies to a greater degree than at other frequencies. In some substances then, the colors of the two waves appear different; hence the name dichroic or "two-colored." In the visible region there are two especially important dichroic substances. One is *tourmaline* (aluminum borosilicate), which preferentially absorbs the ordinary ray. The other is *herapathite* (sulfate of iodoquinine), which has the inconvenience that its crystals are very brittle and therefore difficult to preserve in appropriate sizes. However, this substance is manufactured in a form called *Polaroid*, which consists of many small crystals oriented parallel to one another. This alignment is accomplished by having the herapathite crystals adhere themselves to substances composed of very long molecules, such as stretched, polyvinyl alcohol sheets. The resultant combination yields a material having optical properties that are very different in the longitudinal and in the transverse directions. Dichroism provides one of the simplest and cheapest ways of producing and analyzing polarized light.

14.5 Double Refraction

We shall now discuss how an electromagnetic wave behaves when it passes into an anisotropic medium. Our discussion will be limited to uniaxial crystals; the reflected wave will not be considered since it does not involve any aspect essentially different from those discussed earlier in the chapter. Consider the case of normal incidence of an unpolarized wave on a plane surface. The optical axis is in the plane of the page (Fig. 14-12). From the definition of a wavefront, both the ordinary and the extraordinary refracted wave surfaces remain parallel to the interface as they propagate through the anisotropic medium. To determine the directions of the ordinary and the extraordinary rays, draw the Fresnel velocity surfaces (as discussed in connection

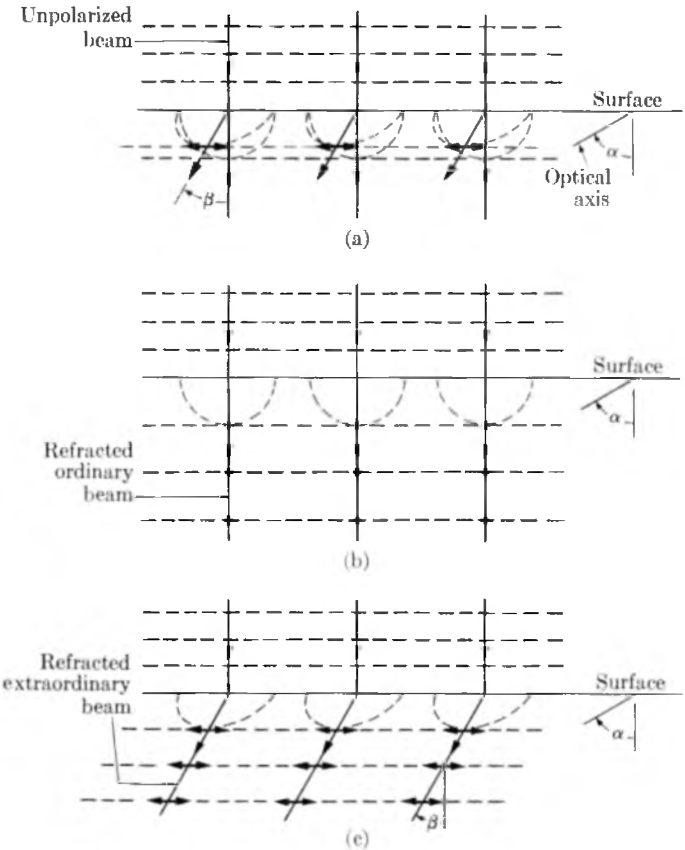


Fig. 14-12. The separation of the ordinary and extraordinary rays as light passes through a doubly refractive material. All rays are in the plane of the page; the optical axis is also in the plane of the page.

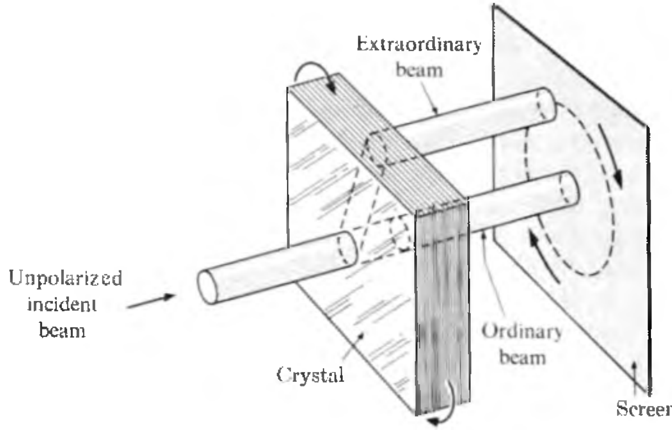


Fig. 14-13. A narrow beam of unpolarized light can be split into two beams by a doubly refracting crystal. If the crystal is rotated, the extraordinary beam rotates around the ordinary ray. The two beams are linearly polarized at right angles with respect to each other.

with Fig. 14-9) at the points of incidence for the incoming rays as shown in Fig. 14-12a. In Fig. 14-12a the surfaces appear as circles (for the ordinary rays) and ellipses (for the extraordinary rays) with their semiaxes parallel to the optical axis. The common tangents of the two sheets of the Fresnel velocity surface give the ordinary and the

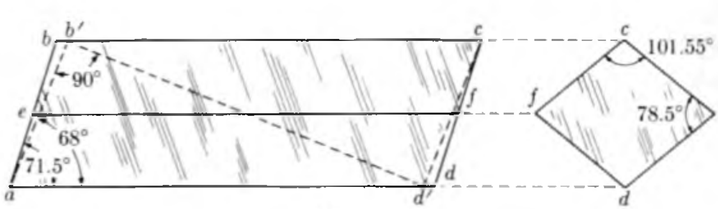
Fig. 14-14. Photograph of the double image produced by a calcite crystal. (Photograph courtesy of W. L. Hyde, Director, Institute of Optics, University of Rochester.)

extraordinary wave fronts. The points of tangency determine the directions of the ordinary and the extraordinary rays. Therefore the ordinary wave will propagate in the direction of incidence and will be linearly polarized perpendicular to the plane of the paper (as indicated by the dots in Fig. 14-12b). However, although remaining parallel to the interface, the extraordinary wave will suffer a sidewise displacement so that the energy flow is along the extraordinary ray at an angle β with respect to the direction of propagation. The extraordinary wave will be polarized in the plane of the paper (as indicated by the bars in Fig. 14-12c).

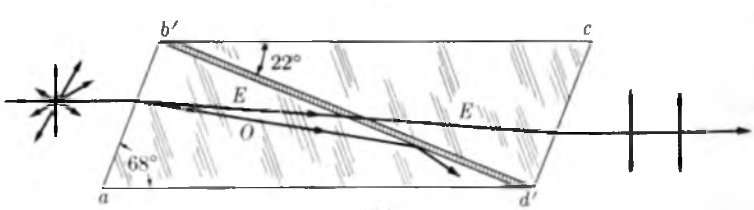
When two refracted rays correspond to a single incident ray, the result is called *double refraction*; and for that reason anisotropic substances are called *birefringent*. When the substance is bounded by two parallel surfaces (Fig. 14-13), the ordinary and the extraordinary rays emerge parallel but separated, and result in a double image as shown in the photograph of a calcite crystal in Fig. 14-14.

When the incidence is oblique, the situation is somewhat more complicated geometrically, but the physical result is the same. That is, for a given incident wave there are two different refracted waves propagating in different directions and polarized at right angles to each other.

Double refraction is a useful research tool in the study of crystal structure, and has many other interesting applications. One practical application consists in producing a beam of plane polarized light by means of a *Nicol prism*. To make a Nicol prism, a calcite crystal whose length is four times its width is cut at the end faces as shown by the dashed lines ab' and cd' in Fig. 14-15a. The crystal is then cut diagonally along line $b'd'$ and the two halves are glued together with Canada balsam. The balsam's index of refraction ($n \simeq 1.55$) has a value between that of the calcite crystal for the ordinary and the extraordinary rays (see Table 14-2). Because of this value and also because of the geometry of the crystal, the ordinary rays are totally reflected at the surface of separation and are deviated out of the prism while the extraordinary rays



(a)



(b)

Fig. 14-15. (a) Natural calcite crystal, called Iceland spar. (b) A Nicol prism.

proceed into the other half of the crystal and emerge at the far end (Fig. 14-15b). Therefore the transmitted light is linearly polarized. Nicol prisms are used in many optical instruments, such as polarimeters.

Example 14.5. A ray of light falls on a calcite crystal cut with its surface parallel to the optical axis. When the plane of incidence is perpendicular to the optical axis, and the angle of incidence is 50°, find the angular separation between the ordinary and the extraordinary rays.

▼ According to Fig. 14-8b, when the propagation of the wave is in a direction perpendicular to the optical axis, the ordinary rays propagate with velocity v_2 corresponding to the refractive index n_2, and the extraordinary waves propagate with velocity v_1 corresponding to the refractive index n_1. Therefore using Snell's law and the principal indexes of refraction from Table 14-2 gives $\sin\theta_i/\sin\theta_o = n_2 = 1.6583$ and $\sin\theta_i/\sin\theta_e = n_1 = 1.4864$. Given that $\theta_i = 50°$, then $\theta_o = 27°\ 31'$ and $\theta_e = 31°\ 1'$. The angular separation of the two rays is thus $\theta_e - \theta_o = 3°\ 30$. ▲

Example 14.6. Fluctuation in the intensity of the transmitted wave when a linearly polarized wave passes through an analyzer and the analyzer is rotated.

▼ An *analyzer* is a device that transmits only that wave component whose electric field is parallel to the axis AA' (Fig. 14-16). When the axis AA' of the analyzer makes an angle θ with the electric field of an incident linearly polarized wave, only the component $\mathscr{E}_A = \mathscr{E}\cos\theta$ is transmitted. Therefore since the intensity of the wave is proportional to the square of the electric field, we have the relation

$$I = I_0 \cos^2\theta \tag{14.13}$$

where I_0 is the intensity of the incident wave and I is that of the transmitted wave. This result is known as *Malus's law*. When $\theta = 0$ or π, the intensity of the transmitted light is maximum; when $\theta = \pi/2$ or $3\pi/2$, it is zero. Therefore when the analyzer is rotated, the intensity of the transmitted wave fluctuates between 0 and I_0. This fluctuation affords a means of determining whether a wave, such as light, is polarized or not. For unpolarized or circularly polarized waves, no fluctuation in intensity is observed. For elliptically polarized waves, the transmitted wave fluctuates between a maximum and a minimum value. These two extremes are obtained when the analyzer

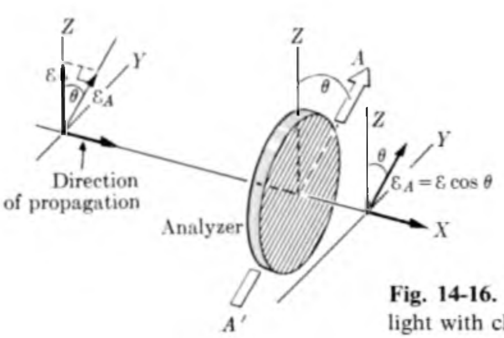


Fig. 14-16. Change in the intensity of linearly polarized light with change in orientation of the analyzer.

is parallel to either the larger or the smaller axis of the polarization ellipse. The degree of polarization of the incident wave is then given by the expression

$$P = \frac{I_{max} - I_{min}}{I_{max} + I_{min}}. \tag{14.14}$$

Note that $P = 1$ for linearly polarized waves and that $P = 0$ for unpolarized waves. ▲

Example 14.7. *Chromatic polarization.*

▼ When linearly polarized white light falls on a plate similar to that considered in Example 14.4, and is analyzed by means of another polarizing device, colored light is observed, the color being dependent on the orientation of the analyzer. Consider the arrangement of Fig. 14-17. Suppose for simplicity that the incident white light is linearly polarized and that the electric field makes an angle of 45° with the optical axis of the plate. According to the results of Example 14.4, the transmitted light will be linearly polarized as either $\mathscr{E}_1$ (parallel to the original polarization) or $\mathscr{E}_2$ (normal to the original polarization vector), depending on whether the wavelength is such that

$$(n_1 - n_2)d = \begin{cases} \text{even integer} \times \lambda/2 & \text{(polarization } \mathscr{E}_1), \\ \text{odd integer} \times \lambda/2 & \text{(polarization } \mathscr{E}_2). \end{cases}$$

For all other wavelengths the transmitted wave is elliptically polarized. If the transmitted light is passed through an analyzer, the light is colored instead of white, and the color changes when the analyzer axis AA' is rotated. This change occurs because according to Example 14.6, when the axis of the analyzer is parallel to $\mathscr{E}_1$, the corresponding wavelengths are transmitted with maximum intensity while those corresponding to $\mathscr{E}_2$ are blocked. With reversed conditions, when the axis of the analyzer is parallel to $\mathscr{E}_2$, the complementary color appears. Therefore as the analyzer is rotated, the shades vary with complementary colors separated by 90°.

This phenomenon has been applied to the stress analysis of structural pieces used in buildings and machines, and has given rise to a branch of applied physics called *photoelasticity*. As mentioned at the end of Section 14.3, when a plastic material is subject to stresses, it becomes birefringent because of the anisotropy resulting from the strains. Therefore if a model made of plastic is subjected to the same stresses as the actual structural piece, the model behaves optically as an

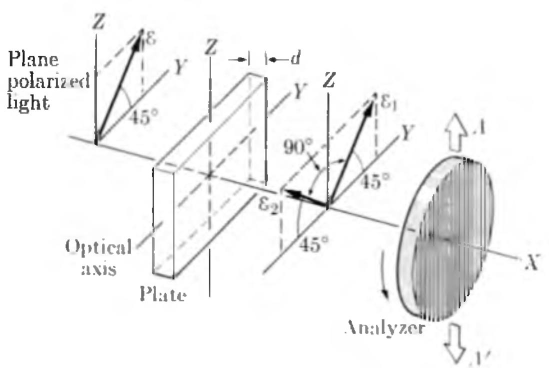


Fig. 14-17. Crystal plate placed between a source of linearly polarized light and an analyzer. The resulting phenomenon is called chromatic polarization.

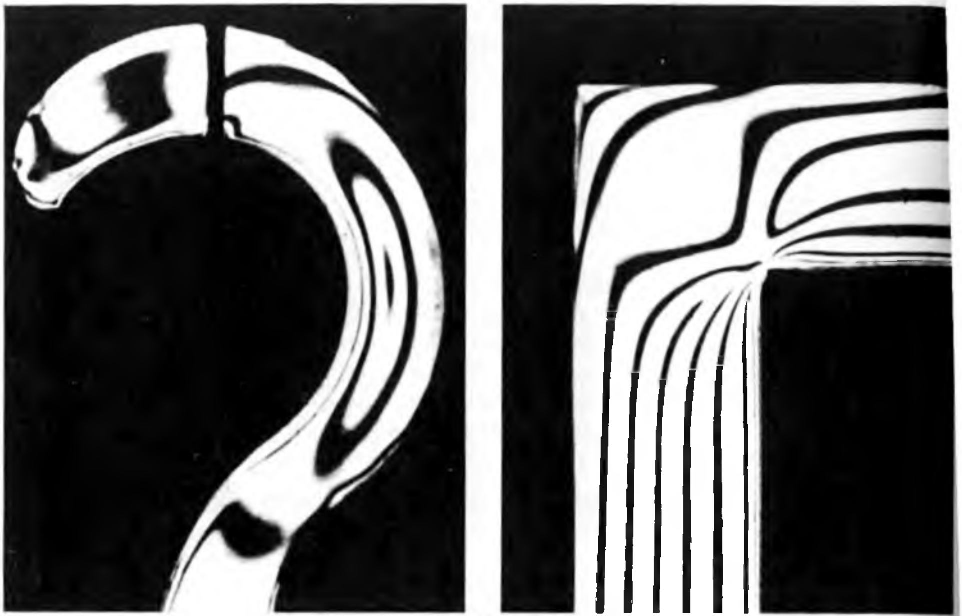


Fig. 14-18. Birefringence induced in a substance by the stresses applied to it. (Photographs courtesy of Klinger Scientific Apparatus Company.)

inhomogeneous birefringent plate. The inhomogeneity is due to the nonuniform distribution of strains in the plastic. When the stressed piece replaces the plate of Fig. 14-17, a (colored) pattern such as those in Fig. 14-18 results. From the pattern the strains can be estimated by the use of special techniques. ▲

14.6 Optical Activity

When a polarized electromagnetic wave passes through certain materials, the plane of polarization is rotated. The rotation of the plane of polarization is called *optical activity*. Thus if a beam of linearly polarized light passes through an optically active substance (Fig. 14-19), the transmitted wave is also linearly polarized but in another plane, making an angle θ with the incident plane. The value of θ is proportional to the length l that the beam traverses through the substance, and depends also on the nature of the substance. From the point of view of an observer receiving the transmitted light, the substances are called *dextrorotatory* (from *dextro*, right) or *levorotatory* (from *levo*, left), depending on whether the rotation of the plane of polarization is clockwise or counterclockwise as seen by the observer.

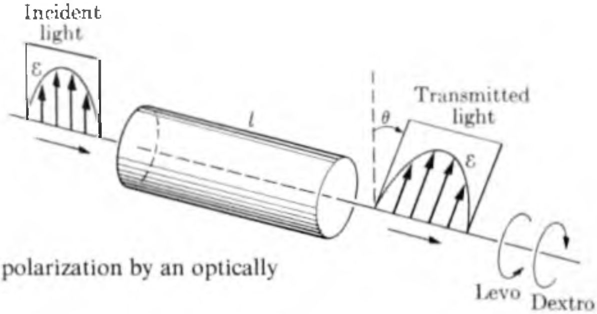


Fig. 14-19. Rotation of the plane of polarization by an optically active substance.

Certain substances exhibit optical activity only in their solid states. Many inorganic crystals, especially quartz, and some organic crystals, such as benzil, are of this type. Upon fusion, solution, or vaporization, these substances lose their optical activity. This loss demonstrates that the optical activity of these substances depends on the special arrangements of atoms or molecules in the crystal; these arrangements disappear when the molecules are oriented at random in the liquid or the gaseous states. Other substances, such as turpentine, sugar, camphor, and tartaric acid, remain optically active in all physical states even when they are in solution. In this latter type of substance, optical activity is associated with the individual molecules and not with their relative arrangement.

Optical activity is a result of a certain twisting of the orbits of the electrons in the molecules or crystals under the action of an external oscillatory electromagnetic field. When the polarization of matter was discussed (Section 2.5), we assumed that the electrons oscillated in a straight line, parallel to the electric field in isotropic substances, and that the electrons oscillated at an angle with the electric field in anisotropic substances (Section 14.3). In certain molecules and crystals, however, the electron motion is along a twisted path, assumed for simplicity to be a helix (Fig. 14-20). Suppose that the molecule (or crystal) is so oriented that the helical electron paths are as shown in Fig. 14-20; i.e., with the helix axis perpendicular to the direction of propagation and parallel to either the electric or the magnetic field of the incoming wave.

Consider first orientation (a) of Fig. 14-20. The oscillating electric field of the wave produces an oscillatory motion of the electrons up and down along the helix; this motion results in an effective oscillating electric dipole moment $\boldsymbol{p}$ parallel to the helix axis. So far the situation is similar to that of ordinary polarization; but because of the twisting of the electronic path, the electronic current along each turn of the helix is equivalent to a magnetic dipole, and the molecule acquires an effective oscillating magnetic dipole moment $\boldsymbol{m}$ also oriented along the helix axis. For orientation (b), the oscillating magnetic field of the wave produces a variable flux through each turn of the helix; by the Faraday-Henry law this flux results in an oscillating electronic current along the helix. This current again produces an oscillating magnetic moment $\boldsymbol{m}$ along the helix axis. However, the electron's back-and-forth motion

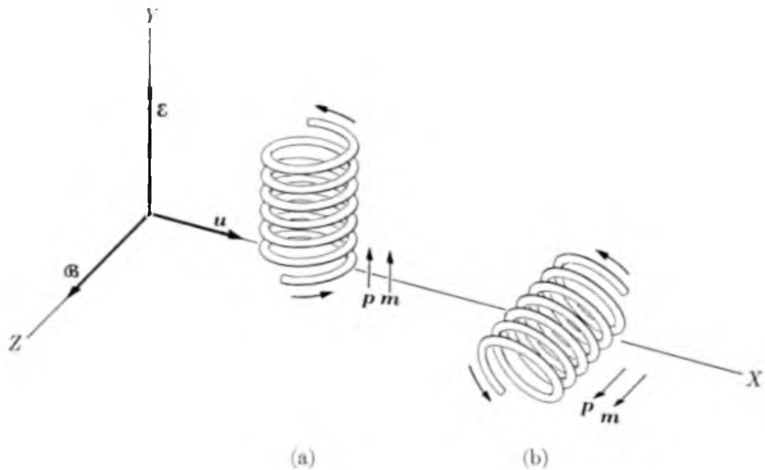


Fig. 14-20. Electric and magnetic dipole moments induced by an electromagnetic wave in a helical molecule.

produces alternate positive and negative charges at the ends of the molecule, and results in an effective oscillatory electric dipole moment $\boldsymbol{p}$ along the helix axis. Therefore for both orientations of the molecule, both an oscillating electric dipole moment $\boldsymbol{p}$ and an oscillating magnetic dipole moment $\boldsymbol{m}$, parallel to the molecular axis, are produced. These dipoles radiate scattered electromagnetic waves in the fashion discussed in Section 12.3, in which only electric dipole scattering was taken into account because the electron's motion was assumed to be in a straight line.

A detailed mathematical analysis of the scattered wave, an analysis here omitted, shows that along the direction of propagation of the incident wave the fields $\mathscr{E}'$ and $\mathscr{B}'$ of the scattered wave are in phase with those of the incident wave, but they oscillate in a different direction because of the differing relative orientation of the $\mathscr{E}$ and $\mathscr{B}$ fields of an electric and a magnetic dipole (Figs. 11-8 and 11-12). An observer along

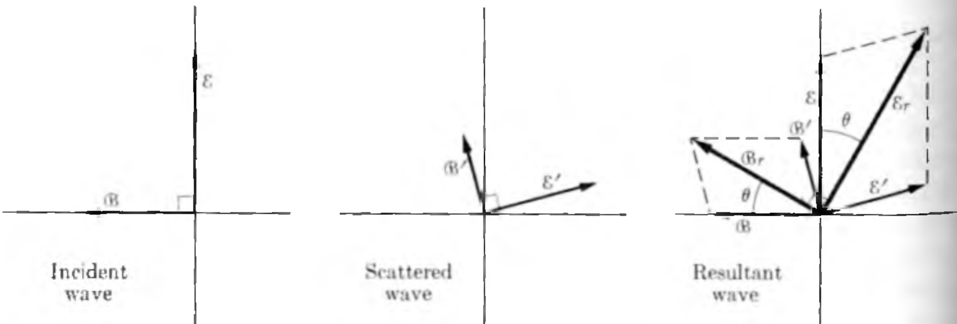


Fig. 14-21. Resultant electric and magnetic fields due to the superposition of incident and scattered waves.

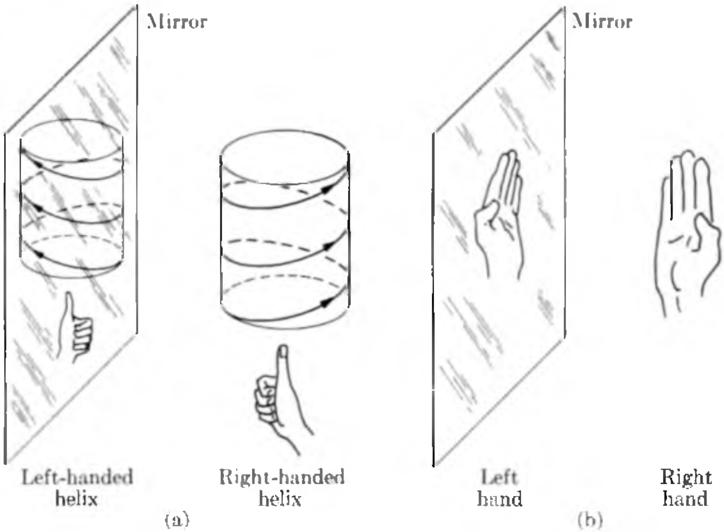


Fig. 14-22. Right-left symmetry. (a) The mirror image of a right-handed helix is left-handed. (b) The mirror image of the right hand is a left hand.

the direction of propagation receives the incident and the scattered waves that interfere and result in linear polarization but in a direction making an angle θ with the original plane of the electric vector (Fig. 14-21). Thus a rotation of the plane of polarization of the wave results. For randomly oriented molecules it can be proved that the effect is always in the same sense although its magnitude depends on the molecular orientation. Thus molecular optical activity persists in any physical state or in solution. In some crystals, however, the effect depends on the molecular arrangement rather than on the individual molecular structure; and therefore the effect disappears when the molecules are disarranged.

The student may realize that there are two kinds of helices, right-handed and left-handed (Fig. 14-22). One helix is the mirror image of the other as the left hand is the mirror image of the right hand. This kind of symmetry is called *enantiomorphism.* Some molecules act like right-handed helixes and others act like left-handed ones. In one case the rotation of the plane of polarization by an optically active substance is in one direction, and in the other case the rotation is in the opposite direction. This fact explains the existence of dextrorotatory and levorotatory substances.

Some substances contain both classes of mirror-image molecules, a property called *stereoisomerism.* For example lactic-acid molecules (CH_3—CHOH—CO_2H) may exist in either one of two mirror-image forms as illustrated in Fig. 14-23. A sample of lactic acid that contains equal amounts of isomers is optically inactive; but if there is more of one kind than of the other, a net rotation for polarized light results.

In the case of quartz (SiO_2), the molecules are all identical; but their space arrangement in the crystal has either a left-handed or a right-handed symmetry as is apparent

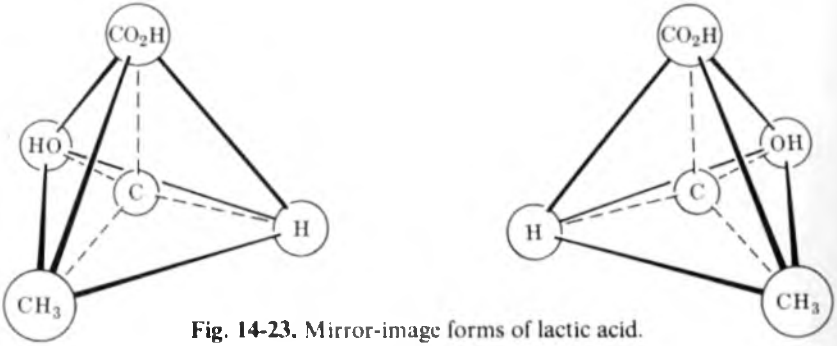


Fig. 14-23. Mirror-image forms of lactic acid.

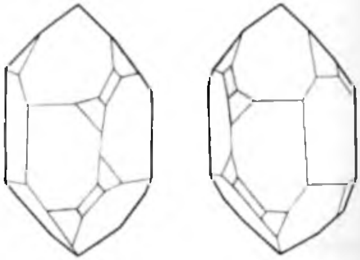


Fig. 14-24. Mirror-image forms of quartz crystal.

from the external appearance of the two kinds of quartz crystals shown in Fig. 14-24. One kind is levo and the other is dextro. When the crystal is melted, the molecular arrangement is destroyed and the optical activity disappears.

When a substance, such as lactic acid, levulose, dextrose, etc., whose molecules are optically active is dissolved in water, the rotation of the plane of polarization depends on the concentration. This result is widely used to determine quantitatively the amount of substance in the solution, such as the concentration of sugar in a syrup or in urine.

Problems

14.1 Linearly polarized light falls on a glass plate ($n = 1.5$) with an angle of incidence of 45°. Find the coefficients of reflection and refraction if the electric field of the incident wave is (a) in the plane of incidence, and (b) normal to the plane of incidence.

14.2 A plane electromagnetic wave falls perpendicularly on a plane surface separating a medium of index n_1 from a medium of index n_2. (a) Using Eq. (14.1) or (14.2), show that the coefficients of reflection and refraction are in this case

$$r = \frac{(n_1 - n_2)}{(n_1 + n_2)}$$

and

$$t = \frac{2n_1}{(n_1 + n_2)}.$$

Note that in this case we do not have to distinguish between π- and σ-components. (b) Draw the electric and magnetic fields in the incident, the reflected, and the refracted waves when $n_1 < n_2$ and when $n_1 > n_2$.

14.3 (a) Light falls perpendicularly on a glass plate ($n = 1.5$). Find the coefficients of reflection and transmission. (b) Repeat the calculation if the light is passing from the glass into the air. (c) Discuss in each case the changes of phase. (*Hint:* Use the results of Problem 14.2.)

14.4 Referring to the situation described in Problem 14.2, and using Eq. (11.16), compute the intensities of the reflected and the refracted waves and show that their sum is equal to the intensity of the incident wave. (*Hint:* Note that in Eq. (11.16) we must now replace c by the velocity $v = c/n$ in the medium, and that $E = \epsilon \mathscr{E}^2$. Also $n \approx \sqrt{\epsilon_r}$.)

14.5 The index of refraction of glass is 1.50. Compute the angles of incidence and of refraction when the light reflected from a glass interface is completely polarized.

14.6 The critical angle of light in a certain substance is 45°. What is the polarizing angle?

14.7 (a) At what angle above the horizontal must the sun be in order that sunlight reflected from the surface of a calm lake shall be completely polarized? (b) What is the plane of the $\mathscr{E}$ vector in the reflected light?

14.8 A plane linearly polarized light wave in air is incident on a medium of index n at the polarizing angle. The electric vector of the incident wave lies in the plane of incidence; its amplitude of oscillation is $\mathscr{E}_0$. Compute (a) the intensity of the incident wave, (b) the amplitude $\mathscr{E}'_0$ of the refracted wave, and (c) the intensity of the refracted wave. (d) Compare (a) with (c) and explain your result.

14.9 (a) Show that $r_\perp$ is positive for an electromagnetic wave if $n_{21} < 1$ and negative if $n_{21} > 1$. (b) Similarly, show that $r_\parallel$ is negative (positive) for angles of incidence smaller (larger) than the polarizing angle when $n_{21} < 1$ and positive (negative) when $n_{21} > 1$.

14.10 If a plane wave is polarized with its electric field making an angle α_i with the plane of incidence, show that the angles the electric field makes with the same plane in the refracted and in the reflected wave are

$$\tan \alpha_r = (t_\perp / t_\parallel) \tan \alpha_i$$

and

$$\tan \alpha'_r = (r_\perp / r_\parallel) \tan \alpha_i,$$

respectively.

14.11 A plane linearly polarized light wave in air ($n = 1$) is incident on a water surface ($n = 1.33$). Determine the amplitudes and phases of the refracted and the reflected waves relative to those of the incident wave for the following cases.

Angle of incidence	Angle between plane of incidence and plane of electric field
(a) 20°	0°
(b) 20°	90°
(c) 75°	0°
(d) 75°	90°

14.12 A plane linearly polarized light wave originating under water ($n = 1.33$) is refracted at the boundary surface between water and air ($n = 1$). For the following cases determine the amplitudes and the phases of the refracted and the reflected waves relative to those of the incident wave.

Angle of incidence	Angle between plane of incidence and plane of electric field
(a) 20°	0°
(b) 20°	90°
(c) 40°	0°
(d) 40°	90°

14.13 A beam of circularly polarized light in air ($n = 1$) is incident on a glass surface ($n = 1.52$) at an angle of 45°. Describe in detail the state of polarization of the reflected beam and of the refracted beam.

14.14 A polarizer and an analyzer are so oriented that the maximum amount of light is transmitted. To what fraction of its maximum value is the intensity of the transmitted light reduced when the analyzer is rotated through (a) 30°, (b) 45°, (c) 60°, (d) 90°, (e) 120°, (f) 135°, (g) 150°, and (h) 180°? (i) Plot I/I_{max} for a complete turn of the analyzer.

14.15 A parallel beam of linearly polarized light of wavelength 5.90×10^{-7} m (in vacuum) is incident on a calcite crystal as in Fig. 14-10. (a) Find the wavelengths of the ordinary and the extraordinary waves in the crystal. (b) Also find the frequency of each ray.

14.16 A beam of plane polarized light falls perpendicularly on a calcite plate (with sides cut parallel to the optical axis as in Fig. 14-10) with the electric vector making an angle of 60° with the optical axis. Find the ratio of (a) the amplitudes and (b) the intensities of the ordinary and the extraordinary beams.

14.17 Find the thickness of a calcite plate needed to produce a phase difference of (a) $\pi/2$ ($\frac{1}{4}$ wave), (b) π ($\frac{1}{2}$ wave), and (c) 2π (full wave) between the ordinary and the extraordinary rays for a wavelength of 6×10^{-7} m.

14.18 In Fig. 14-25 A and C are sheets of Polaroid whose transmission directions are as indicated. B is a sheet of doubly refractive material whose optical axis is vertical. All three sheets are parallel. Unpolarized light enters from the left. Discuss the state of polarization of the light at points (2), (3), and (4).

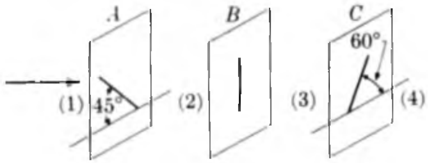


Figure 14-25

14.19 The wave described in Problem 11.5 is perpendicularly incident on a polarizer, which is rotated in its plane until the transmitted intensity is a maximum. (a) In which direction does the transmission axis of the polarizer lie? (b) In which direction does the transmission axis lie for minimum transmission? (c) Compute the ratio of the transmitted intensities for the positions found in (a) and (b).

14.20 A beam of white linearly polarized light is perpendicularly incident on a plate of quartz 0.865 mm thick, cut parallel to the optic axis, as in Fig. 14-10. The plane of the electric field is at an angle of 45° to the axis of the plate. The principal indexes of refraction of quartz for sodium light are listed in Table 14-2. Disregard the variation of $n_1 - n_2$ with wavelength. (a) Which wavelengths between 6.0×10^{-7} m and 7.0×10^{-7} m emerge from the plate linearly polarized? (b) Which wavelengths emerge circularly polarized? (c) Suppose that the beam emerging from the plate passes through an analyzer whose transmission axis is perpendicular to the plane of vibration of the incident light. Which wavelengths are missing in the transmitted beam?

14.21 It is known experimentally that for each kg of sugar dissolved in one m^3 of water, the rotation of the plane of polarization of a linearly polarized electromagnetic wave is $+6.65°$ per m of path. A tube 0.3 m long contains a sugar solution with 150 kg of sugar per m^3 of solution. Find the angle of rotation of polarized light.

14.22 Find the amount of sugar in a cylindrical tube 0.3 m long and 2×10^{-4} m^2 in cross section if the plane of polarization is rotated 39.7°. (*Hint:* See preceding problem.)

CHALLENGING PROBLEMS

14.23 A glass plate (index n_g) is coated with a thin plastic film (index n_c) (Fig. 14-26). (a) Designating the index of air by n_a, show that at normal incidence the reflection coefficients at the interface between air and coating and between coating and glass are equal if $n_c = \sqrt{n_g n_a}$. (b) Find the ratio of the reflection coefficients when the angle of incidence is 10° and n_g is 1.52.

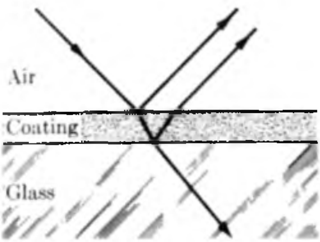


Figure 14-26

14.24 Consider two transparent media, (1) and (2), separated by a plane surface (Fig. 14-27). If r and t are the coefficients of reflection and refraction for a ray incident in medium (1), and r′ and t′ the same coefficients when the ray is incident in medium (2), show that $t't = 1 - r^2$ and $r = -r'$. These are called *Stokes' relations*. The second relation indicates that the reflection coefficients are of opposite sign: and if for one of the reflections there is no phase change, for the other there must be a phase change of π.

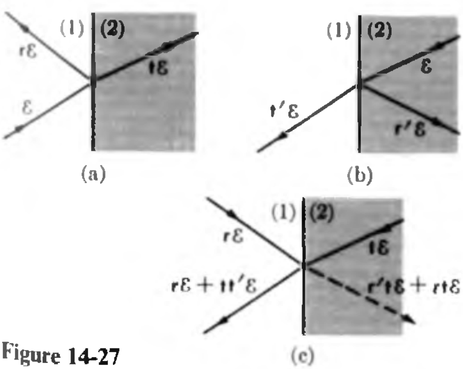


Figure 14-27

(*Hint:* Assume that rays r$\mathscr{E}$ and t$\mathscr{E}$ are reversed in direction as shown in Fig. 14-27(c), and recognize that in this case the final ray in medium (1) must be $\mathscr{E}$ and that no final ray must exist in medium (2).)

14.25 What is the state of polarization of the light transmitted by a *quarter-wave plate* when the electric vector of the incident linearly polarized light makes an angle of 30° with the optic axis?

14.26 A *Babinet compensator* (Fig. 14-28) consists of two quartz wedges that can slide over each other. The wedges are so cut that their optical axes are perpendicular. Therefore the ordinary ray in one of them is the extraordinary ray in the other. Show that for any ray the phase difference is

$$\delta = (2\pi/\lambda)(n_1 - n_2)(l - l')$$

where $l = AB$ and $l' = BC$. Therefore, if one slides one wedge along the other, the phase difference can be varied continuously.

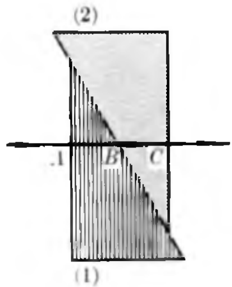


Figure 14-28

14.27 In a Babinet compensator, the effective width l of one wedge is 2 mm. Find the width the other must have in order to produce a phase difference of $2\pi/3$ in either direction when using light with wavelength 5.7×10^{-7} m.

14.28 Figure 14-29 represents a *Wollaston prism* made of two prisms of quartz cemented together. The optic axis of the right-hand prism is perpendicular to the page whereas that of the left-hand prism is parallel. The incident light is

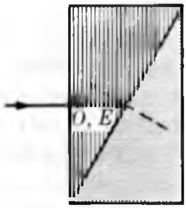


Figure 14-29

normal to the surface and gives rise to ordinary and extraordinary rays that travel in the left-hand prism along the same path but with different speeds. Copy Fig. 14-29 and show on your diagram how the ordinary and the extraordinary rays are bent in going into the right-hand prism and thence into the air.

14.29 A beam of light, after passing through a *Nicol prism* N_1, traverses a cell containing a scattering medium. The cell is observed at right angles through another Nicol prism, N_2. Originally, the Nicol prisms are oriented until the brightness of the field seen by the observer is a maximum. (a) Prism N_2 is rotated through 90°. Is extinction produced? (b) Prism N_1 is now rotated through 90°. Is the field through N_2 bright or dark? (c) Prism N_2 is then restored to its original position. Is the field through N_2 bright or dark?

CHAPTER FIFTEEN

WAVE GEOMETRY

15.1 Introduction

In previous chapters we discussed some phenomena that occur when a wave passes from one medium to another in which the propagation is different. We not only analyzed what happens to the wave front, but also introduced the concept of a *ray* as one that is very useful for geometrical constructions. In this chapter we are going to elaborate more fully on the phenomena of reflection and refraction from the geometrical point of view, using the ray concept as the tool for describing the processes that occur at the surfaces of discontinuity. We shall also assume that the processes are only reflections and refractions and that no other changes occur at the wave surfaces. (We shall defer consideration of diffraction and scattering until Chapter 17.)

This way of looking at the subject is called *wave geometry*, or *ray tracing*. In particular for electromagnetic waves in the visible and near-visible regions, this viewpoint constitutes *geometrical optics*, a very important branch of applied physics.

The geometrical treatment developed in this chapter is adequate so long as the surface irregularities and other discontinuities encountered by the wave during its propagation have dimensions very large compared with the wavelength. As long as this condition is fulfilled, our treatment applies equally well to light waves, acoustical waves (especially ultrasonic), earthquake waves, etc. However, in our discussions we shall consider light waves, except when otherwise stated.

A characteristic example of the use of rays is the image produced by a *pinhole camera* (Fig. 15-1). Such a camera consists of a box with a very small hole in one side. If an object AB emitting light waves is placed in front of it, the rays Bb and Aa will form an image ab on the opposite side. This image is well defined when the hole is very small so that only a small fraction of the wave fronts pass through it, and therefore, for each point of the object, there is a corresponding point of the image. If the hole is too large, the image appears blurred because to each point of the object there corresponds a spot in the image. Further, the hole must not be so small that its radius is comparable with the wavelength of light because then diffraction effects begin to appear and the image ab again appears blurred (as will be discussed in Chapter 17).

15.2 Reflection at a Spherical Surface

We begin by considering the reflection of waves at a spherical surface. We must first establish certain definitions and sign conventions. The center of curvature C is the center of the spherical surface (Fig. 15-2) and the vertex O is the pole of the spherical cap. The line passing through O and C is called the *principal axis*. If we take our origin of coordinates at O, all quantities measured to the right of O are positive and all those to the left are negative.

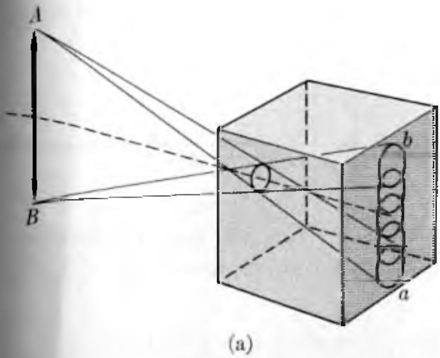
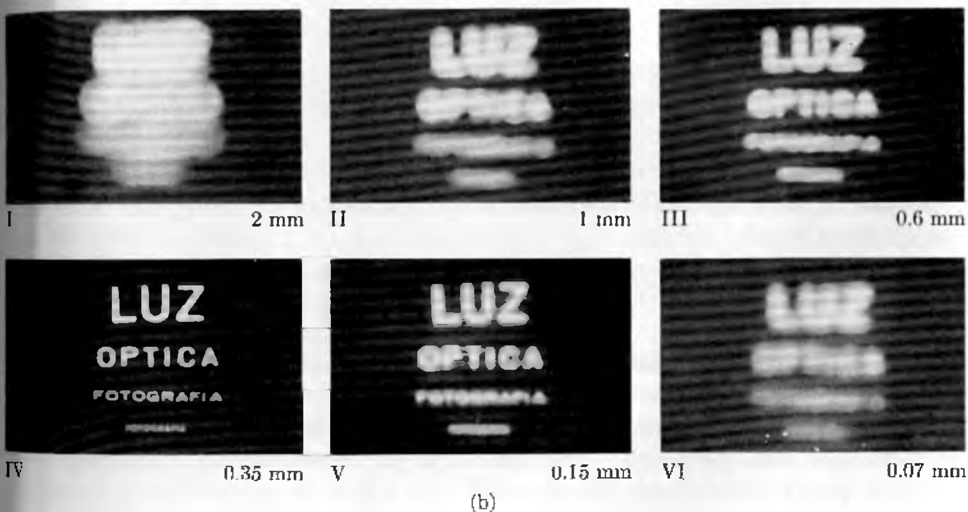


Fig. 15-1. (a) Image formation by a pinhole camera. The line drawing illustrates the ray paths. (b) The series of photographs shows the change in the clarity of the image as the diameter of the hole is decreased. Note that there is an optimum diameter for image clarity. (Photograph courtesy of Dr. N. Joel, UNESCO Pilot Project for the teaching of physics.)



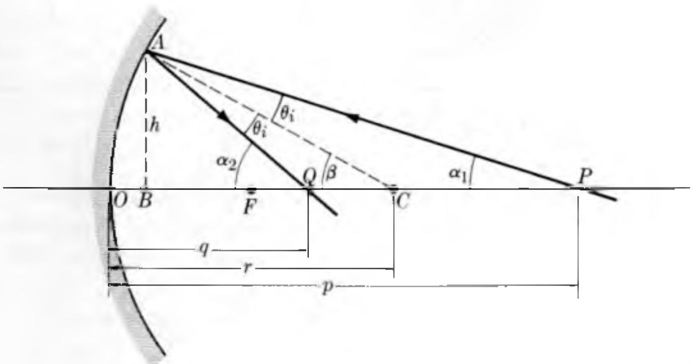


Fig. 15-2. Path of a ray reflected at a spherical surface.

Suppose that point P is a source of spherical waves. The ray PA is reflected as the ray AQ, and since the angles of incidence and reflection are equal, we have

$$\beta=\theta_i+\alpha_1 \qquad \text{and} \qquad \alpha_2=\beta+\theta_i,$$

resulting in

$$\alpha_1+\alpha_2=2\beta. \tag{15.1}$$

Assuming that the angles α_1, α_2, and β are very small (i.e., the rays are *paraxial*) and that the distance OB is therefore very small relative to the distances OQ, OC, and OP, we may write with good approximation

$$\alpha_1 \approx \tan \alpha_1=\frac{AB}{BP}\approx\frac{h}{p},$$

$$\alpha_2 \approx \tan \alpha_2=\frac{AB}{BQ}\approx\frac{h}{q},$$

$$\beta \approx \tan \beta=\frac{AB}{BC}\approx\frac{h}{r}.$$

Substituting these approximate values for α_1, α_2, and β in Eq. (15.1), and canceling the common factor h, we get

$$\frac{1}{p}+\frac{1}{q}=\frac{2}{r}, \tag{15.2}$$

which is *Descartes' formula for reflection at a spherical surface*. In Eq. (15.2), p is called the *object distance*, q is called the *image distance*, and r is the radius of the surface. This formula implies that under the approximation used in its derivation, all incident rays passing through P will go through Q after reflection at the surface. We say then that Q is the *image* of the *object P*.

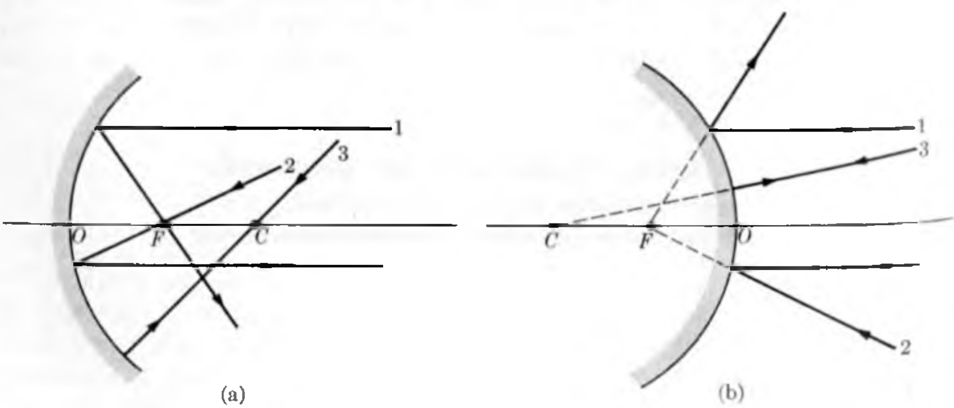


Fig. 15-3. Principal rays in spherical mirrors. (a) Concave and (b) convex.

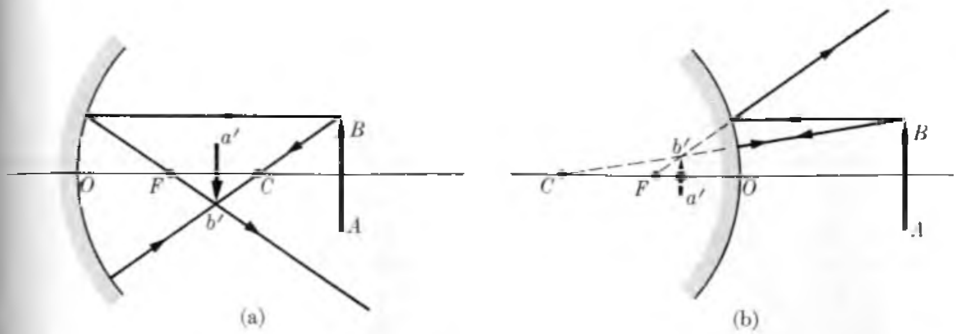


Fig. 15-4. Image construction in spherical mirrors. (a) Concave and (b) convex.

For the special case in which the incident ray is parallel to the principal axis, a situation which is equivalent to placing the object at a very large distance from the mirror, the object distance is infinite; that is, $p=\infty$. Then Eq. (15.2) becomes $1/q=2/r$ and the image falls at the point F, at a distance from the mirror given by $q=r/2$. Point F is called the *focus* of the spherical mirror, and its distance OF from the mirror is called the *focal length*, designated by f, so that $f=r/2$. Then Eq. (15.2) can be written in the form

$$\frac{1}{p}+\frac{1}{q}=\frac{1}{f}. \tag{15.3}$$

Since f can be determined experimentally by observing the point of convergence of rays that are parallel to the principal axis, it is not necessary to know the radius r in order to apply Eq. (15.3). Note that when the object is placed at the focal point so that $p=f$, then $q=\infty$. That is, all incident rays that pass through the focus F are reflected parallel to the principal axis. Although the derivation has been made for concave mirrors, the Descartes formula is equally valid for convex mirrors as long as the sign conventions for p, q, and r are taken into account. According to our sign convention, concave surfaces have a positive value for r, and convex surfaces have a negative value for r. Therefore, the signs of the corresponding focal lengths are positive and negative, respectively.

Figure 15-3 shows those three rays called *principal rays* for both a concave and a convex surface:

Ray 1 is a parallel ray that passes through the focal point after reflection or seems to have come from the focus after reflection;
Ray 2 is a focal ray that is parallel to the principal axis after reflection;
Ray 3 is a central ray that passes through the center of curvature on its way to the surface and after reflection as well. This ray strikes the surface at normal incidence ($\theta_i=0$).

In Fig. 15-4 these rays are used to illustrate the formation of an image by a spherical

Table 15-1. Sign Conventions in Spherical Mirrors

	+	−
Radius r	Concave	Convex
Focus f	Convergent	Divergent
Object p	Real	Virtual
Image q	Real	Virtual

reflecting surface. The *object* is AB and the *image* is $a'b'$. In Fig. 15-4(a) the image is *real* (since the reflected rays do cross), and in Fig. 15-4(b) the image is *virtual* (because the rays only seem to have crossed behind the mirror). Table 15-1 lists the sign conventions for mirrors that are used in this text. According to our convention, when the object or the image is *real* its respective distance to the mirror is *positive* because the image or the object is in front of the mirror; when either image or object is *virtual*, the distance is *negative* because it is behind the mirror.

When the aperture of the mirror is large, so that it accepts rays of large inclination, Eq. (15.3) is no longer a good approximation. In such a case, there is not a well-defined point image corresponding to a point object, but an infinite number of images; hence the image of an extended object appears blurred. Figure 15-5 shows the rays coming from the point P and reflected at the mirror. We see that the rays intersect not at the same point, but on a segment QQ' along the axis, an effect called *spherical aberration*. The point Q, which corresponds to the rays that make a very small angle with the axis, is determined by Eq. (15.3); Q' corresponds to an image of P from the rays that make a large inclination with the axis. The reflected rays also intersect along a surface, a section of which is indicated by the heavy line QS. This surface is called the *reflection caustic*.

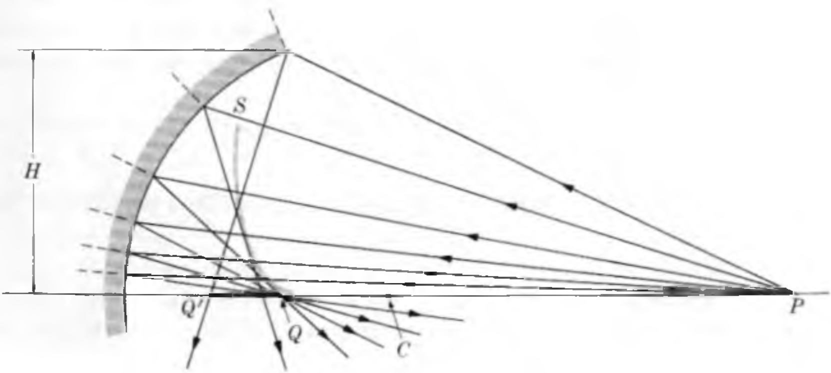


Fig. 15-5. Spherical aberration in a concave mirror.

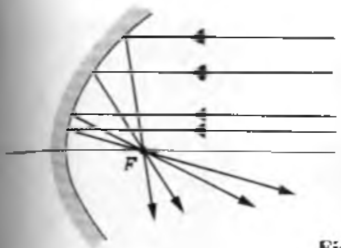


Fig. 15-6. Reflection of parallel rays from a parabolic mirror.

Spherical aberration cannot be completely eliminated. However, by proper design of the surface, it can be suppressed for certain positions, called *anastigmatic*. For a point object at the center of a spherical mirror, the image is exactly a point (also at the center), and it has no spherical aberration. Therefore the center of a spherical mirror is an *anastigmatic* position. Anastigmatic positions can be modified by changing the shape of the surface. For example, because of the property of the parabola, a parabolic mirror produces no aberration for rays that are parallel to the principal axis; they must all pass through the focus of the parabola (Fig. 15-6). Therefore parabolic mirrors are used in telescopes, not only for receiving rays in the visible region of the electromagnetic spectrum but also for receiving rays in the radiofrequency region, as in radiotelescopes (Fig. 15-7).

Other defects beside spherical aberration appear in the images produced by reflection (or refraction) at spherical surfaces. However, we shall not discuss these defects since they belong to rather specialized branches of optics.

Example 15.1. Image formation for a mirror whose aperture is large.

▼ When a spherical mirror has a large aperture and can accept rays of large inclination, the approximation made in obtaining Eq. (15.2) is no longer valid since the replacement of α by $\tan\alpha$ would then not be a very good approximation. It is not difficult to obtain another expression more precise than Eq. (15.2). When we apply the law of sines to triangles ACP and AQC (Fig. 15-8), we obtain

$$\frac{CP}{AP}=\frac{\sin\theta_i}{\sin(\pi-\beta)}, \qquad \frac{QC}{AQ}=\frac{\sin\theta_i}{\sin\beta}.$$

Then since $\sin(\pi-\beta)=\sin\beta$, we can combine the two expressions as

$$\frac{CP}{AP}=\frac{QC}{AQ} \quad \text{or} \quad \frac{p-r}{AP}=\frac{r-q}{AQ}.$$

Multiplying and dividing the left-hand side by pr and multiplying and dividing the right-hand side by rq give

$$\left(\frac{p-r}{pr}\right)\frac{pr}{AP}=\left(\frac{r-q}{rq}\right)\frac{rq}{AQ},$$

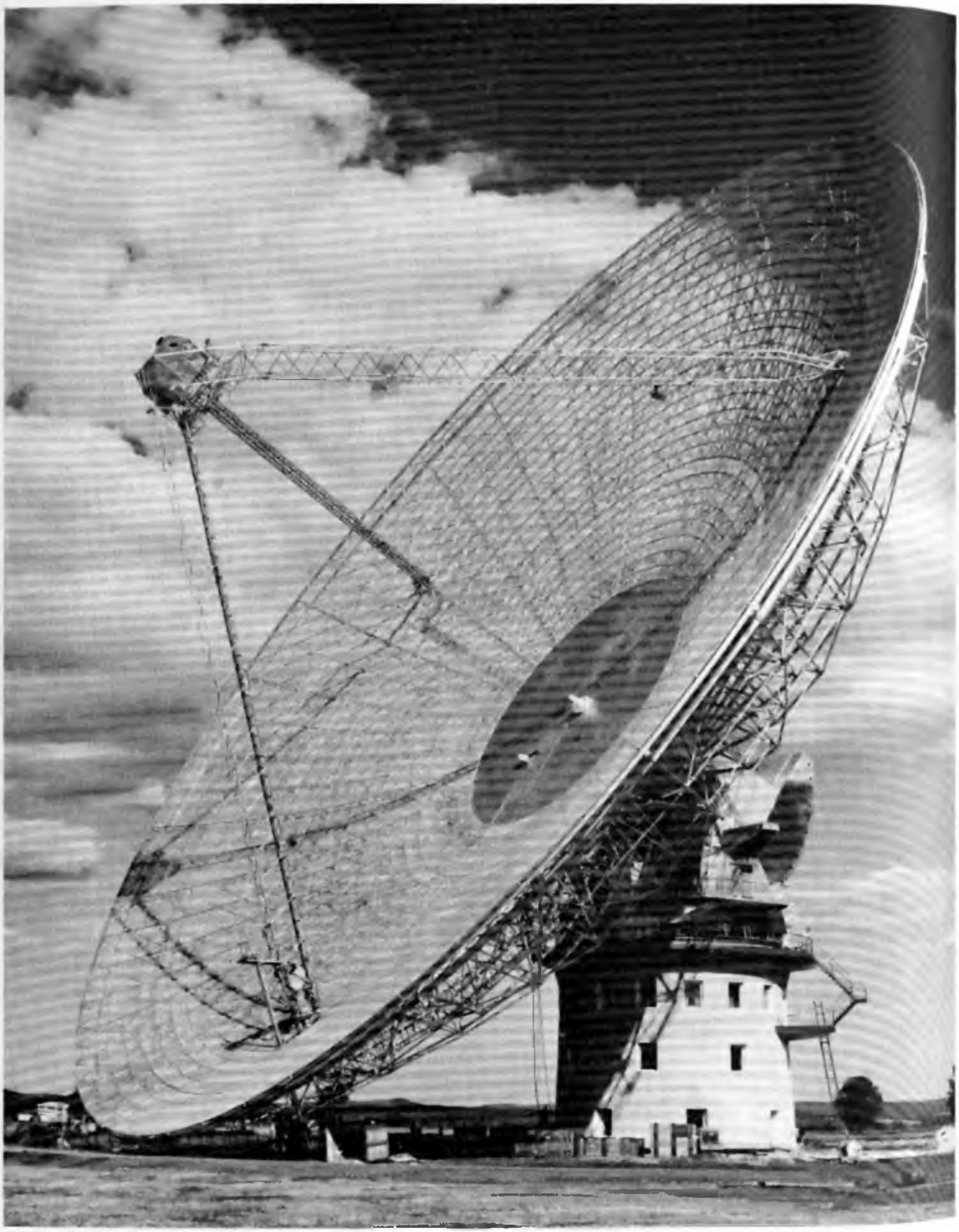


Fig. 15-7. Reflecting radiotelescope at Parkes, New South Wales, Australia. The reflector is 210 (64 m) in diameter. The reflector can rotate about the vertical as well as change in zenith angle, and is steerable over most of the visible sky. The telescope is designed for optimum performance at the 21-cm hydrogen-line wavelength, although it is still sensitive down to wavelengths of a few centimeters. The location, 340 km from Sydney, was chosen so that there would be minimum electrical interference. A radiotelescope consists of a metallic mirror formed by a wire mesh. A dipole receiving antenna is placed at the focus of the mirror. The signals received by the antenna correspond to electromagnetic waves propagating in a direction parallel to the axis of the mirror, and are transmitted to the laboratory for analysis.
(Photograph courtesy of the Australian News and Information Bureau.)

which may be written in the form

$$\left(\frac{1}{r}-\frac{1}{p}\right)\frac{p}{AP}=\left(\frac{1}{q}-\frac{1}{r}\right)\frac{q}{AQ}. \tag{15.4}$$

This relation involves no approximation. If α_1 and α_2 are very small, we can make the approximation $p \cong AP$ and $q \cong AQ$, recovering Eq. (15.2). However, we can go one step further before making such an approximation; that is, we will require only that $h \ll r$. From the triangle ACP and use of the law of cosines we get

$$\begin{aligned}
AP^2 &= r^2+(p-r)^2+2r(p-r)\cos\beta\\
&= p^2-2r(p-r)(1-\cos\beta)\\
&= p^2-4r(p-r)\sin^2\tfrac{1}{2}\beta\\
&= p^2\left[1-4\frac{r^2}{p}\left(\frac{1}{r}-\frac{1}{p}\right)\sin^2\tfrac{1}{2}\beta\right]\\
&\approx p^2\left[1-\frac{h^2}{p}\left(\frac{1}{r}-\frac{1}{p}\right)\right]
\end{aligned}$$

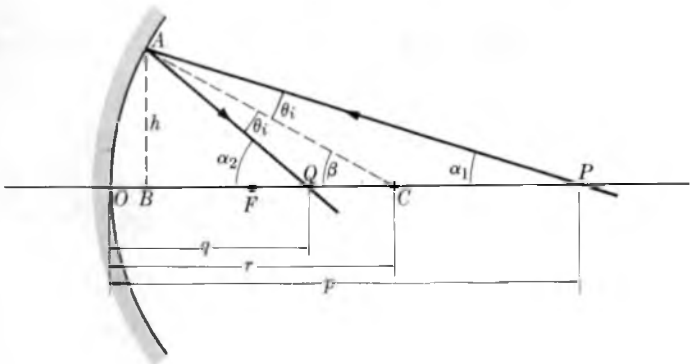


Figure 15-8

where in the last line the approximation $\sin \frac{1}{2}\beta \simeq \frac{1}{2}\beta \simeq h/2r$ has been made. Then

$$\frac{p}{AP}=\left[1-\frac{h^2}{p}\left(\frac{1}{r}-\frac{1}{p}\right)\right]^{-1/2}=1+\frac{h^2}{2p}\left(\frac{1}{r}-\frac{1}{p}\right) \tag{15.5}$$

where the approximation $(1-x)^{-1/2}=1+\frac{1}{2}x$ has been made. In the same way, using the triangle AQC, we have

$$\frac{q}{AQ}=1+\frac{h^2}{2q}\left(\frac{1}{r}-\frac{1}{q}\right). \tag{15.6}$$

Then substituting Eqs. (15.5) and (15.6) in the exact Eq. (15.4), we get

$$\left(\frac{1}{r}-\frac{1}{p}\right)\left[1+\frac{h^2}{2p}\left(\frac{1}{r}-\frac{1}{p}\right)\right]=\left(\frac{1}{q}-\frac{1}{r}\right)\left[1+\frac{h^2}{2q}\left(\frac{1}{r}-\frac{1}{q}\right)\right].$$

Multiplying and grouping terms, we obtain

$$\frac{1}{p}+\frac{1}{q}=\frac{2}{r}+\frac{h^2}{2}\left[\frac{1}{p}\left(\frac{1}{r}-\frac{1}{p}\right)^2+\frac{1}{q}\left(\frac{1}{r}-\frac{1}{q}\right)^2\right].$$

Since the second term on the right-hand side is a corrective one, we may use Eq. (15.2) to eliminate q in that term. This results in

$$\frac{1}{p}+\frac{1}{q}=\frac{2}{r}+\frac{h^2}{r}\left(\frac{1}{r}-\frac{1}{p}\right)^2. \tag{15.7}$$

The distance h is determined by the inclination of the rays coming from P: and for a given p, the larger h, the smaller q. Therefore, all rays emanating from a point P (Fig. 15-5) on the principal axis intersect not at the same point, but on a segment QS as previously indicated. We obtain the point Q by using Eq. (15.2) or by setting $h=0$ in Eq. (15.7). The point Q' of Fig. 15-5 is found by making $h=H$ where H is the radius of the base of the spherical surface. ▲

Example 15.2. A concave mirror has a radius of 0.600 m. An object is placed 1.000 m from the mirror. Assuming that its aperture is 20°, find the closest and farthest images produced by the mirror.

▼ In this case we have $r=+0.600$ m and $p=+1.000$ m. Therefore, for paraxial rays, we have, using Eq. (15.2), that

$$\frac{1}{1.000}+\frac{1}{q}=\frac{2}{0.600} \qquad \text{or} \qquad q=+0.429 \text{ m}.$$

The rays with maximum inclination produce an image obtained by using Eq. (15.7), with $h=r\sin\beta$ and $\beta=\frac{1}{2}(20°)=10°$. Therefore, $h=0.600 \sin 10°=0.104$ m and $h^2=0.011$. Thus

$$\frac{1}{1.000}+\frac{1}{q}=\frac{2}{0.600}+\frac{0.011}{0.600}\left(\frac{1}{0.600}-\frac{1}{1.000}\right)^2$$

or $q=+0.427$ m. Therefore the images occupy a small segment, of length about 0.002 m = 2 mm, along the principal axis. ▲

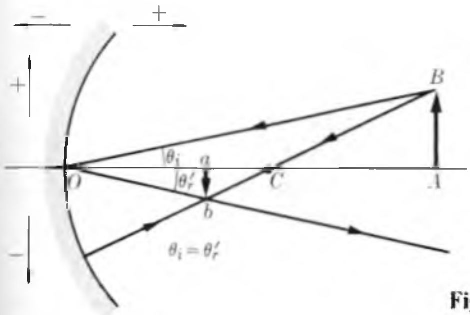


Fig. 15-9. Calculation of magnification by a spherical mirror,

Example 15.3. Magnification produced by a spherical mirror.

▼ The *magnification* M of an optical system is defined as the ratio of the size of the image to that of the object. That is, $M = ab/AB$. From Fig. 15-9 we see that

$$\tan \theta_i = \frac{AB}{OA} = \frac{AB}{p}.$$

$$\tan \theta_r' = \frac{ab}{Oa} = \frac{ab}{q}.$$

Thus, considering that $\theta_i = \theta_r'$, we have

$$\frac{\tan \theta_r'}{\tan \theta_i} = 1 = \left(\frac{ab}{q}\right)\frac{p}{AB}$$

or

$$M = \frac{ab}{AB} = \frac{q}{p}. \quad ▲ \tag{15.8}$$

Example 15.4. Use of Fermat's principle to discuss reflection at a spherical surface.

▼ In order that a point source P (Fig. 15-10) in front of a reflecting surface will produce an image at Q, the shape of the surface must be such that according to Fermat's principle, all rays require

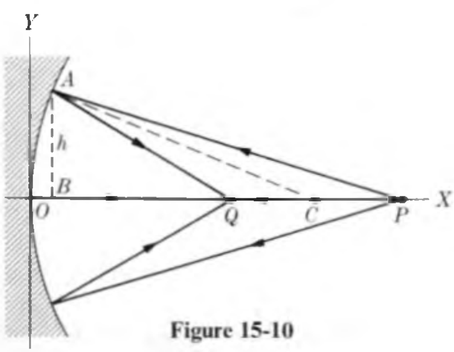


Figure 15-10

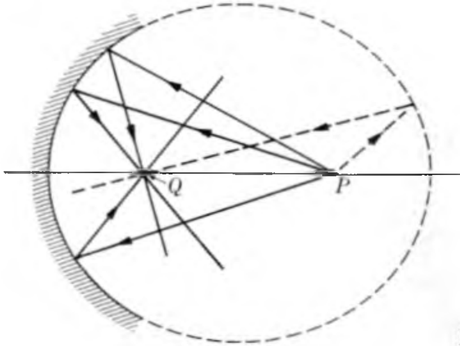


Fig. 15-11. Elliptical mirror.

the same amount of time to travel from P to Q. The time necessary for a ray to travel along the principal axis is

$$t=\frac{1}{c}(PO+OQ).$$

For a ray incident on the surface at A, we have that

$$t'=\frac{1}{c}(PA+AQ),$$

and we require that $t=t'$, at least in the first order of approximation. Note that this would be impossible if the surface were plane (such as OY) because if A were on OY, then $PA>PO$ and $AQ>OQ$ always, resulting in $t'>t$. But by curving the surface, both PA and AQ may be adjusted so that $t=t'$ holds. For this to be possible with a spherical surface, the requirement is that

$$PA+AQ=PO+OQ. \tag{15.9}$$

From triangle ABP, we have

$$h^2=PA^2-BP^2=(PA-BP)(PA+BP). \tag{15.10}$$

But if A is sufficiently close to O, we have that PA is slightly larger than PO and BP slightly smaller. Therefore we may write as a good approximation $PA+BP\approx 2PO=2p$. Substituting this value in Eq. (15.10) and solving for PA give

$$PA\approx BP+\frac{h^2}{2p}.$$

Similarly, from triangle ABQ we obtain

$$AQ\approx BQ+\frac{h^2}{2q}.$$

Substituting these equations into Eq. (15.9), we find that the requirement $t=t'$ is equivalent to

$$\left(BP+\frac{h^2}{2p}\right)+\left(BQ+\frac{h^2}{2q}\right)=PO+OQ$$

or

$$\frac{h^2}{2p}+\frac{h^2}{2q}=(PO-BP)+(OQ-BQ)=2OB. \tag{15.11}$$

But from triangle ABC, we have that $OB=h^2/2r$, if we neglect OB^2 compared with r^2. This is acceptable so long as A is close to O (i.e., if rays are paraxial). Therefore if we substitute in Eq. (15.11) and eliminate the common factor $\frac{1}{2}h^2$, we obtain Descartes' formula:

$$\frac{1}{p}+\frac{1}{q}=\frac{2}{r}.$$

A further step would be to see if, by means of a suitable surface, we could satisfy Eq. (15.9) rigorously, at least for a pair of points P and Q. We note that in this case Eq. (15.9) would be equivalent to $PA+AQ=\text{const.}$ This is the equation of an ellipsoid of revolution whose foci are at P and Q, as indicated in Fig. 15-11, and is the shape of the reflecting surface for which the image of P is rigorously at Q (that is, there is no spherical aberration for this pair of points). For all other points there will be aberration, the amount depending on the distance of the object point from one of these two chosen points. ▲

15.3 Refraction at a Spherical Surface

We shall now consider the refraction of waves at a spherical surface separating two media whose absolute indexes of refraction are n_1 and n_2 (Fig. 15-12). The fundamental geometric elements are the same as those defined in the previous section for spherical mirrors. The sign conventions are essentially the same as those used for spherical mirrors.

1. The radius of the refracting surface is *positive* if the surface is *concave* toward the object, and *negative* if it is *convex*.
2. The distance p of the object is *positive* if the object is *real* (that is, the object is to the right of the surface), and *negative* if the object is *virtual* (that is, the object is to the left of the surface).

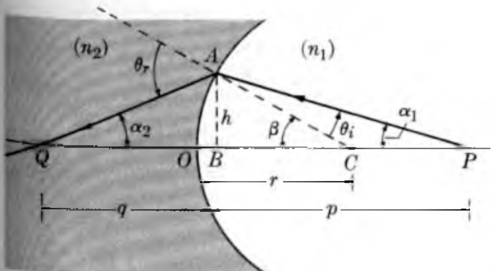


Fig. 15-12. Image formation by refraction at a convergent spherical surface.

3. The distance q of the image is *positive* if the image is *real* (that is, the image falls to the *left* of the surface), and *negative* if the image is *virtual* (that is, the image falls to the *right* of the surface).

These sign conventions are summarized in Table 15-2.

Let us consider a concave refracting surface. An incident ray such as PA is refracted along AD and intersects the principal axis at Q. From Fig. 15-12 we observe that in this case p, q, and r are all positive quantities. Also from the figure we have that $\beta=\theta_i+\alpha_1$ and $\beta=\theta_r-\alpha_2$*. From Snell's law, $n_1 \sin\theta_i=n_2 \sin\theta_r$. Assume, as we did in the preceding section, that the rays have a very small inclination. Then the angles θ_i, θ_r, α_1, α_2, and β are all very small, and we may use $\sin\theta_i \approx\theta_i$ and $\sin\theta_r \approx\theta_r$ so that Snell's law becomes $n_1\theta_i=n_2\theta_r$ or

$$n_1(\beta-\alpha_1)=n_2(\beta+\alpha_2). \tag{15.12}$$

From Fig. 15-12 we make the approximations

$$\tan\alpha_1=\alpha_1\approx\frac{h}{p},\quad \tan\alpha_2\approx\alpha_2\approx\frac{h}{q},\quad \tan\beta\approx\beta\approx\frac{h}{r}$$

so that when we substitute in Eq. (15.12), cancel common factors, and rearrange terms, we get

$$\frac{n_1}{p}+\frac{n_2}{q}=\frac{n_1-n_2}{r}, \tag{15.13}$$

which is *Descartes' formula for refraction at a spherical surface*. Although we have derived Eq. (15.13) for a concave refracting surface, this formula is valid also for convex refracting surfaces except that for convex surfaces, r is a negative number.

The *object focus* F_o, also called the *first focal point* of a spherical refracting surface, is the position of a point object on the principal axis such that the refracted rays are parallel to the principal axis. Thus the image of the point will be formed at infinity, or $q=\infty$ (Fig. 15-13a). The distance of the first focal point from the spherical surface is called the *object focal length*, designated by f_o. Setting $p=f_o$ and $q=\infty$ in Eq. (15.13), we have $n_1/f_o=(n_1-n_2)/r$, or

$$f_o=\left(\frac{n_1}{n_1-n_2}\right)r. \tag{15.14}$$

The focal length f_0 is *positive* when the object focus is *real* and is in front of the spherical surface. Then the system is called *convergent*; when the object focus is *virtual* the focal length f_o is *negative* and the system is called *divergent*. Table 15-2 lists the sign conventions used in this book.

Similarly, when the incident rays are parallel to the principal axis, a situation which is equivalent to having the object at a very large distance from the spherical surface

*Note that α_2 is *negative* to reflect the fact that q is negative according to our sign convention.

Table 15-2. Sign Conventions for a Spherical Refracting Surface

	+	−
Radius r	Concave	Convex
Focus f_o	Convergent	Divergent
Object p	Real	Virtual
Image q	Real	Virtual

($p=\infty$), the refracted rays pass through a point F_i on the principal axis called the *image focus* or *second focal point*. In this case the distance of the second focal point from the spherical surface is called the *image focal length*, designated by f_i. Making $p=\infty$ and $q=f_i$ in Eq. (15.13), we have $n_2/f_i=(n_1-n_2)/r$, or

$$f_i=\left(\frac{n_2}{n_1-n_2}\right)r, \tag{15.15}$$

and is positive when the image focus is real and is to the left of the surface (Fig. 15-13b). Note that $f_o-f_i=r$ and that the two focal points are always on opposite sides of the refracting surface.

The construction of the image of an object for a case in which $r>0$ and $n_1>n_2$ is shown in Fig. 15-14, in which the three *principal rays* for a single refracting surface

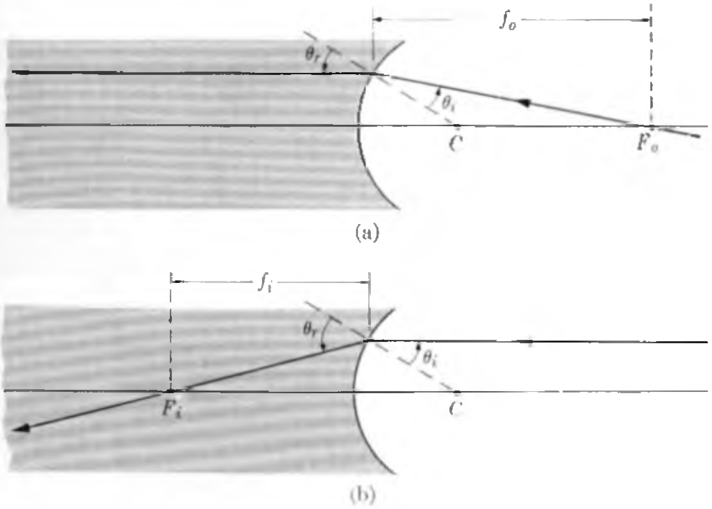


Figure 15-13

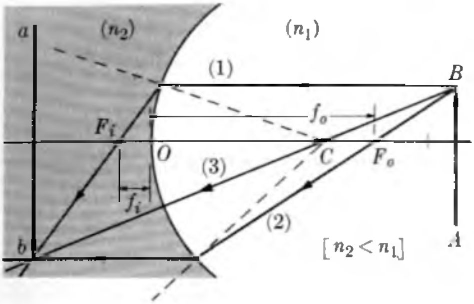


Fig. 15-14. Image formation by refraction at a convergent spherical surface using principal rays.

have been shown. The student should draw similar figures for the three remaining cases; that is, $r > 0$ and $n_1 < n_2$, $r < 0$, $n_1 \gtrless n_2$.

Equation (15.13) also indicates that for each point object there is a unique point image. This is acceptable so long as the spherical surface is of small aperture, admitting only rays of very small inclination so that our approximations are valid. For refracting spherical surfaces of large aperture, the situation is similar to that encountered in Fig. 15-5 for a spherical mirror, and results in the same phenomenon of spherical aberration discussed previously for spherical mirrors.

Example 15.5. Image formation by a refracting surface of large aperture.

▼ The procedure in this case is similar to that for a mirror. From triangles ACP and ACQ (Fig. 15-12), the law of sines allows us to write

$$\frac{CP}{AP} = \frac{\sin\theta_i}{\sin(\pi-\beta)} = \frac{\sin\theta_i}{\sin\beta}, \quad \frac{CQ}{AQ} = \frac{\sin(\pi-\theta_r)}{\sin\beta} = \frac{\sin\theta_r}{\sin\beta}.$$

Solving these equations for $\sin\theta_i$ and $\sin\theta_r$ and substituting their values in Snell's law, $n_1 \sin\theta_i = n_2 \sin\theta_r$, we get

$$n_1 \frac{CP}{AP} = n_2 \frac{CQ}{AQ} \quad \text{or} \quad n_1 \frac{p-r}{AP} = n_2 \frac{q+r}{AQ}.$$

This may be written in the form

$$n_1 \left(\frac{1}{r} - \frac{1}{p}\right) \frac{p}{AP} = n_2 \left(\frac{1}{r} + \frac{1}{q}\right) \frac{q}{AQ}, \tag{15.16}$$

which should be compared with Eq. (15.4). If α_1 and α_2 are very small, we can make the approximation $p \cong AP$ and $q \cong AQ$, recovering Eq. (15.13). However, we can go a step further by making use of the approximations equivalent to Eqs. (15.5) and (15.6) which are valid when $h \ll r$. These yield

$$\frac{p}{AP} = 1 + \frac{h^2}{2p}\left(\frac{1}{r} - \frac{1}{p}\right)$$

and

$$\frac{q}{AQ}=1-\frac{h^2}{2q}\left(\frac{1}{r}+\frac{1}{q}\right).$$

By substituting these two equations into Eq. (15.16), we get

$$n_1\left(\frac{1}{r}-\frac{1}{p}\right)\left[1+\frac{h^2}{2p}\left(\frac{1}{r}-\frac{1}{p}\right)\right]=n_2\left(\frac{1}{r}+\frac{1}{q}\right)\left[1-\frac{h^2}{2q}\left(\frac{1}{r}+\frac{1}{q}\right)\right].$$

Multiplying and grouping terms, we obtain

$$\frac{n_1}{p}+\frac{n_2}{q}=\frac{n_1-n_2}{r}+\frac{h^2}{2}\left[\frac{n_1}{p}\left(\frac{1}{r}-\frac{1}{p}\right)^2+\frac{n_2}{q}\left(\frac{1}{r}+\frac{1}{q}\right)^2\right].$$

Using Eq. (15.13) to eliminate q in the last corrective term, we finally obtain

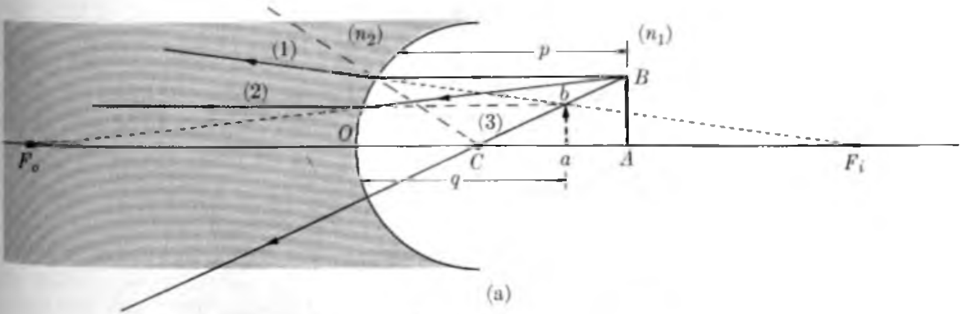
$$\frac{n_1}{p}+\frac{n_2}{q}=\frac{n_1-n_2}{r}+\frac{h^2}{2}\left(\frac{n_1-n_2}{n_2^2}\right)\left[\frac{n_1^2}{r}-\frac{n_1(n_1+n_2)}{p}\right]\left(\frac{1}{r}-\frac{1}{p}\right)^2. \tag{15.17}$$

As in the case of a spherical mirror, the position of the image depends on the value of h or on the slope of the incident ray. Therefore the image of a point originally on the principal axis is not necessarily another point on the principal axis. ▲

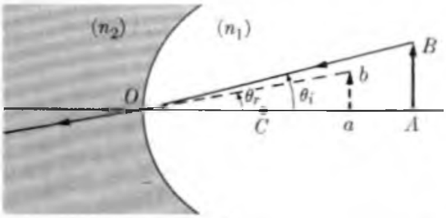
Example 15.6. Magnification produced by a spherical refracting surface.

▼ This problem is similar to that of Example 15.3. Considering Fig. 15-15a, in which AB is an object and ab is its (virtual) image, we have that $M=ab/AB$. We also have (Fig. 15-15b)

$$\tan\theta_i=\frac{AB}{OA}=\frac{AB}{p}, \qquad \tan\theta_r=\frac{ab}{Oa}=\frac{ab}{q},$$



(a)



(b)

Figure 15-15

and therefore

$$M=\frac{ab}{AB}=\frac{q\tan\theta_r}{p\tan\theta_i}\approx\frac{q\sin\theta_r}{p\sin\theta_i}$$

where the last approximation is valid whenever the angles are small. and we can replace the tangents by sines. Then, using Snell's law, $n_1 \sin\theta_i=n_2 \sin\theta_r$, we have

$$M=\frac{n_1 q}{n_2 p}, \tag{15.18}$$

We remind the reader that in using this relation, the distances q and p are used only with their absolute values. ▲

Example 15.7. A concave surface whose radius is $+0.50$ m separates a medium whose index of refraction is 1.20 from another whose index is 1.60. An object is placed in the first medium, 0.80 m from the surface. Determine the focal lengths of the system as well as the position of the image and its magnification.

▼ In this case $r=+0.50$ m, $n_1=1.20$ and $n_2=1.60$. Therefore, using Eqs. (15.14) and (15.15), we obtain

$$f_o=\frac{n_1 r}{n_1-n_2}=-1.50\text{ m},\qquad f_i=\frac{n_2 r}{n_1-n_2}=-2.00\text{ m}.$$

The system is therefore divergent. Using Eq. (15.13), we find that

$$\frac{1.20}{0.80}+\frac{1.60}{q}=\frac{1.20-1.60}{0.50}\qquad\text{or}\qquad q=-0.70\text{ m}.$$

The negative sign indicates that the image is virtual and the construction shown in Fig. 15-15a is approximately correct for this example. For the magnification we use the result of Example 15.6, using only absolute values for p and q.

$$M=\frac{1.20\times0.70\text{ m}}{1.60\times0.80\text{ m}}=0.65. \;▲$$

15.4 Lenses

A lens is a transparent medium bounded by two curved (usually spherical) surfaces, although one of the faces of the lens may be plane. An incident wave therefore suffers two refractions in going through the lens. For simplicity assume that the medium on both sides of the lens is the same and has an index of refraction of one (such as air) and the index of refraction of the lens is n. We shall also consider only thin lenses, i.e., lenses in which the thickness is very small compared with the radii.

The principal axis of a lens is the line determined by the two centers C_1 and C_2 (Fig. 15-16). Consider the incident ray PA. At the first surface the incident ray is refracted along ray AB. If extended, the ray AB would pass through Q', which is

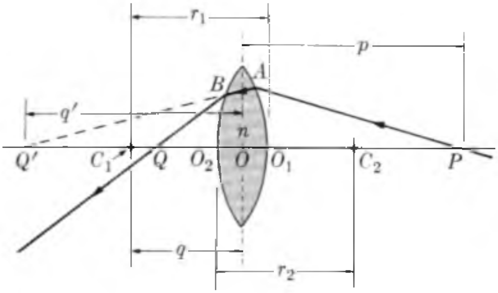


Fig. 15-16. Path of a ray through a lens.

therefore the image of P produced by the first refracting surface. The distances of the object and the image should be measured from either O_1 or O_2; but if the lens is very thin, we can neglect the thickness O_1O_2 and measure all distances from the common center point, O. The distance q' of Q' from O is obtained by applying Eq. (15.13); that is,

$$\frac{1}{p}+\frac{n}{q'}=\frac{1-n}{r_1}. \tag{15.19}$$

At B the ray suffers a second refraction and becomes ray BQ. We say that Q is the final image of P produced by the system of the two refracting surfaces that constitute the lens. Regarding the refraction at B, the object is the first image Q' and thus is a virtual object at a distance $-q'$ from the lens; and the image is Q, at a distance q from O. Therefore again applying Eq. (15.13) with p replaced by $-q'$, we have

$$\frac{n}{(-q')}+\frac{1}{q}=\frac{n-1}{r_2}. \tag{15.20}$$

Note that the order of the indexes of refraction has been reversed because in the second refraction the ray goes from the lens into the air. Combining Eqs. (15.19) and (15.20) to eliminate q', we find

$$\frac{1}{p}+\frac{1}{q}=(n-1)\left(\frac{1}{r_2}-\frac{1}{r_1}\right), \tag{15.21}$$

which is *Descartes' formula for a thin lens.* In writing this equation we must use for the radii r_1 and r_2 the same convention given in Table 15-2; that is, the radii are positive if the surface is concave and negative if it is convex when viewed from the side from which light reaches the lens.

The point O in Fig. 15-16 is chosen so that it coincides with the *optical center* of the lens. The optical center is a point defined such that any ray passing through it emerges in a direction parallel to the incident ray. To see that such a point exists, consider two parallel radii C_1A_1 and C_2A_2 in the lens of Fig. 15-17. Draw the corresponding tangents T_1 and T_2. For ray R_1A_1, whose direction is such that the refracted ray is A_1A_2, the emergent ray A_2R_2 is parallel to A_1R_1. From the similarity of triangles

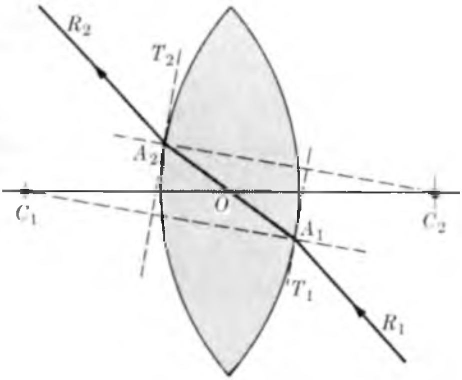


Fig. 15-17. Optical center of a lens.

C_1A_1O and C_2A_2O, we see that the point O is positioned so that

$$\frac{C_1O}{OC_2}=\frac{C_1A_1}{A_2C_2}=\frac{R_1}{R_2};$$

and therefore its position is independent of the particular ray chosen. Therefore all incident rays that pass through point O emerge without angular deviation; they will, however, have suffered some lateral displacement.

As in the case of a single refracting surface, the *object focus* F_o, or first focal point, of a lens is the position of the object for which the rays emerge parallel to the principal axis ($q=\infty$) after traversing the lens. The distance of F_o from the lens is called the *object focal length*, designated by f. Then setting $p=f$ and $q=\infty$ in Eq. (15.21), we obtain the object focal length as

$$\frac{1}{f}=(n-1)\left(\frac{1}{r_2}-\frac{1}{r_1}\right), \tag{15.22}$$

which is sometimes called the *lensmaker's equation*. Substituting Eq. (15.22) for the right-hand side of Eq. (15.21) we have

$$\frac{1}{p}+\frac{1}{q}=\frac{1}{f}. \tag{15.23}$$

This expression gives us a certain advantage in that if we determine f experimentally, we may use a lens without necessarily knowing its index of refraction or its radii.

For an incident ray parallel to the principal axis ($p=\infty$), the emergent ray passes through a point F_i having $q=f$ and called the *image focus*, or second focal point. Therefore, in a thin lens, the two foci are symmetrically located on both sides. If f is positive the lens is called *convergent*, and if it is negative, *divergent*. In the first case the object focus is real, and in the second case it is virtual. The sign conventions are the same as those given in Table 15-2 for a spherical refracting surface.

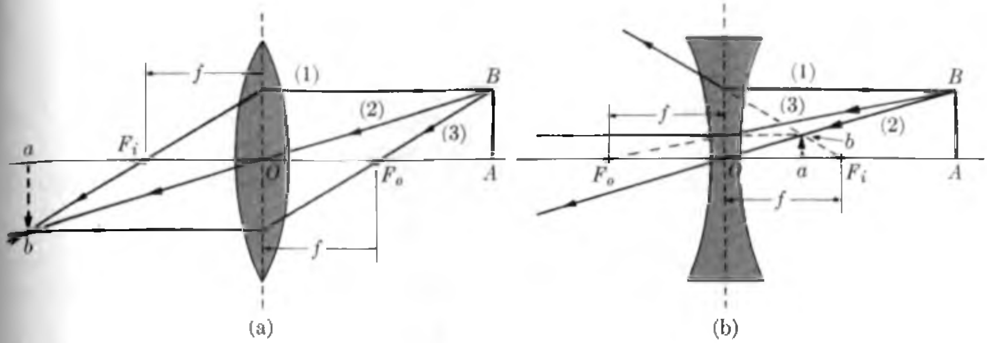


Fig. 15-18. Principal rays for (a) convergent and (b) divergent lenses.

For purposes of ray tracing, we may represent a thin lens by a plane perpendicular to the principal axis passing through O. Figure 15-18 shows the construction of the principal rays for a convergent and for a divergent lens.

The theory we have developed is correct only as long as the rays have a very small inclination so that spherical aberration is negligible. For lenses having a large diameter, the image of a point is not a point but a line segment of the principal axis. In particular, incident rays that are parallel to the principal axis intersect at different points, depending on their distance from that axis. *Spherical aberration* is measured by the difference $f' - f$ between the focal distance for a marginal ray and for an axial ray (Fig. 15-19). The refracted rays intersect over a conical surface called the *refraction caustic.*

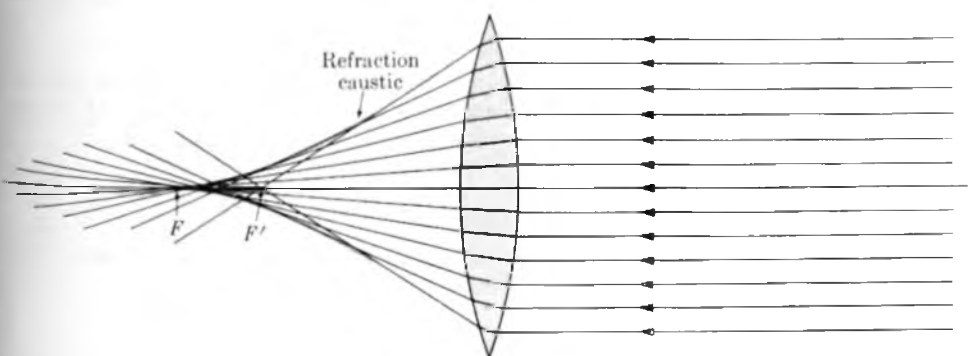


Fig. 15-19. Spherical aberration of a lens.

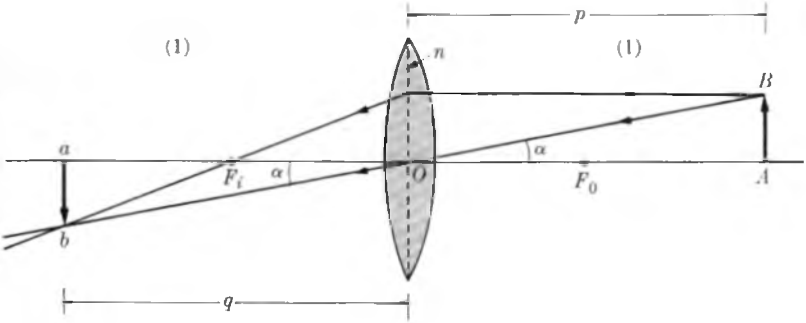


Fig. 15-20. Magnification produced by a lens.

Example 15.8. Magnification produced by a lens.

▼ As before, the magnification is defined as $M = ab/AB$. But from Fig. 15-20, if O is the optical center of the lens, we have that $\tan\alpha = AB/OA$ and $\tan\alpha = ab/Oa$. Therefore $ab/AB = Oa/OA$, or

$$M = \frac{q}{p}.$$

This relation could have been obtained by using the result of Example 15.6 for a spherical refracting surface since referring to Fig. 15-16, the magnification produced by the refraction at the first surface is $M_1 = q'/np$ while the magnification produced by the refraction at the second surface is $M_2 = nq/q'$. Therefore the total magnification is

$$M = M_1 M_2 = \frac{q'}{np} \times \frac{nq}{q'} = \frac{q}{p}. \quad \blacktriangle$$

Example 15.9. A spherical lens has two convex surfaces of radii 0.80 m and 1.20 m. Its index of refraction is $n = 1.50$. Find its focal length and the position of the image of a point 2.00 m from the lens.

▼ According to the sign conventions of Table 15-2, since the first surface looks convex and the second concave as seen from the side of the object that is placed on the right (see Fig. 15-16), we must write $r_1 = O_1C_1 = -0.80$ m, $r_2 = +1.20$ m. Therefore, using Eq. (15.22), we have

$$\frac{1}{f} = (1.50 - 1)\left(\frac{1}{1.20} - \frac{1}{-0.80}\right) \quad \text{or} \quad f = +0.96 \text{ m}.$$

The fact that f is positive indicates that this lens is convergent. To obtain the position of the image, we use Eq. (15.23) with $p = 2.00$ m and the above value of f, which yields

$$\frac{1}{2.00} + \frac{1}{q} = \frac{1}{0.96} \quad \text{or} \quad q = +1.85 \text{ m}.$$

The positive sign of q indicates that the image falls on the left side of the lens and is thus real.

Finally the magnification is

$$M = q/p = 0.92.$$

Since M is less than one, the image is smaller than the object. ▲

Example 15.10. Positions of the foci of a system of two thin lenses separated a distance L.

▼ The system of thin lenses illustrated in Fig. 15-21 shows in (a) the path of a ray passing through point P. The image of P produced by the first lens is Q'. Let p be the distance of the object from the first lens. Then the position of Q' is determined by

$$\frac{1}{p} + \frac{1}{q'} = \frac{1}{f_1}$$

where in this case q' is negative because Q' is a virtual image. Point Q' acts as a real object with respect to the second lens, producing a final image at Q. Since the distance of Q' from the second lens is $L + |q'| = L - q'$ (recall q' is negative in this case), we have that

$$\frac{1}{L - q'} + \frac{1}{q} = \frac{1}{f_2}$$

where q is the distance of the final image from the second lens. The set of equations above allows us to obtain the position of the image corresponding to any position of the object.

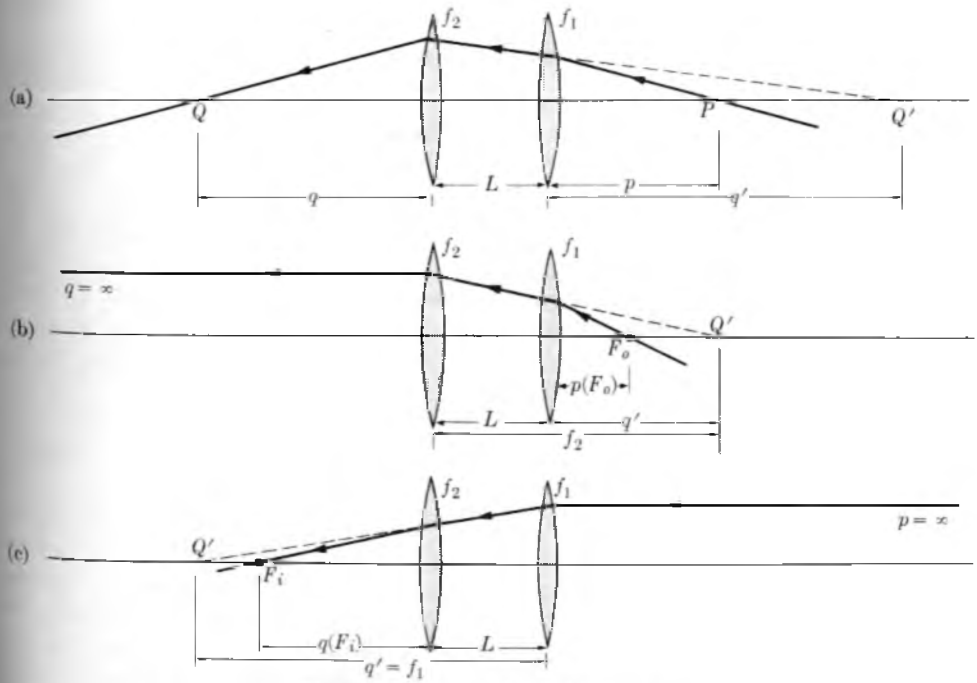


Fig. 15-21. System of two thin lenses.

The object focus F_o (Fig. 15-21b) of the lens system is the position of the object for which the image Q is at infinity ($q=\infty$). Designating the distance from F_o to the first lens by $p(F_o)$, we have from the second relation that $q'=L-f_2$, which when substituted in the first relation gives

$$p(F_o)=\frac{f_1(f_2-L)}{f_1+f_2-L}. \tag{15.24}$$

Similarly, for the position of the image focus F_i (Fig. 15-21c), designated by $q(F_i)$, we make $p=\infty$, resulting in $q'=f_1$ and having a positive value since now Q' is real:

$$q(F_i)=\frac{f_2(f_1-L)}{f_1+f_2-L}. \tag{15.25}$$

An important situation occurs when the two lenses are in contact so that L can be neglected. Then the equation relating q' and q becomes

$$\frac{1}{(-q')}+\frac{1}{q}=\frac{1}{f_2},$$

which, combined with the first equation, becomes

$$\frac{1}{p}+\frac{1}{q}=\frac{1}{f_1}+\frac{1}{f_2}.$$

This shows that a set of thin lenses in contact is equivalent to a single lens of focal length F given by

$$\frac{1}{F}=\frac{1}{f_1}+\frac{1}{f_2}. \qquad \text{(two lenses in contact)} \tag{15.26}$$

The same result can be obtained by making $L=0$ in the expression for $p(F_o)$. ▲

15.5 The Microscope

A microscope is a lens system producing an enlarged virtual image of a small object. The simplest microscope is a single convergent lens, commonly called a *magnifying glass*. The object AB (Fig. 15-22) is placed between the lens and the focus F_o, so that the image is virtual and falls at a distance q equal to the minimal distance of distinct vision, δ, which for a normal person is about 0.25 m. Since p is almost equal to f, particularly if f is very small, we may write for the magnification

$$M=\frac{q}{p}\approx\frac{\delta}{f}. \tag{15.27}$$

The *compound microscope* is more elaborate (Fig. 15-23). It consists of two convergent lenses, called the *objective* and the *eyepiece*, each of a small focal length. The focal length f of the objective is much smaller than the focal length f' of the eyepiece. Both f and f' are much smaller than the distance between the objective and the eyepiece. The object AB is placed at a distance from the objective slightly greater than f. The objective forms a real image $a'b'$, that acts as the object for the eyepiece. The image $a'b'$ must be at a distance from the eyepiece slightly less than

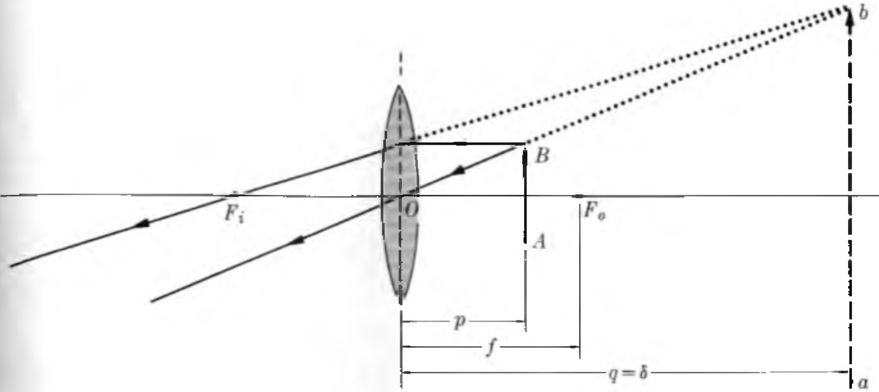


Fig. 15-22. Ray tracing in a magnifying glass.

f'. The final image, ab, is virtual, inverted, and much larger than the object. The object AB is so placed that ab is at a distance from the eyepiece equal to the minimal distance of distinct vision, δ (about 0.25 m). This condition is attained by the operation called *focusing*, which consists in moving the whole microscope relative to the object. The magnification of the objective is

$$M_O = \frac{a'b'}{AB} \approx \frac{L}{f},$$

and that of the eyepiece is

$$M_E = \frac{ab}{a'b'} \approx \frac{\delta}{f'}.$$

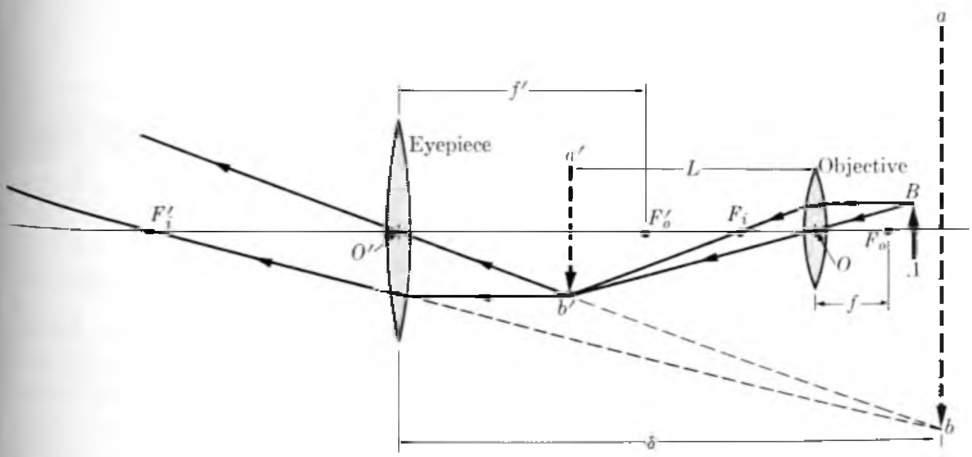


Fig. 15-23. Ray tracing in a compound microscope.

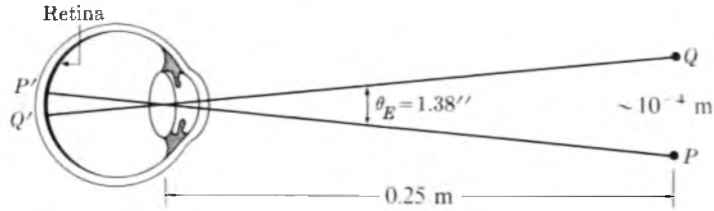


Fig. 15-24. Resolving power of the eye.

Therefore the total magnification is

$$M = M_O M_E = \frac{ab}{AB} = \frac{\delta L}{ff'}. \tag{15.28}$$

In an actual microscope, L is practically the same as the distance between the objective and the eyepiece.

The useful magnification in a microscope is limited by its *resolving power*; that is, the minimum distance between two points in the object that can be seen as distinct in the image. This resolving power is in turn determined by diffraction at the objective lens (see Chapter 17). A detailed calculation that is not reproduced here gives the resolving power as

$$R = \frac{\lambda}{2n \sin \theta} \tag{15.29}$$

where λ is the wavelength; n, the index of refraction of the medium in which the object is immersed; and θ, the angle a marginal ray makes with the axis of the microscope. In general $2n \sin \theta$ is about three so that $R \approx \frac{1}{3}\lambda$. On the other hand, the resolving power of the eye is about 10^{-4} m for an object at about 0.25 m (Fig. 15-24). Therefore the maximum useful magnification is

$$M = \frac{10^{-4}\ \text{m}}{\frac{1}{3}\lambda} \approx \frac{3 \times 10^{-4}\ \text{m}}{\lambda}.$$

For example, for light with $\lambda = 5 \times 10^{-7}$ m, which is about the center of the visible spectrum, M is about 600.

15.6 The Telescope

Another important optical instrument is the telescope, used to observe very distant objects. In the *refracting telescope*, the objective (Fig. 15-25) is a convergent lens having a very large focal length f, sometimes of several meters. Since the object AB is very distant, its image $a'b'$, produced by the objective, falls at its focus F_o. We have indicated only the central rays Bb' and Aa' since they are all that is necessary because we know the position of the image. The eyepiece, also a convergent lens, but

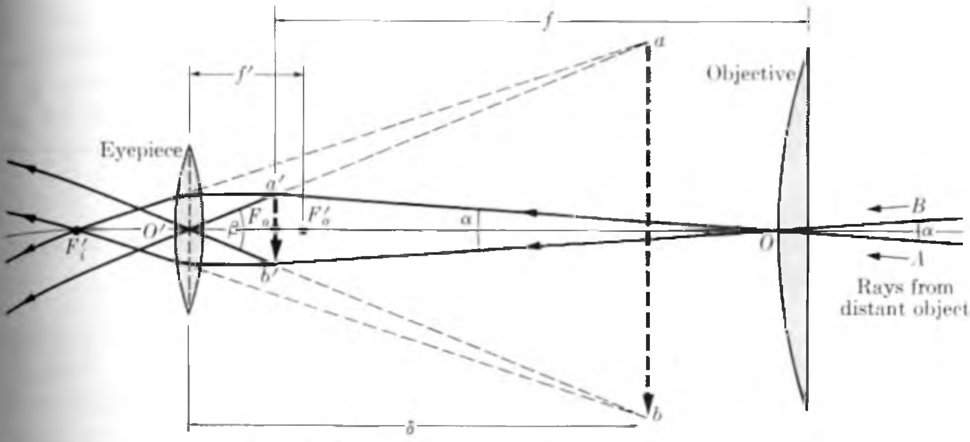


Fig. 15-25. Ray tracing in a refracting telescope.

of a much smaller focal length f' is placed so that the intermediate image $a'b'$ falls between O' and F_o, and the final (virtual) image ab is formed at the minimum distance for distinct vision. Focusing is performed by moving the eyepiece lens only since nothing is gained in this case by moving the objective lens.

The linear magnification produced by this instrument is not a useful concept because the image is usually much smaller than the object. Instead an *angular* magnification is defined as the ratio between the angle β subtended at the eye by the image when the telescope is used and the angle α subtended at the eye by the object when no instrument is used. This is written as

$$M=\frac{\beta}{\alpha}. \tag{15.30}$$

As shown in Fig. 15-25, the angle β is much larger than α, producing the sensation of magnification. From Fig. 15-25, considering that angles α and β are small, we may write

$$\alpha \approx \tan\alpha=\frac{a'b'}{f}, \qquad \beta=\tan\beta \approx \frac{a'b'}{f'}$$

because the distance from $a'b'$ to O' is practically f'. Substituting into Eq. (15.30), we get

$$M=\frac{f}{f'}. \tag{15.31}$$

Therefore, to obtain a large magnification, the focal length of the objective should be very large compared to that of the eyepiece. Practically, the length of the instrument is determined by the focal length f of the objective.

The magnification of an *astronomical telescope* is limited by the resolving power of the objective and of the eye of the observer. For an objective lens whose diameter is D, the resolving power (i.e., the minimum angle subtended by two points of the object AB that appear as distinct or different in the image $a'b'$) is, as will be shown in Chapter 17,

$$\theta \approx 1.22 \frac{\lambda}{D}. \tag{15.32}$$

On the other hand, the resolving power of the eye (Fig. 15-24), expressed in terms of an angle, is equal to

$$\theta_E = \frac{10^{-4}\ \text{m}}{0.25\ \text{m}} = 4 \times 10^{-4}\ \text{rad} = 1.38''.$$

Therefore the maximum useful magnification of a telescope is

$$M = \frac{\beta}{\alpha} = \frac{\theta_E}{\theta} = \frac{4 \times 10^{-4} D}{1.22\lambda} \approx 3.3 \times 10^{-4} \frac{D}{\lambda}. \tag{15.33}$$

A larger magnification means either a smaller value of α, which implies less detail in the image, or a larger value of β, which essentially does not reveal any new detail in the final image ab since the detail was not present in the intermediate image $a'b'$. For example, for light of $\lambda = 5 \times 10^{-7}$ m, we have $M \approx 660D$ where D is in meters. By increasing the diameter D of the objective, we can therefore increase the magnification. The Yerkes telescope, which is the largest existing refracting telescope, has a diameter of about 1 m, resulting in a magnification of about 660 and a resolving power of 10^{-1} second of arc.

In the *reflecting* telescope, the objective is a concave parabolic mirror. Thus an image free from spherical aberration is formed at the focus. The second largest reflecting telescope is at Mount Palomar. This telescope has a diameter of about 5 m and a magnification of the order of 3300.

15.7 The Prism

A prism is a medium bounded by two plane surfaces making an angle A (Fig. 15-26). We assume that the medium has an index of refraction n and that it is surrounded by a medium such as air, having unity index. The incident ray (Fig. 15-26) PQ suffers two refractions and emerges deviated an angle δ relative to the incident direction. From the figure the following relations may be seen to hold:

$$\sin i = n \sin r, \tag{15.34}$$

$$\sin i' = n \sin r', \tag{15.35}$$

$$r + r' = A, \tag{15.36}$$

$$\delta = i + i' - A. \tag{15.37}$$

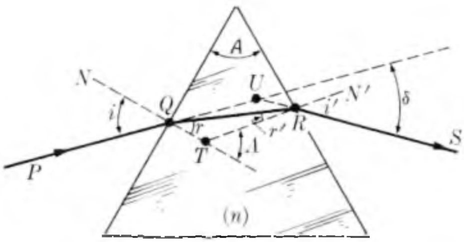


Fig. 15-26. Path of a ray through a prism.

The first and second equations are simply Snell's law applied to the refractions at Q and R. The third follows when we use triangle QTR, and the fourth when we use triangle QRU. The first three equations serve to trace the path of the ray, and the last allows us to find the deviation.

There is one particular path for which the deviation has a minimum value. This is obtained by making $d\delta/di=0$. From Eq. (15.37), we have

$$\frac{d\delta}{di}=1+\frac{di'}{di},$$

and for $d\delta/di=0$ we must have

$$\frac{di'}{di}=-1. \tag{15.38}$$

From Eqs. (15.35) and (15.36), since $dA=0$, we have

$$\cos i\, di=n\cos r\, dr, \quad \cos i'\, di'=n\cos r'\, dr', \quad \text{and} \quad dr=-dr'.$$

Therefore

$$\frac{di'}{di}=-\frac{\cos i\cos r'}{\cos i'\cos r}. \tag{15.39}$$

Since the four angles i, r, i', and r' are smaller than $\frac{1}{2}\pi$ and satisfy the symmetric conditions (15.34) and (15.35), Eqs. (15.38) and (15.39) can be satisfied simultaneously only if $i=i'$ and $r=r'$, which requires that

$$i=\tfrac{1}{2}(\delta_{\min}+A) \qquad \text{and} \qquad r=\tfrac{1}{2}A \tag{15.40}$$

where $\delta_{\min}$ is the value of the minimum deviation. Note that in this case the path of the ray is symmetric with respect to the two faces of the prism. Introducing Eq. (15.40) in Eq. (15.34), we obtain

$$n=\frac{\sin\frac{1}{2}(\delta_{\min}+A)}{\sin\frac{1}{2}A}. \tag{15.41}$$

Equation (15.41) is a convenient formula for measuring the index of refraction of a substance by finding $\delta_{\min}$ experimentally in a prism of known angle A.

15.8 Dispersion

When a wave is refracted into a dispersive medium whose index of refraction depends on the frequency (or wavelength), the angle of refraction will also depend on the frequency or wavelength. If the incident wave, instead of being composed of a single frequency (or *monochromatic*), is composed of several frequencies or wavelengths superposed, each component wavelength will be refracted through a different angle, a phenomenon called *dispersion*. (We introduced the subject of dispersion of electromagnetic waves in matter in Section 12.7.)

We remind the student that colors are associated with wavelength in a small interval of the electromagnetic spectrum. Therefore white light is decomposed into colors when refracted from air into another substance such as water or glass. If a piece of glass is in the form of a plate with parallel sides, the rays that emerge are parallel and the different colors are superposed again (Fig. 15-27), and no dispersion is observed except at the very edges of the image. Even so, this effect is not normally noticeable.

If the light passes through a prism (Fig. 15-28), the emerging rays are not parallel for the different colors and the dispersion is noticeable, especially at the edges. For that reason prisms are widely used for analyzing light in instruments called *spectroscopes*. A simple type of spectroscope is illustrated in Fig. 15-29. Light emitted by a source S is limited by a slit. The light is transformed into parallel rays by the lens L, set at a distance equal to its focal length from the slit. After being dispersed by the prism, the rays of different colors pass through another lens L'. Since all rays of the same color (or wavelength) are parallel, they are focused on the same point of the screen, set at a distance equal to the focal length of L'. But rays that differ in color (or wavelength) are not parallel; therefore different colors are focused on different points of the screen. The different colors or wavelengths emitted by the source S appear displayed on the screen in what is called the *spectrum* of the light coming from S. If the deviation δ varies rapidly with the wavelength λ, the colors appear widely spaced on the screen. For each wavelength of the source, a line on the screen appears that is the image of the slit. If the source emits the full spectrum of visible light, a *continuous* spectrum will appear on the screen.

The dispersion of a prism is defined by

$$D = \frac{d\delta}{d\lambda} = \frac{d\delta}{dn}\frac{dn}{d\lambda}. \tag{15.42}$$

The factor $d\delta/dn$ depends primarily on the geometry of the system: the factor $dn/d\lambda$ depends on the material composing the prism. By differentiating Eq. (15.41), we find that when the prism is arranged for average minimum deviation, we have

$$\frac{d\delta}{dn} = \frac{2 \sin \frac{1}{2}A}{\cos \frac{1}{2}(\delta_{\min} + A)}. \tag{15.43}$$

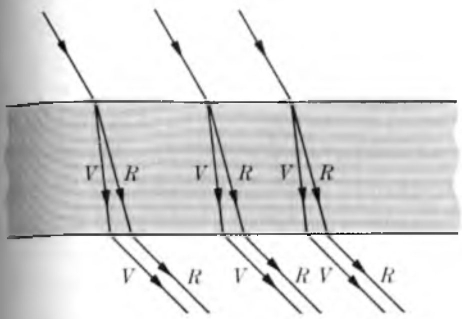


Fig. 15-27. Dispersion when light passes through a plate with parallel sides.

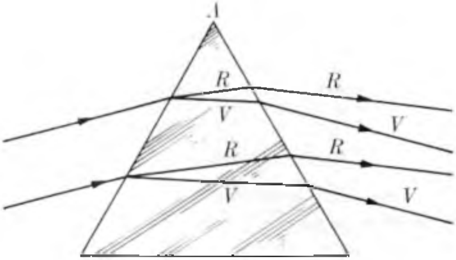


Fig. 15-28. Dispersion when light passes through a prism.

The second factor $dn/d\lambda$ in Eq. (15.42) depends on the nature of the waves and the medium. For electromagnetic waves in general and for light in particular, a satisfactory approximate expression for the index of refraction as a function of the wavelength is given by *Cauchy's formula*,

$$n = A_0 + \frac{B_0}{\lambda^2} \tag{15.44}$$

where A_0 and B_0 are constants characteristic of each substance. The variation of n with λ for various materials transparent in the optical region is shown in Fig. 15-30. From Eq. (15.44) we obtain

$$\frac{dn}{d\lambda} = -\frac{2B_0}{\lambda^3}.$$

The dispersion in a prism is then

$$D = \frac{d\delta}{d\lambda} = \frac{2 \sin \frac{1}{2}A}{\cos \frac{1}{2}(\delta_{min} + A)} \left(-\frac{2B_0}{\lambda^3} \right). \tag{15.45}$$

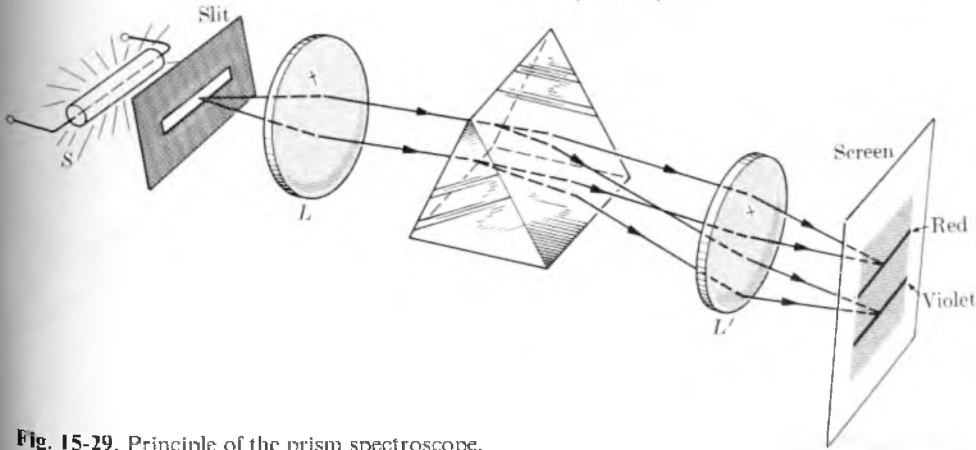


Fig. 15-29. Principle of the prism spectroscope.

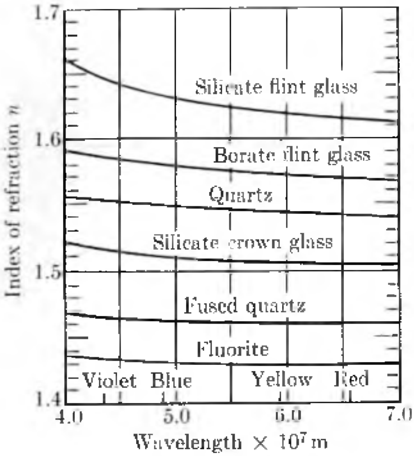


Fig. 15-30. Variation of refractive index with wavelength, in the visible region, for some materials.

The negative sign means that the deviation decreases when the wavelength increases so that the red is deviated less than the violet.

NOTE: Justification of Cauchy's formula.

In Eq. (12.17) we obtained an expression that relates the index of refraction as a function of the frequency of the electromagnetic waves and the characteristic frequencies of the substance. Assuming for simplicity that there is only one atomic frequency ω_0 and that $\omega \ll \omega_0$, we obtain

$$n^2 = 1 + \frac{Ne^2}{\epsilon_0 m(\omega_0^2 - \omega^2)}$$

so that using the binomial expansion, we get

$$\begin{aligned} n &= \left(1 + \frac{Ne^2}{\epsilon_0 m(\omega_0^2 - \omega^2)}\right)^{1/2} \approx 1 + \frac{Ne^2}{2\epsilon_0 m(\omega_0^2 - \omega^2)} \\ &\approx 1 + \frac{Ne^2}{2\epsilon_0 m\omega_0^2}\left(1 - \frac{\omega^2}{\omega_0^2}\right)^{-1} \\ &\approx 1 + \frac{Ne^2}{2\epsilon_0 m\omega_0^2}\left(1 + \frac{\omega^2}{\omega_0^2}\right). \end{aligned}$$

And since $\omega = 2\pi c/\lambda$, we have

$$n = A_0 + \frac{B_0}{\lambda^2}$$

where

$$A_0 = 1 + \frac{Ne^2}{2\epsilon_0 m\omega_0^2} \qquad \text{and} \qquad B_0 = \frac{2\pi^2 c^2 Ne^2}{\epsilon_0 m\omega_0^4}.$$

[The student should estimate the order of magnitude of A_0 and B_0.]

15.9 Chromatic Aberration

When light (such as white light) composed of several wavelengths passes through a lens, the light suffers dispersion, and the edges of the image produced by the lens appear colored. This effect is called *chromatic aberration*. It is easy to understand the reason for this effect when we recognize that a lens is equivalent to two prisms attached at their bases (for a convergent lens) or their vertexes (for a divergent lens).

A lens has a focus for each color or wavelength as is seen from Eq. (15.22) since f is determined by the index of refraction n, and n depends on the wavelength. For transparent substances whose index of refraction decreases with increasing wavelength in the visible region (see Fig. 15-30), violet has a shorter focal length than red. In Fig. 15-31 the chromatic aberration of a convergent and a divergent lens is shown for such a material.

The *longitudinal chromatic aberration*, A, in a lens is defined by the difference $f_C - f_F$ between the focal distances corresponding to the wavelengths 6.563×10^{-7} m and 4.862×10^{-7} m, emitted by hydrogen and designated as the C- and F-Fraunhofer lines. Using Eq. (15.22), we have

$$\frac{1}{f_C} = (n_C - 1)\left(\frac{1}{r_2} - \frac{1}{r_1}\right) \quad \text{and} \quad \frac{1}{f_F} = (n_F - 1)\left(\frac{1}{r_2} - \frac{1}{r_1}\right),$$

and thus

$$\frac{1}{f_F} - \frac{1}{f_C} = (n_F - n_C)\left(\frac{1}{r_2} - \frac{1}{r_1}\right). \tag{15.46}$$

The D-Fraunhofer line, with wavelength 5.890×10^{-7} m, corresponds approximately to the average index of refraction, designated by n_D where

$$\frac{1}{f_D} = (n_D - 1)\left(\frac{1}{r_2} - \frac{1}{r_1}\right). \tag{15.47}$$

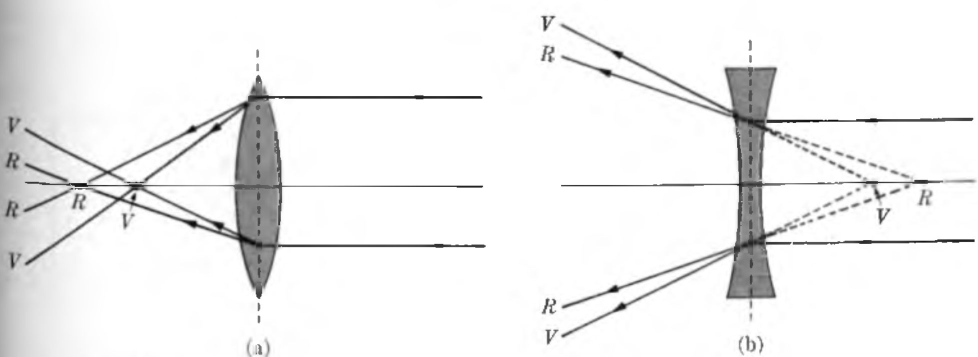


Fig. 15-31. Chromatic aberration in lenses. (a) Convergent and (b) divergent lens.

Table 15-3. Indexes of Refraction and Dispersive Power

Fraunhofer line	C	D	F	Dispersive power, ω
Wavelength $\times 10^7$ m	6.563	5.890	4.862	
Crown glass	1.514	1.517	1.523	0.0174
Flint glass	1.622	1.627	1.639	0.0271
Alcohol	1.361	1.363	1.367	0.0165
Benzene	1.497	1.503	1.514	0.0338
Water	1.332	1.334	1.338	0.0180

Using Eq. (15.47) we can eliminate the dependence of Eq. (15.46) on the radii of the lens and write

$$\frac{1}{f_F}-\frac{1}{f_C}=\left(\frac{n_F-n_C}{n_D-1}\right)\frac{1}{f_D}.$$

Also the left-hand side of this equation can be written as

$$\frac{f_C-f_F}{f_C f_F}\approx\frac{f_C-f_F}{f_D^2}$$

since to good approximation $f_C f_F \approx f_D^2$. Therefore the longitudinal chromatic aberration of the lens is

$$A=f_C-f_F=\left(\frac{n_F-n_C}{n_D-1}\right)f_D. \tag{15.48}$$

The quantity

$$\omega=\frac{f_C-f_F}{f_D}=\frac{n_F-n_C}{n_D-1} \tag{15.49}$$

is called the *dispersive power* of the material. Table 15-3 gives the indexes of refraction of some transparent materials at the C-, D-, and F-Fraunhofer lines.

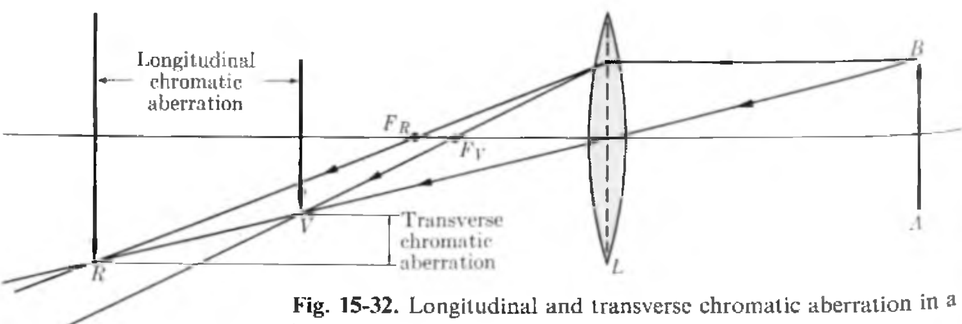


Fig. 15-32. Longitudinal and transverse chromatic aberration in a lens.

The kind of chromatic aberration we have discussed so far for lenses is called *longitudinal* because it is measured along the principal axis. There is also a *transverse* chromatic aberration. Consider an object AB in front of a lens L (Fig. 15-32). Unless the light from the object is monochromatic, the light will be dispersed as it goes through the lens; and instead of one image, a series of images differing in size will be formed, one for each wavelength or color. The figure shows only the extreme images corresponding to red and violet, and their separation has been greatly exaggerated. Because of this lateral dispersion, the edges of the images will appear colored. Transverse chromatic aberration can be expressed in terms of the different magnification for the C- and F-lines.

Example 15.11. Discussion of achromatic lenses.

Chromatic aberration can be reduced or even eliminated by combining lenses of different materials, resulting in an *achromatic system*. To see how this can be done, suppose we have the lens system in Fig. 15-33, in which, for example, lens L is made of crown glass and lens L' of flint glass. Call f and f' their respective focal lengths and ω and ω' their respective dispersive powers. Lens L would have the chromatic aberration indicated by the segment VR. But if the divergent lens L' is properly designed, rays of all wavelengths should focus at F. To see how, we recall from Example 15.10 and Eq. (15.26) that for each wavelength

$$\frac{1}{F_C}=\frac{1}{f_C}+\frac{1}{f'_C} \qquad \text{and} \qquad \frac{1}{F_F}=\frac{1}{f_F}+\frac{1}{f'_F}$$

where F_C and F_F are the corresponding focal lengths of the lens combination. Therefore, subtracting these two equations and remembering that $F_C F_F \approx F_D^2$, etc., we have

$$\frac{1}{F_F}-\frac{1}{F_C}=\frac{F_C-F_F}{F_D^2}=\frac{f_C-f_F}{f_D^2}+\frac{f'_C-f'_F}{f'^2_D}=\frac{\omega}{f_D}+\frac{\omega'}{f'_D}.$$

When there is no longitudinal chromatic aberration, we must have $F_C=F_F$ so that

$$\frac{\omega}{f_D}+\frac{\omega'}{f'_D}=0. \tag{15.50}$$

Since ω and ω' are positive, we conclude that f_D and f'_D are of opposite signs. Therefore one lens of an *achromat* is convergent and the other is divergent. ▲

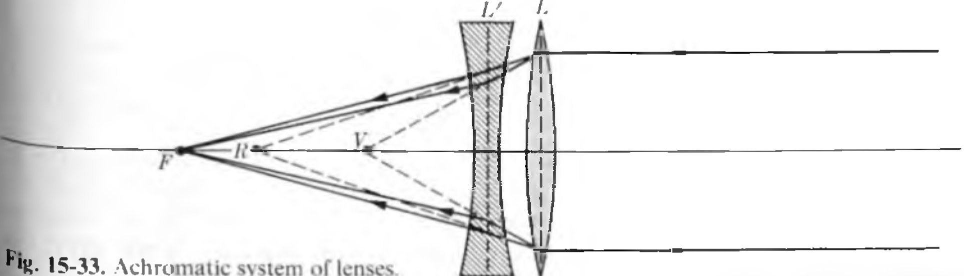


Fig. 15-33. Achromatic system of lenses.

Example 15.12. Design an achromatic converging lens system having a focal length of 0.350 m and made of two lenses, one of crown glass and the other of flint glass.

▼ From the data of Table 15-3, we have that the dispersive power ω of crown glass is 0.0193 and that of flint glass is 0.0271. Therefore the requirement of achromaticity imposed by Eq. (15.50) is expressed by

$$\frac{0.0193}{f_1}+\frac{0.0271}{f_2}=0 \qquad \text{or} \qquad \frac{f_2}{f_1}=-1.402.$$

A second requirement, using the result of Example 15.10 with $F=+0.35$ m, is that

$$\frac{1}{f_1}+\frac{1}{f_2}=\frac{1}{0.350}.$$

Combining the two equations, we then get

$$f_1=+0.1007 \text{ m}, \qquad f_2=-0.1414 \text{ m}.$$

Thus the crown lens is convergent and the flint lens is divergent. Assuming that the system is plano-convex, with the plane face corresponding to the flint lens, we have, using Eq. (15.22) with the value n_D for the index of refraction, that the radius of the face common to the two lenses is 0.089 m and the radius of the other face of the crown lens is 0.126 m. ▲

Problems

15.1 A concave mirror has a radius of 1.00 m. Find the position and the magnification of the image of an object if the object is at a distance from the mirror equal to (a) 1.40 m, (b) 1.00 m, (c) 0.80 m, (d) 0.50 m, and (e) 0.30 m. (f) Consider also a virtual object at a distance of 0.60 m.

15.2 A convex mirror has a radius of 1.00 m. Find (a) the position of the image of an object and (b) the magnification if the distance from the object to the mirror is 0.60 m. Also consider a virtual object at a distance of (c) 0.30 m, and (d) 0.80 m.

15.3 Determine the focal length and the nature of a spherical mirror if an object placed 1.20 m from the mirror produces an image that is (a) real and situated 0.80 m from the mirror, (b) virtual and situated 3.20 m from the mirror, (c) virtual and situated 0.60 m from the mirror, (d) real and twice as large, (e) virtual and twice as large, (f) real and one-third as large and (g) virtual and one-third as large.

15.4 A concave spherical mirror has a radius of 1.60 m. Find the position of the object if the image is (a) real and three times larger, (b) real and one-third as large, and (c) virtual and three times larger.

15.5 Repeat Problem 15.4 for a mirror that is convex.

15.6 A concave shaving mirror has a focal length of 0.15 m. Find the optimum distance of a person from the mirror if the distance of distinct vision is 0.25 m. What is the magnification?

15.7 A concave mirror produces a real, inverted image three times larger than the object at a distance of 0.28 m from the object. Find the focal length of the mirror.

15.8 When an object that is initially 0.60 m from a concave mirror moves 0.10 m closer to it, the separation between the object and its image becomes five-halves larger. Determine the focal length of the mirror.

15.9 (a) If x_1 and x_2 are the distances of the object and its image, measured from the focus of a spherical mirror, show that Eq. (15.3) gives

$$x_1 x_2 = f^2.$$

This is called *Newton's equation.* (b) Can you then conclude that the object and its image are always on the same side of the focus? (*Hint:* Note that $x_1 = FP = OP - OF$, and similarly for x_2.)

15.10 A transparent substance is limited by a concave spherical surface with a radius equal to 0.60 m. Its index of refraction is 1.5. (a) Determine the focal lengths. (b) Determine the position and the magnification of the image of an object placed at a distance from the surface equal to (i) 2.40 m, (ii) 1.60 m, and (iii) 0.60 m.

15.11 Repeat Problem 15.10 for a surface that is convex.

15.12 A cylindrical glass rod whose index of refraction is 1.5 is terminated by two convex spherical surfaces, with radii of curvature of 0.10 and 0.20 m, respectively (Fig. 15-34). The length of the rod between vertexes is 0.50 m. An arrow 1 mm long lies in front of the first spherical surface, at right angles to the axis of the cylinder and 0.25 m from the vertex. Compute (a) the position and (b) the length of the image of the arrow formed by the first surface, and (c) the position and (d) the length of the image of the arrow formed by both surfaces. Specify whether the images are real or virtual.

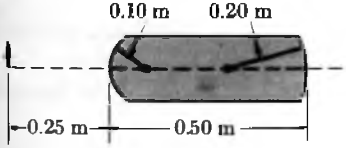


Figure 15-34

15.13 (a) Determine the focal points of the system described in Problem 15.12. (b) Solve this problem graphically.

15.14 A transparent rod 0.40 m long is cut flat at one end and rounded to a hemispherical surface of 0.12 m radius at the other end. An object is placed on the axis of the rod, 0.10 m from the hemispherical end. (a) What is the position of the final image? (b) What is its magnification? Assume the refractive index is 1.50.

15.15 The rod in Problem 15.14 is shortened until there is a distance of 0.10 m between its vertexes; the curvatures of its ends remain the same. (a) What is the object distance for the second surface? (b) Is the object real or virtual? (c) Is it upright or inverted, with respect to the original object? (d) What is the height of the final image?

15.16 A glass rod of refractive index 1.50 is ground and polished at both ends to convex hemispherical surfaces of 0.05 m radius. When an object is placed on the axis of the rod and 0.20 m from one end, the final real image is formed 0.40 m from the opposite end. What is the length of the rod?

15.17 (a) Sketch the various possible thin lenses obtainable by combining two surfaces whose radii of curvature, in absolute magnitude, are 0.10 m and 0.20 m. (b) Which are converging and which are diverging? (c) Find the focal length of each lens if each is made of glass with refractive index 1.50.

15.18 A double convex lens has an index of refraction of 1.5 and its radii are 0.20 m and 0.30 m. (a) Find the focal length. Determine (b) the position of the image and (c) the magnification of an object located at a distance from the lens equal to (i) 0.80 m, (ii) 0.48 m, (iii) 0.40 m, (iv) 0.24 m, and (v) 0.20 m. (d) Consider also the case of a virtual object located 0.20 m behind the lens.

15.19 A double concave lens has an index of refraction of 1.5 and its radii are 0.20 m and 0.30 m. (a) Find the focal length. Determine (b) the position of the image and (c) the magnification of an object located 0.20 m from the lens. Consider also a virtual object at a distance of (d) 0.40 m, and (e) 0.20 m.

15.20 A lens casts an image of an object on a screen placed 0.12 m from the lens. When the lens is moved 2×10^{-2} m farther from the object, the screen must be moved 2×10^{-2} m closer to the object in order to bring it into focus. What is the focal length of the lens?

15.21 An object is placed 0.18 m from a screen. (a) At what points between object and screen may a lens whose focal length is 4×10^{-2} m be placed to obtain an image on the screen? (b) What is the magnification of the image for these positions of the lens?

15.22 A convergent lens has a focal length equal to 0.40 m. Determine the position of an object and the nature of the image if the magnification is (a) -0.6, (b) -1.5, (c) -1, (d) 3, and (e) 0.8. The minus signs imply inversion of the image.

15.23 A convergent lens has a focal length of 0.60 m. Find the position of the object that will produce an image (a) real and three times larger, (b) real and one-third as large, (c) virtual and three times larger.

15.24 Determine the focal length and the nature of a lens that produces an image of an object located 1.20 m from the lens that is (a) real and 0.80 m from the lens, (b) virtual and 3.20 m from the lens, (c) virtual and 0.60 m from the lens, (d) real and twice as large, (e) virtual and twice as large, (f) real and one-third as large, and (g) virtual and one-third as large.

15.25 Show that for a thin spherical lens

$$x_1 x_2 = -f^2,$$

where x_1 is the distance of the object from the first focus and x_2 is the distance of the image from the second focus. [See Problem 15.9.]

15.26 Rays from a lens converge at a point image P, as in Fig. 15-35. What thickness t of glass with a refractive index of 1.50 must be interposed, as shown in the figure, in order that the image be formed at P'?

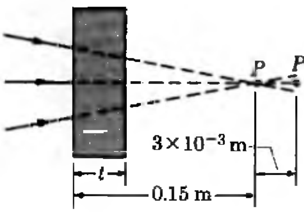


Figure 15-35

15.27 A lens system is composed of two convergent lenses, with focal lengths of 0.3 m and 0.60 m. Find the image focus of the system for the cases in which the separation is (a) 0.20 m, (b) 0.50 m, (c) 0.90 m, and (d) 1.20 m.

15.28 Repeat Problem 15.27 assuming the first lens is divergent.

15.29 The ocular (or eyepiece) of an optical instrument is composed of two identical convergent lenses, with focal lengths of 5×10^{-2} m each, that are separated 2.5×10^{-2} m. Find the position of the foci of the system as measured from the lens closer to the object.

15.30 The objective of a microscope has a focal length of 4 mm. The image formed by this objective is 180 mm from its second focal point. The eyepiece has a focal length of 31.25 mm. (a) What is the magnification of the microscope? The unaided eye can distinguish two points as being separate if they are about 10^{-4} m apart. (b) What is the minimum separation that can be perceived with the aid of this microscope?

15.31 The diameter of the moon is 3.5×10^6 m and its distance from the earth is 3.8×10^8 m. Find the angular diameter of the image of the moon formed by a telescope if the focal length of the objective is 4.0 m and that of the eyepiece 0.10 m.

15.32 A prism has an index of refraction of 1.5 and an angle of 60°. (a) Determine the deviation of a ray incident at an angle of 40°. Find (b) the minimum deviation and (c) the corresponding angle of incidence.

15.33 The minimum deviation of a prism is 30°. The angle of the prism is 50°. Find (a) its index of refraction and (b) the angle of incidence for minimum deviation.

15.34 From the data of Table 15-3, obtain the coefficients A_0 and B_0 appearing in Cauchy's formula for the index of refraction in the case of crown glass.

15.35 Using the result of the preceding problem, determine the angular separation corresponding to the C- and F-Fraunhofer lines in a crown-glass prism having an angle of 50° in the case of a ray whose angle of incidence is 30°.

15.36 A lens system is composed of two lenses in contact, one plano-concave and made of flint glass and another double convex and made of crown glass. The radius of the common face is 0.20 m and the radius of the other face of the crown lens is 0.12 m. Find (a) the focal length of the system and (b) the chromatic aberration.

CHALLENGING PROBLEMS

15.37 A lighted object is placed 0.15 m from a thin converging lens that has a focal length of 0.10 m. Find the position of the image and trace two light rays from the object through the lens. [AP-B; 1970]

15.38 An object 1 cm high is placed 4 cm away from a converging lens having a focal length of 3 cm. (a) Sketch a principal ray diagram for this situation. (b) Find the location of the image by a numerical calculation. (c) Determine the size of the image. [AP-B; 1974]

15.39 An object 1 cm high is placed 6 cm to the left of a converging lens whose focal length is 8 cm, as shown in Fig. 15-36. (a) Calculate the position of the image. Is it to the left or right of the lens? Is it real or virtual? (b) Calculate the size of the image. Is it upright or inverted? (c) Locate the image by ray tracing. (d) What simple optical instrument uses this sort of object-image relationship? [AP-B; 1976]

15.40 An object 6 cm high is placed 30 cm from a concave mirror of focal length 10 cm, as shown in Fig. 15-37. (a) On a diagram drawn to scale, locate the image by tracing two rays that begin at point P and pass through the focal point F. Is the image real or virtual? Is it located to the left or to the right of the mirror? (b) Calculate the position of the image. (c) Calculate the size of the image. (d) Indicate on a diagram how the ray from point P to point Q is reflected, if aberrations are negligible. [AP-B; 1978]

15.41 An object O is placed 0.18 m from the center of a converging lens of focal length 6×10^{-2} m as illustrated in Fig. 15-38a. (a)

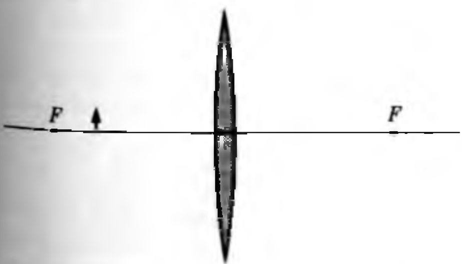


Figure 15-36

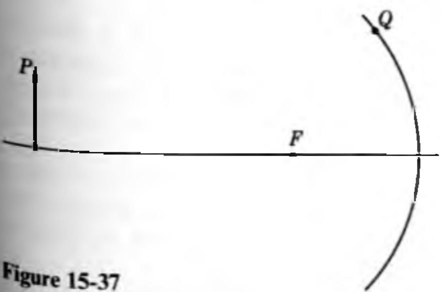


Figure 15-37

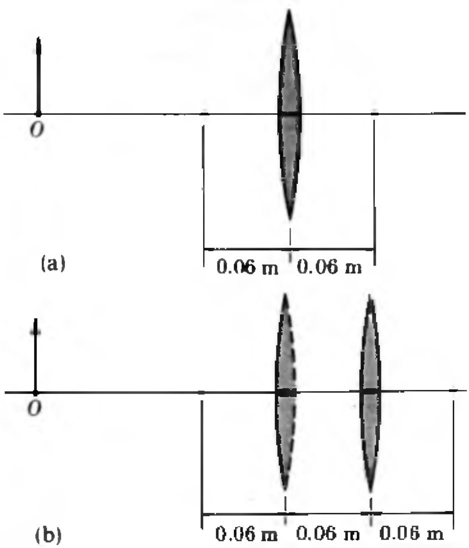


Figure 15-38

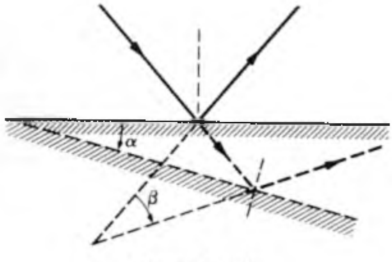


Figure 15-39

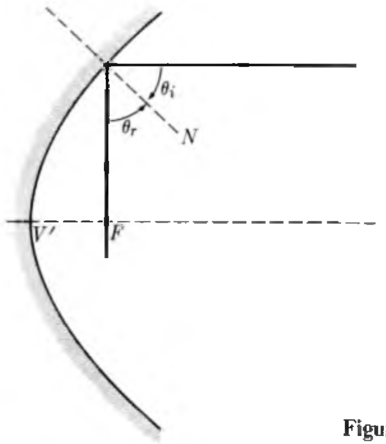


Figure 15-40

Sketch a ray diagram to locate the image. (b) Is the image real or virtual? Explain your choice. (c) Using the lens equation, compute the distance of the image from the lens. A second coverging lens, also of focal length 6×10^{-2} m, is placed 6×10^{-2} m to the right of the original lens as illustrated in Fig. 15-38b. (d) Sketch a ray diagram to locate the final image that now will be formed. Clearly indicate the final image. [AP-B; 1981]

15.42 Show that when a plane mirror is rotated an angle α, the reflected ray rotates through an angle twice as large; that is, $\beta=2\alpha$ in Fig. 15-39.

15.43 Show that if a plane mirror is displaced parallel to itself a distance x along the normal, the image moves a distance $2x$.

15.44 The *spherical aberration* of a (spherical) mirror is defined as the difference between the focal length f for a ray close to the mirror's axis and the focal length f' for a ray close to the edge. Show that

$$f-f'\approx\frac{H^2}{2r},$$

where H is the radius of the mirror's base.

15.45 A concave mirror has a radius of 0.10 m. The base of the mirror has a radius of 0.08 m. Find the spherical aberration of the mirror and compare with its focal length.

15.46 An object moves toward a spherical mirror with a constant velocity v. (a) Find the velocity of the image as a function of the distance p. (b) Plot the velocity of the image versus p. (c) Repeat the problem for a spherical lens.

15.47 (a) Show that if u_1 and u_2 are the distances of an object and its image from the center of a spherical mirror, the relation

$$\frac{1}{u_1}+\frac{1}{u_2}=-\frac{2}{r}$$

holds. (b) Show that in this case the magnification is given by

$$M=\frac{u_2}{u_1}.$$

15.48 Given the focal length f of a spherical mirror and the magnification M, show that the positions of the object and the image are

$$p=\frac{f(M-1)}{M}$$

and

$$q=-f(M-1).$$

15.49 Prove that all rays parallel to the axis of a parabolic mirror (Fig. 15-40) pass through the focus after the reflection, irrespective of their distances from the axis.

15.50 A glass sphere 2×10^{-2} m in diameter contains a small air bubble at a distance of 5×10^{-3} m from the center. Find (a) the position and (b) the magnification of the image of the bubble, as seen by a person looking from one or the other of the two opposite directions along the line connecting the center of the sphere with the bubble. The index of refraction of the glass is 1.50.

15.51 A transparent sphere with a refraction index of n relative to air has a radius r. An object is placed at a distance $4r$ from the center of the sphere. (a) Find the position of its final image. (b) Plot the path of a ray through the sphere.

15.52 Both ends of a glass rod 0.10 m in diameter, with a refraction index of 1.50, are ground and polished to convex hemispherical surfaces of radius 5×10^{-2} m at the right end and radius 0.10 m at the left end. The length of the rod between vertexes is 0.60 m. An arrow 1 mm long, at right angles to the axis and 0.20 m to the right of the first vertex, constitutes the object for the first surface. (a) What constitutes the object for the second surface? (b) What is the object distance for the second surface? (c) Is this object real or virtual? (d) What is the position of the image formed by the second surface? (e) What is the height of the final image?

15.53 A solid glass sphere of radius R and refraction index 1.50 is silvered over one hemisphere, as in Fig. 15-41. A small object is located on the line passing through the center of the sphere and the pole of the hemisphere, at a distance of $2R$ from the pole of the unsilvered hemisphere. Find the position of the final image after all refractions and reflections have taken place.

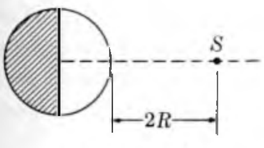


Figure 15-41

15.54 Consider a glass sphere of radius R and index of refraction n, cut by a plane through a point S at a distance x from the center O and perpendicular to OS (Fig. 15-42). Show that if $x = R/n$, all rays entering the glass sphere from a point source at S will emerge from the sphere along lines diverging from a point S', collinear with O and S at a distance $x' = nR$ from O. (*Hint:* Show that the refracted ray, when extended backward, passes through S' for all values of ϕ and the given values of x and x'.)

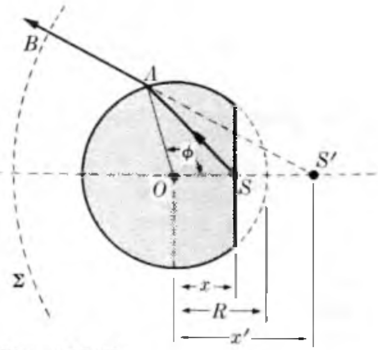


Figure 15-42

15.55 Show that for a spherical refracting surface separating two substances of indexes of refraction n_1 and n_2, the relation

$$x_1 x_2 = f_0 f_i$$

holds, where x_1 is the distance of the object from the first focus and x_2 the distance of the image from the second focus.

15.56 Show that for refraction at a spherical surface the following relation holds:

$$n_1 y_1 \sin \alpha_1 = n_2 y_2 \sin \alpha_2$$

where α_1 and α_2 are as shown in Fig. 15-12 and y_1 and y_2 are the sizes of the object and its image. Therefore, for a ray passing through several refracting surfaces, all having their centers on the same line, the relation

$$ny \sin \alpha = \text{const}$$

holds. This relation is called the *Helmholtz law*. (*Hint:* Referring to Fig. 15-12, apply the law of sines combined with Snell's law and the relation obtained from the similarity of triangle Cab and CAB in Fig. 15-15.)

15.57 A thin lens with a refraction index of n_2 is surrounded by two media, with refraction indexes of n_1 and n_3, respectively. Show that the equation relating the position of the object and of the image is

$$\frac{n_1}{p} - \frac{n_3}{q} = \frac{(n_1 - n_2)}{r_1} + \frac{(n_2 - n_3)}{r_2}.$$

15.58 A tank filled with water has an opening in one wall covered by a double convex lens with a refraction index of 1.5 and radius equal to 0.30 m. (a) Find the focal length of a ray that approaches the lens parallel to the axis from inside the tank and from outside the tank. (b) Determine the position of the image of a light source located inside the tank at (i) 0.30 m, and (ii) 0.45 m from the lens. The index of refraction for water is 1.33.

15.59 An equiconvex thin lens made of glass with a refraction index of 1.50 has a focal length in air of 0.30 m. The lens is sealed into an opening in one end of a tank filled with water (index = 1.33). At the end of the tank opposite the lens is a plane mirror, 0.80 m distant from the lens. Find the position of the image formed by the lens-water-tank for an object placed on the lens axis 0.90 m from the lens and outside the tank.

15.60 (a) Using expression (15.17) for computing the refraction at each surface of a spherical lens, show that the focal length is given by

$$\frac{1}{f}=(n-1)\left(\frac{1}{r_2}-\frac{1}{r_1}\right)+\frac{h^2}{2}\frac{n-1}{n^2}\times\left[\left(\frac{1}{r_2}+\frac{n+1}{f}\right)\left(\frac{1}{r_2}+\frac{1}{f}\right)^2-\frac{1}{r_1^3}\right],$$

where f is given by Eq. (15.22). (b) Estimate from this result the value of the *spherical aberration* of the lens, defined as the difference in focal length for a ray close to the axis and a ray close to the edge of the lens.

15.61 A double-convex lens has a refraction index of 1.5. The two surfaces have the same radius, equal to 0.10 m. The lens has a radius of 0.08 m. Find (a) the focal length and (b) the spherical aberration.

15.62 Plot q as a function of p for (a) (i) a spherical mirror satisfying Eq. (15.3), and (ii) a spherical lens satisfying Eq. (15.23). (b) Verify that in each case an equilateral hyperbola results. (c) Also plot the magnification as a function of p in each case.

15.63 (a) Show that for refraction due to a prism,

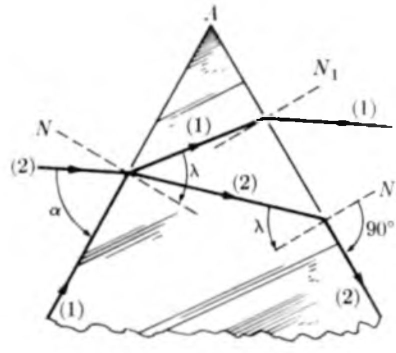


Figure 15-43

$$\sin\tfrac{1}{2}(\delta+A)=n\sin\tfrac{1}{2}A\,\frac{\cos\frac{1}{2}(r-r')}{\cos\frac{1}{2}(i-i')}.$$

(b) Show also that $\cos\frac{1}{2}(r-r')$ is never smaller than $\cos\frac{1}{2}(i-i')$. (c) Conclude then that the condition for minimum deviation is $i=i'$.

15.64 Show that if the angle of a prism is very small and the incident rays fall almost perpendicular to one of the faces, the deviation is $\delta=(n-1)A$.

15.65 If a ray reaches the second surface of a prism at an angle larger than the critical angle, total reflection occurs and the ray is reflected back instead of passing out of the prism. This principle is used in many optical instruments. (a) Show that if $n>1$, the condition that at least one ray emerge is that $A\leqslant 2\lambda$, where λ is the critical angle. (b) Discuss the range of values of the angle of incidence i if the ray is to emerge on the other side. This range is given by the angle α shown in Fig. 15-43. (c) Prove that the range is given by

$$\cos\alpha=n\sin(A-\lambda).$$

(d) Discuss the variation of α with A.

15.66 Apply the discussion of the preceding problem to the case of a prism having a refracting angle of 45° and a refraction index of 1.5. (a) Obtain the value of α. (b) Discuss the path of a ray that falls perpendicular to one of the faces. (c) Consider also the case in which the prism's angle is 35°.

CHAPTER SIXTEEN

INTERFERENCE

16.1 Introduction

A very important characteristic of wave motion is the phenomenon of interference. This occurs when two or more wave motions coincide in space and time. The theory of superposition of two simple harmonic motions developed earlier in Volume I can be applied directly to our present problem for the case of harmonic or monochromatic waves. One place in which interference occurs is the spatial region in which reflected and incident waves coincide. In fact, this is one of the most common methods of producing interference. Another important case of interference is found in a wave motion confined to a limited region of space; for example, a string with its two ends fixed, or a liquid in a channel, or an electromagnetic wave in a metallic cavity. The interference then gives rise to *standing waves.*

In order to apply the formulas previously developed, we shall write for a harmonic wave moving in the $+X$-direction

$$\xi = \xi_0 \sin(\omega t - kx), \tag{16.1}$$

and for one moving in the $-X$-direction

$$\xi = \xi_0 \sin(\omega t + kx), \tag{16.2}$$

instead of Eqs. (10.5) and (10.9). This involves only a change of sign; and as long as we are consistent, it is a correct procedure as was indicated at the end of Section 10.2 in Eq. (10.10) and as used in Example 13.2.

As mentioned in previous chapters, the theory developed here is applicable to any kind of wave motion; but in general our examples and applications will refer to electromagnetic waves.

16.2 Interference of Waves Produced by Two Synchronous Sources

Consider two point sources S_1 and S_2 (Fig. 16-1) that oscillate in phase with the same angular frequency ω and have amplitudes ξ_{01} and ξ_{02}. Their respective spherical waves are

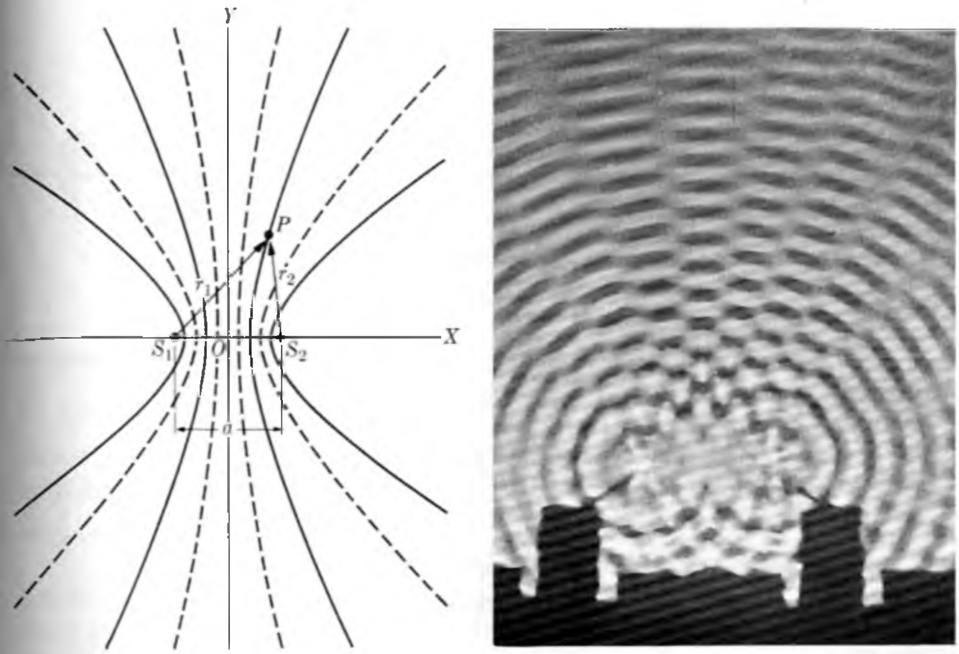
$$\xi_1 = \xi_{01} \sin(\omega t - kr_1) \tag{16.3}$$

and

$$\xi_2 = \xi_{02} \sin(\omega t - kr_2) \tag{16.4}$$

where r_1 and r_2 are the distances from any point to S_1 and S_2, respectively. Note that although the two sources are identical, they do not produce the same amplitude at P if r_1 and r_2 are different because the amplitude of a spherical wave decreases according to a $1/r$ law.

Suppose that ξ is a scalar property, such as a pressure disturbance. If ξ corresponds to a vector quantity, we assume that ξ_1 and ξ_2 are in the same direction so



(a) (b)

Fig. 16-1. (a) Nodal and ventral lines resulting from interference of waves produced by two identical sources. (b) Actual interference pattern of waves on the surface of water. (Photograph courtesy of Educational Services, Inc.)

that the combination of the two waves can be treated in a scalar manner. When we compare Eqs. (16.3) and (16.4) with the expression for simple harmonic motion $\xi = A \sin(\omega t + \alpha)$ where α is the initial phase, we recognize that the quantities kr_1 and kr_2 play the same role as initial phases. Then the phase difference between the two wave motions at any point P (if we remember that $k = 2\pi/\lambda$) is

$$\delta = kr_1 - kr_2 = \frac{2\pi}{\lambda}(r_1 - r_2). \tag{16.5}$$

When we use the technique of rotating vectors (which is explained in Chapter 12 of Volume I), the two interfering wave motions can be represented by rotating vectors of length ξ_{01} and ξ_{02}, respectively, which make angles $\alpha_1 = kr_1$ and $\alpha_2 = kr_2$ with the X-axis (Fig. 16-2). The amplitude ξ_0 and phase α of the resulting wave motion are given by their vector resultant. Therefore we may express the amplitude of the resulting disturbance at P by

$$\xi_0 = \sqrt{\xi_{01}^2 + \xi_{02}^2 + 2\xi_{01}\xi_{02}\cos\delta}. \tag{16.6}$$

From Eq. (16.6), we see that ξ falls between the values $\xi_{01} + \xi_{02}$ when $\cos\delta = 1$ or $\delta = 2n\pi$ and $\xi_{01} - \xi_{02}$ when $\cos\delta = -1$ or $\delta = (2n+1)\pi$ where n is either a positive or a

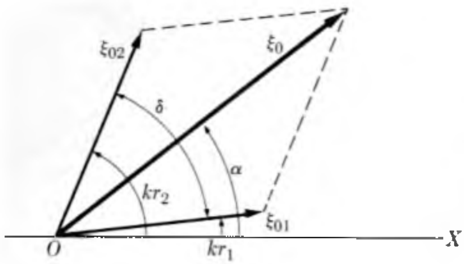


Fig. 16-2. Resultant amplitude of two interfering waves. The X-axis has been taken as the reference line.

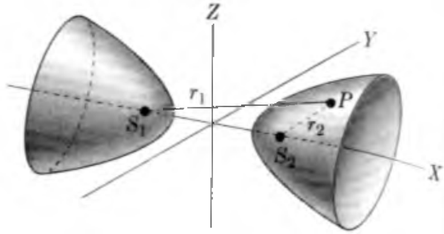


Fig. 16-3. Surfaces of constant phase difference for spherical waves produced by two coherent point sources S_1 and S_2.

negative integer. In the first case we have maximum reinforcement of the two wave motions, or *constructive interference*, and in the second case maximum attenuation, or *destructive interference*. That is,

$$\delta = \begin{cases} 2n\pi & \text{constructive interference,} \\ (2n+1)\pi & \text{destructive interference.} \end{cases}$$

Using Eq. (16.5), we may then write

$$\frac{2\pi}{\lambda}(r_1 - r_2) = \begin{cases} 2n\pi & \text{constructive interference,} \\ (2n+1)\pi & \text{destructive interference,} \end{cases} \tag{16.7}$$

or

$$r_1 - r_2 = \begin{cases} n\lambda & \text{constructive interference,} \\ (2n+1)\dfrac{\lambda}{2} & \text{destructive interference.} \end{cases} \tag{16.8}$$

But $r_1 - r_2 = \text{constant}$ defines a hyperbola whose foci are S_1 and S_2; or since the problem is actually in space, this equation defines hyperbolic surfaces of revolution, such as those in Fig. 16-3. Therefore we conclude from Eq. (16.8) that at the hyperbolic surfaces whose equations are $r_1 - r_2 = \pm\lambda, \pm 2\lambda, \pm 3\lambda, \ldots$, the two wave motions interfere constructively. These surfaces are called *ventral* or *antinodal surfaces*. At the hyperbolic surfaces whose equations are $r_1 - r_2 = \pm\frac{1}{2}\lambda, \pm\frac{3}{2}\lambda, \ldots$, the two wave motions interfere destructively. These surfaces are called *nodal surfaces*. The overall pattern is thus a succession of alternate ventral and nodal surfaces. The intersections of these surfaces with a plane passing through the X-axis are the hyperbolas illustrated in Fig. 16-1.

The situation described is such that at each point of space the resulting wave motion has a characteristic amplitude, given by Eq. (16.6), so that

$$\xi = \xi_0 \sin(\omega t - \alpha)$$

where α is an initial phase angle as indicated in Fig. 16-2. Therefore the result of the

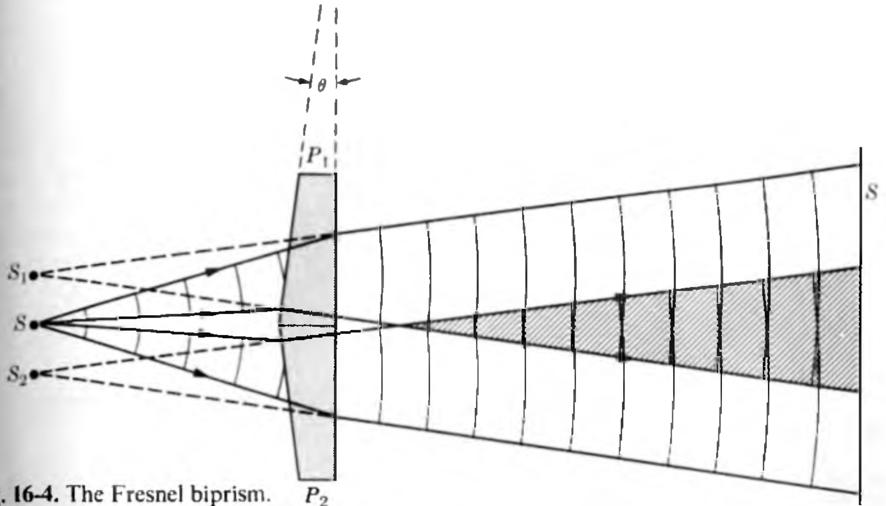


Fig. 16-4. The Fresnel biprism. P_2

interference does not have the appearance of a progressive wave motion; rather, there is a *stationary* situation at each point of space where the oscillatory motion has a fixed amplitude. The reason for this stationary condition is that the two sources oscillate with the same frequency, maintain a constant phase difference, and hence are said to be *coherent*. If the sources are not of the same frequency, or if their phase difference changes erratically with time, no stationary interference pattern is observed, and the sources are said to be *incoherent*. This is the normal situation for two light sources composed of the same kind of atoms, emitting light of the same frequency. Since there are many atoms involved in each source and they do not oscillate in phase, no definite interference pattern is observed.

To circumvent this difficulty and produce two coherent beams of light, several devices have been designed. A common device is *Fresnel's biprism*, illustrated in Fig. 16-4 and named after its developer, the French physicist Augustin Fresnel (1788–1827). It is composed of two prisms, P_1 and P_2 set with their bases joined; each prism has a very small refracting angle θ. Light coming from the source S is refracted in each prism and separated into two beams that apparently proceed from two coherent sources, S_1 and S_2. These are the images of S produced by each prism. Coherency is assured in this case because the two beams proceed from the same source. The beams interfere in the shaded region. For large phase differences, the coherence is destroyed because the interfering beams are produced by the source at two relatively widely separated times so that microscopically speaking, the source is not the same.

An even simpler device is the one used by the English scientist Thomas Young (1773–1829), who in his early experiments on light interference proved that light was a wave phenomenon. His arrangement (Fig. 16-5) consists of two small, closely spaced holes or slits, S_1 and S_2, in an opaque surface with a light source S placed

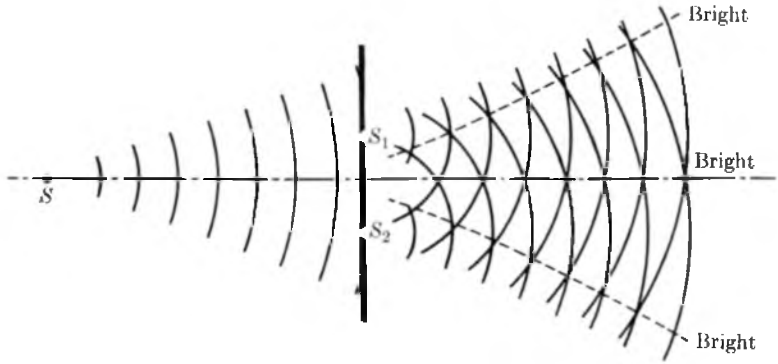


Fig. 16-5. The interference of two coherent sources. Young's double-slit experiment.

behind it. According to Huygens' principle, S_1 and S_2 behave as secondary and coherent sources whose waves interfere on the far side of the surface.

In the case of light, an interference pattern is observed on a screen placed parallel to the two sources S_1 and S_2 as indicated in Fig. 16-6a. A series of alternate bright and dark fringes appear on the screen as shown in Fig. 16-6b, and are caused by the intersection of the screen with the hyperbolic ventral and nodal surfaces. For other regions of the electromagnetic spectrum, different kinds of detectors are used to observe the interference pattern.

If the separation a of the sources S_1 and S_2 is small compared with the distance D, we may neglect the small difference between r_1 and r_2 and assume the amplitudes ξ_{01} and ξ_{02} are practically equal. We may then rewrite Eq. (16.6) as

$$\xi_0 = \xi_{01}\sqrt{2(1+\cos\delta)} = 2\xi_{01}\cos\tfrac{1}{2}\delta.$$

Now from the geometry of Fig. 16-6, considering that θ is a small angle so that $\sin\theta \approx \tan\theta = x/D$, we have $r_1 - r_2 = S_1B = a\sin\theta = ax/D$; and rewriting Eq. (16.7), we get

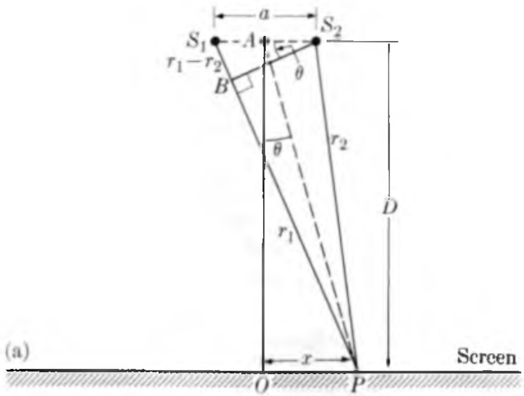
$$\delta = \frac{2\pi}{\lambda}(r_1 - r_2) = \frac{2\pi}{\lambda}a\sin\theta = \frac{2\pi ax}{D\lambda}. \tag{16.9}$$

The intensity of the resulting motion on the points of the screen is proportional to ξ_0^2. Therefore

$$I = I_0\cos^2(\tfrac{1}{2}\delta) = I_0\cos^2\left(\frac{\pi a\sin\theta}{\lambda}\right) = I_0\cos^2\left(\frac{\pi ax}{D\lambda}\right) \tag{16.10}$$

where I_0 is the intensity for $\theta = 0$. This $\cos^2$ intensity distribution is illustrated in Fig. 16-7. The points of maximum intensity correspond to

$$\frac{\pi a\sin\theta}{\lambda} = n\pi \qquad \text{or} \qquad \frac{\pi ax}{D\lambda} = n\pi.$$



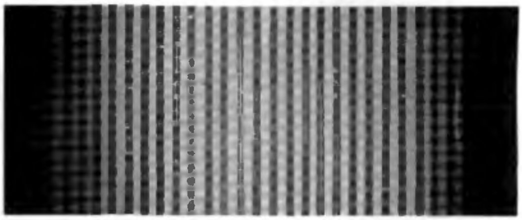
(b)


Fig. 16-6. (a) Schematic diagram for determining the intensity of the resultant wave motion on a screen caused by the interference of two coherent sources. (b) Photograph of the interference fringes produced on a screen by a pair of slits illuminated by a point monochromatic light source. Note that there is a fading of the fringes near the edges, because of the loss of coherence.

These two results for maximum intensity may be written as

$$\sin\theta = n\frac{\lambda}{a} \qquad \text{or} \qquad x = n\frac{D\lambda}{a} \tag{16.11}$$

where n can be either a positive or negative integer, and θ and x are shown in Fig. 16-6. The separation between two successive bright fringes on the screen is $\Delta x = (D/a)\lambda$. Therefore measuring Δx, D, and a serves to determine the wavelength λ. This method is, in fact, one of the standard methods of measuring wavelengths. Note that if the light source is composed of more than one wavelength, each wavelength will separately produce its own interference pattern on the screen.

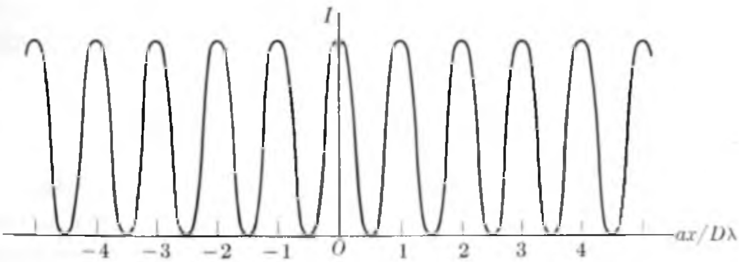


Fig. 16-7. Intensity distribution in the interference pattern produced by two coherent sources.

Example 16.1. Determination of the separation between two successive bright or dark fringes in a two slit experiment.

▼ Consider an experiment, similar to Young's, in which two slits are separated a distance of 0.8 mm. The slits are illuminated with monochromatic light of wavelength 5.9×10^{-7} m, and the interference pattern is observed on a screen at a distance of 0.50 m from the slits. The quantities appearing in Eq. (16.11) are, in this case, $a = 0.8$ mm $= 8 \times 10^{-4}$ m, $D = 5 \times 10^{-1}$ m, and $\lambda = 5.9 \times 10^{-7}$ m. Therefore the positions of the bright fringes are $x = n(D\lambda/a) = 3.7 \times 10^{-4} n$ m $= 0.37n$ mm. In general the fringes have to be observed with a magnifying glass. The separation between successive bright fringes is 0.37 mm and is the same as the separation between two dark fringes. ▲

Example 16.2. The interference pattern of two incoherent sources of the same frequency.

▼ Incoherence in this case is due to a variable phase difference. Therefore instead of Eqs. (16.3) and (16.4), we write

$$\xi_1 = \xi_{01} \sin(\omega t - kr_1 - \phi) \qquad \text{and} \qquad \xi_2 = \xi_{02} \sin(\omega t - kr_2)$$

where ϕ is the additional phase difference that varies with time in a random manner. Then the phase difference is $\delta = 2\pi(r_1 - r_2)/\lambda + \phi$, and the resulting amplitude at any point in the interference pattern is

$$\xi_0^2 = \xi_{01}^2 + \xi_{02}^2 + 2\xi_{01}\xi_{02} \cos\left[\frac{2\pi}{\lambda}(r_1 - r_2) + \phi\right].$$

But ξ_0 is not constant in time because of the changes in ϕ; therefore we must find $(\xi_0^2)_{\text{ave}}$ instead. Because of the random variations in ϕ, we have that

$$\left\{\cos\left[\frac{2\pi}{\lambda}(r_1 - r_2) + \phi\right]\right\}_{\text{ave}} = 0.$$

Therefore

$$(\xi_0^2)_{\text{ave}} = (\xi_{01}^2)_{\text{ave}} + (\xi_{02}^2)_{\text{ave}};$$

and since the intensity is proportional to the square of the amplitude,

$$I_{\text{ave}} = I_1 + I_2.$$

Therefore the resultant average intensity is the sum of the individual intensities, and no fluctuations of intensity with position are observed. For this reason we do not observe interference fringes from two electric bulbs, for example, since the phase differences of their respective radiating atoms are distributed at random even if both emit monochromatic light of the same wavelength. ▲

16.3 Interference of Several Synchronous Sources

Consider now the case of several synchronous and identical monochromatic sources arranged linearly as illustrated in Fig. 16-8. To simplify our discussion, assume that we observe the resulting wave motion at a distance very large compared with the

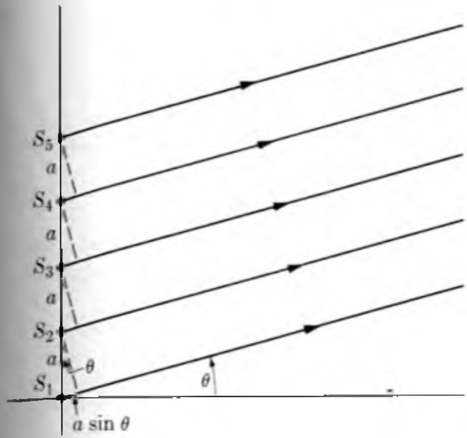


Fig. 16-8. Linear series of equally spaced coherent sources.

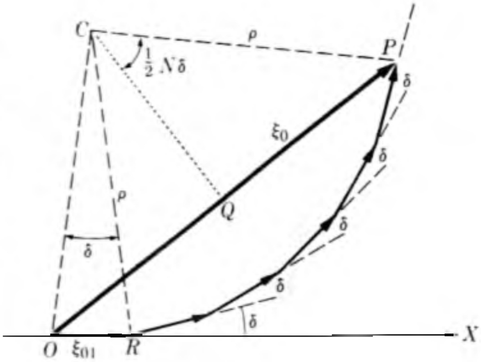


Fig. 16-9. Resultant amplitude at an arbitrary point due to the interference of waves generated by equally spaced linear coherent sources.

separation of the sources so that effectively the interfering rays are parallel. From the figure it may be seen that between successive rays there is a constant phase difference given by

$$\delta = \frac{2\pi}{\lambda} a \sin \theta \tag{16.12}$$

where a is the distance between the sources, λ is the wavelength, and θ is the angle a particular set of rays makes with the normal to the plane of the sources. To obtain the resulting amplitude for the direction of observation, given by the angle θ, we must evaluate the *vector sum* of the corresponding rotating vectors for each source. If all the sources are alike, their rotating vectors all have the same amplitude, ξ_{01}; and successive vectors are deviated from the previous vector by the same angle δ as indicated in Fig. 16-9. Designating the number of sources by N, we then have an $N+1$-sided polygon having a center C and a "radius" ρ where the angle OCP is $N\delta$. The amplitude of the resultant vector is given by the line OP that closes the polygon. From the figure we see that

$$\xi_0 = OP = 2QP = 2\rho \sin \tfrac{1}{2}N\delta.$$

Also from triangle COR we have

$$\xi_{01} = 2\rho \sin \tfrac{1}{2}\delta.$$

Dividing the two relations to eliminate ρ, we obtain

$$\xi_0 = \xi_{01} \frac{\sin \frac{1}{2}N\delta}{\sin \frac{1}{2}\delta}. \tag{16.13}$$

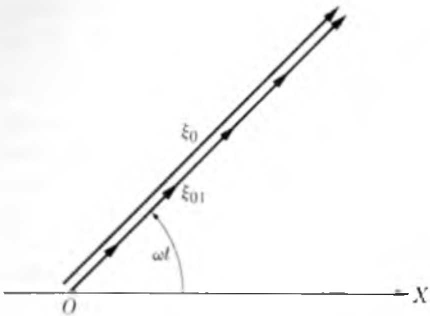


Fig. 16-10. Resultant amplitude of interference of equally spaced linear sources at a point of zero phase difference.

For $N=2$, we get $\xi_0=2\xi_{01}\cos\frac{1}{2}\delta$, in agreement with the previous result for two equal sources derived in Section 16.2.

The intensity of the resulting wave, since it is proportional to ξ_0^2, is then

$$I=I_0\left(\frac{\sin\frac{1}{2}N\delta}{\sin\frac{1}{2}\delta}\right)^2=I_0\left[\frac{\sin(N\pi a\sin\theta/\lambda)}{\sin(\pi a\sin\theta/\lambda)}\right]^2 \tag{16.14}$$

where I_0 is the intensity of each source, proportional to ξ_{01}^2. Expression (16.14) has very pronounced maxima, equal to N^2I_0 when $\delta=2n\pi$ because $\sin N\alpha/\sin\alpha=\pm N$ for $\alpha=n\pi$, and in our case $\alpha=\frac{1}{2}\delta$. That is, there is constructive interference when

$$a\sin\theta=n\lambda \tag{16.15}$$

where n is any integer, including zero. The value of I is then $I=N^2I_0$ whenever Eq. (16.15) is satisfied. This result is understandable because when $\delta=2n\pi$, all source vectors ξ_{01} are parallel as indicated in Fig. 16-10; and the resulting amplitude is $\xi_0=N\xi_{01}$ as given by Eq. (16.13). Expression (16.15) agrees with Eq. (16.11) derived for two sources under similar assumptions. The intensity of the resulting wave is zero whenever $\frac{1}{2}N\delta=n'\pi$. That is, we have destructive interference when

$$a\sin\theta=\frac{n'\lambda}{N} \tag{16.16}$$

where n' is any integer excluding $0, N, 2N, \ldots$, because then Eq. (16.16) would transform into Eq. (16.15) in which we have shown that all sources are in phase. Between two minima there must always be a maximum. Therefore we conclude that there are also $N-2$ additional maxima between the principal maxima given by Eq. (16.15). Their amplitudes are, however, relatively much smaller, especially if N is large. The principal maxima, however, do correspond to the directions for which the waves emitted by adjacent sources are in phase.

The graph of I/I_0 in terms of δ is shown in Fig. 16-11 for $N=2, 4, 8$, and very large. We see that as N increases, the system becomes highly directional, since the resulting wave motion is important only for narrow bands of values of δ, or for narrow bands of values of the angle θ.

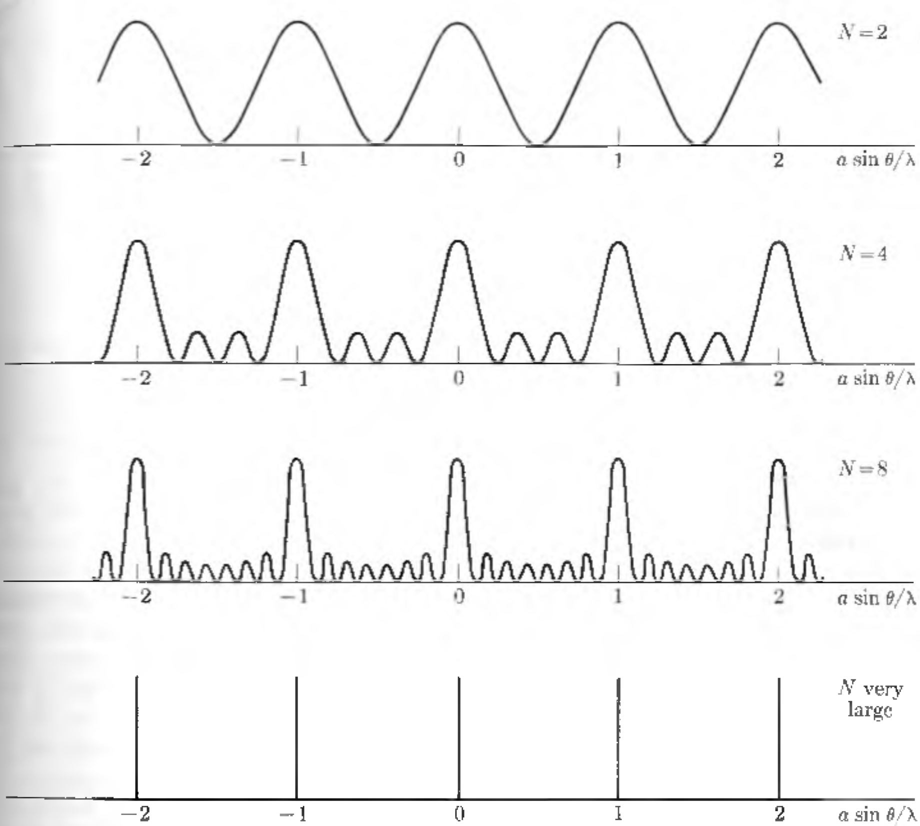


Fig. 16-11. Intensity of the interference pattern for 2, 4, 8, and very many sources. The source spacing a is kept constant.

These results are widely used in broadcasting or receiving stations when a directional effect is desired. In this case several antennas are arranged in such a form that the intensity of the radiation emitted (or received) is maximum only for certain directions, given by Eq. (16.15). For example, given four antennas in a straight line and separated by $a=\lambda/2$, Eq. (16.15) gives $\sin\theta=2n$. Then only $n=0$ is possible for the principal maxima, giving $\theta=0$ and π. For the zeros, or nodal planes, Eq. (16.16) gives $\sin\theta=\frac{1}{2}n'$, allowing for $n'=\pm 1$ and ± 2 or $\theta=\pm\pi/3$ and $\pm\pi/2$. The situation is illustrated in the polar diagram of Fig. 16-12, in which the intensity is plotted in terms of the angle. The antenna arrangement of Fig. 16-12 then transmits and receives preferentially in a direction perpendicular to the line joining the sources, and is therefore called a *broadside array*. The same directional effect is used in radio telescopes. Several parabolic antennas are placed at equal distances along a straight line and with their axes parallel. For a given spacing and orientation of the axes, the wavelength of the radio waves received is determined by Eq. (16.15). (See Problem 16.12.)

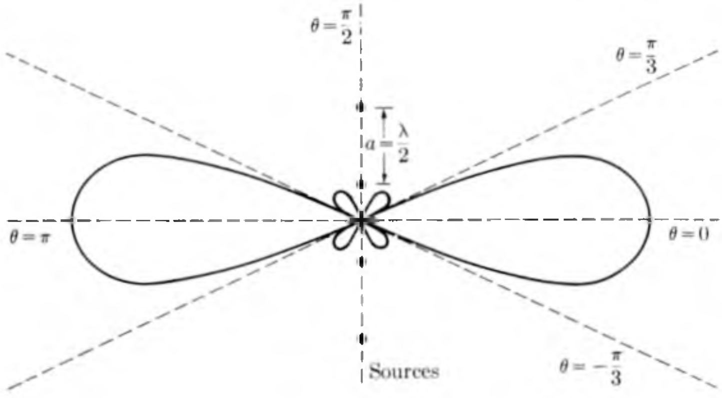


Fig. 16-12. Angular distribution of intensity in the interference pattern for waves generated by four coherent linear sources spaced a half wavelength apart.

Example 16.3. Interference by reflection from or transmission through thin films.

▼ The discussion of the previous section can be applied to the case of light reflected or transmitted by thin films. Consider (Fig. 16-13) a thin film of thickness a with plane waves falling on it at an angle of incidence θ_i. Part of a ray such as AB is reflected along BG, and part is refracted along BC. Ray BC in turn is partly reflected at C along CD and partly transmitted along CH. Ray CD again is partly reflected at D along DK, being superposed on the refracted ray of FD and partly transmitted along DE, superposed with the reflected ray of FD. Similarly the reflected ray BG also contains contributions from several rays to the left. Therefore interference phenomena occur along the reflected and refracted rays. The situation is then similar to the case illustrated in Section 16.3, with N very large, but with an important difference: the interfering rays do not all have the same intensity because each successive reflection or refraction decreases the intensity.

If we neglect this change in intensity, the maxima for interference by reflection or refraction occur when the phase difference δ between successive rays satisfies the equation $\delta = 2m\pi$ where m is an integer. To compute δ for interference by reflection, consider rays AB and FD. If we draw the wave front BB', the phase difference along DE is due to the difference in times required for

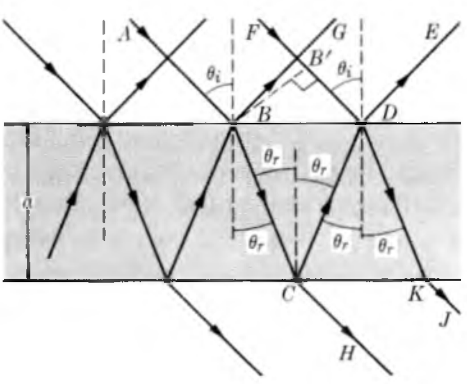


Fig. 16-13. Interference by reflection or on refraction through a thin film.

following paths $B'D$ and BCD. Now from the figure we see that $B'D = BD \sin \theta_i$ and $BD = 2a \tan \theta_r$. Therefore

$$B'D = 2a \tan \theta_r \sin \theta_i = \frac{2an \sin^2 \theta_r}{\cos \theta_r}$$

because by Snell's law $\sin \theta_i = n \sin \theta_r$. Also $BCD = 2BC = 2a/\cos \theta_r$. Then the times $t_1 = B'D/c = 2an \sin^2 \theta_r / c \cos \theta_r$ and $t_2 = BCD/v = 2an/c \cos \theta_r$ because $v = c/n$. The time difference is

$$t_2 - t_1 = \frac{2an \cos \theta_r}{c}$$

and the phase difference is

$$\delta = \omega(t_2 - t_1) = \frac{2a\omega n \cos \theta_r}{c} = \frac{4\pi an \cos \theta_r}{\lambda}$$

where the relation $\lambda = 2\pi c/\omega$ has been used. This may not be the entire phase difference because as we saw in Section 14.2, sometimes on reflection there is an additional phase difference of π radians. For example in the case of electromagnetic waves, when light polarized perpendicular to the plane of incidence goes from a medium in which the velocity is larger to another in which it is smaller, the phase changes by π radians. So in this case if $n > 1$, there is a phase shift of π for ray FD when it is reflected at D, but not for ray BC when it is reflected at C; the reverse occurs when $n < 1$. Thus in either case we must write

$$\delta = \frac{4\pi an \cos \theta_r}{\lambda} + \pi;$$

and setting $\delta = 2m\pi$ where m is an integer, we get

$$2an \cos \theta_r = \tfrac{1}{2}(2m - 1)\lambda \qquad \text{(maximum reflection, minimum transmission)}, \tag{16.17}$$

as the condition for interference by reinforcement in the reflected waves. The student may verify by similar calculations that for the transmitted wave along DK the condition for maximum intensity is

$$2an \cos \theta_r = m\lambda \qquad \text{(maximum transmission, minimum reflection)}. \tag{16.18}$$

The phase shift of π radians does not occur in this case because the ray has undergone two internal reflections. Equation (16.17) also gives the condition for minimum transmission and Eq. (16.18) gives the condition for minimum reflection. It is interesting to note, then, that the color we observe because by Snell's law $\sin \theta_i = n \sin \theta_r$. Also $BCD = 2BC = 2a/\cos \theta_r$. Then the time $t_1 = B'D/c = 2an \sin^2 \theta_r / c \cos \theta_r$ and $t_2 = BCD/v = 2an/c \cos \theta_r$ because $v = c/n$. The time difference is

If the incident light is not monochromatic, Eqs. (16.17) and (16.18) give different values of θ_r, and then θ_i, for each λ. The different values explain the colors we observe in thin oil films on water surfaces. Also if the film is of variable thickness, conditions (16.17) and (16.18) are not fulfilled at all points for a given wavelength; the result in the case of monochromatic light is a succession of dark and bright bands, and in the case of white light a succession of colored bands. This effect can easily be seen by placing a plano-convex lens on a plane glass plate as shown in Fig. 16-14a. The space between the lens and the glass plate is an air layer of varying thickness. The resulting

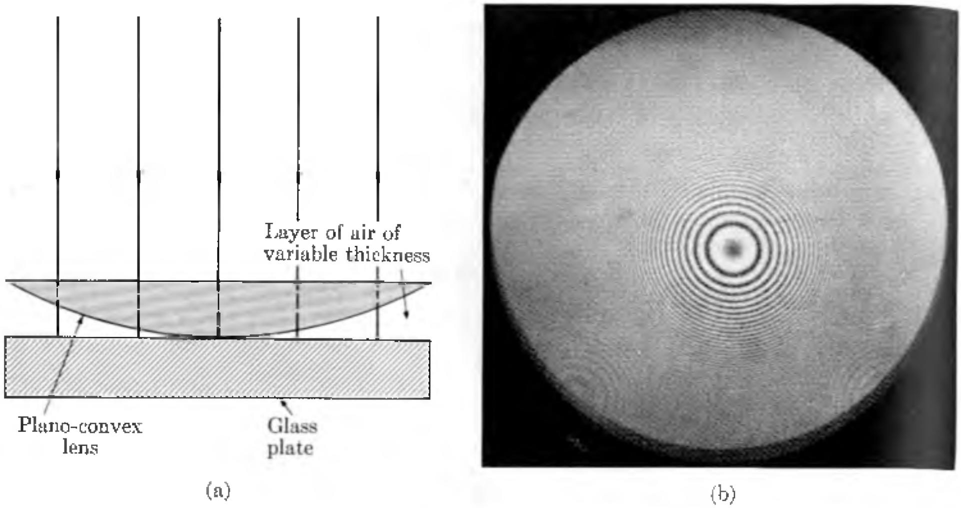


Fig. 16-14. Newton's rings, formed by interference in the air film between a convex and a plane surface. (a) Schematic diagram. (b) Photograph of rings. (Courtesy Bausch and Lomb Optical Co.)

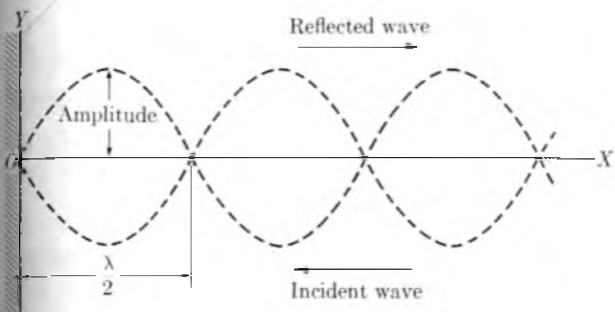
interference pattern consists of a series of concentric colored rings, known as *Newton's rings*, shown in Fig. 16-14b. ▲

Example 16.4. A given film has an index of refraction of 1.42. Determine its minimum thickness if it is to appear black by (a) reflection and (b) transmission when illuminated with sodium light (wavelength 5.9×10^{-7} m).

▼ From Eqs. (16.17) and (16.18) we see that the minimum value of a occurs at the maximum value of $\cos \theta_r$; that is, for $\theta_r = 0°$ and thus also $\theta_i = 0°$. This situation corresponds to normal incidence. In this case (with $m = 1$), the conditions become $a = \lambda/4n$ for no transmission and $a = \lambda/2n$ for no reflection. So the corresponding values are $a = 1.04 \times 10^{-7}$ m and $a = 2.08 \times 10^{-7}$ m. Since the separation between atoms is of the order of 10^{-9} m, the minimum film thickness in each case is only a few hundred atomic layers. ▲

16.4 Standing Waves in One Dimension

In Example 13.2 we discussed transmitted and reflected waves on a string when the string had a discontinuity at a certain point, such as a change in diameter or in material. Now consider the situation when one end of the string is fixed at point O as indicated in Fig. 16-15. An incident transverse wave moving to the left and having the equation $\xi = \xi_0 \sin(\omega t + kx)$ is reflected at O, and produces a new wave propagating to the right and having as its equation $\xi = \xi_0' \sin(\omega t - kx)$. The displacement at any



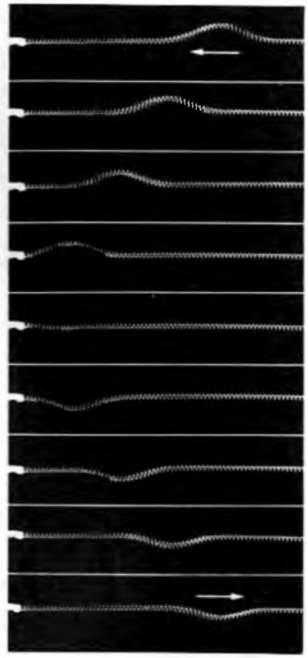


Fig. 16-15. Change in phase of a reflected wave on a string having one end fixed. (Photograph courtesy Educational Services, Inc.)

point of the string is the result of the interference or superposition of the two waves; that is,

$$\xi = \xi_0 \sin(\omega t + kx) + \xi_0' \sin(\omega t - kx). \tag{16.19}$$

At O, $x=0$ so that

$$\xi_{(x=0)} = (\xi_0 + \xi_0') \sin \omega t.$$

But that O is fixed means that $\xi_{(x=0)}=0$ at all times. Therefore $\xi_0' = -\xi_0$. In other words, the wave undergoes a phase change of π radians when it is reflected at the fixed end. We have encountered this phase change many times before (Examples 13.2 and 14.11). The phase change may be seen in the series of photographs of Fig. 16-15, which show an incident and a reflected pulse. Then Eq. (16.19) becomes

$$\xi = \xi_0[\sin(\omega t + kx) - \sin(\omega t - kx)].$$

Using the trigonometric relation $\sin\alpha - \sin\beta = 2 \sin \frac{1}{2}(\alpha-\beta) \cos[\frac{1}{2}(\alpha+\beta)]$, we obtain

$$\xi = 2\xi_0 \sin kx \cos \omega t. \tag{16.20}$$

The expressions $\omega t \pm kx$ no longer appear and Eq. (16.20) does not represent a traveling wave but rather a simple harmonic motion of angular frequency ω whose amplitude varies from point to point and is given by

$$A = 2\xi_0 \sin kx. \tag{16.21}$$

This amplitude has been indicated by the dashed lines in Fig. 16-15. The amplitude is zero for $kx=n\pi$ where n is an integer. This result may also be written as

$$x=\tfrac{1}{2}n\lambda. \tag{16.22}$$

These points of zero amplitude are called *nodes*. Successive nodes are separated by a distance of $\frac{1}{2}\lambda$. When we remember expression (10.37), $v=\sqrt{T/\sigma}$, for the velocity of propagation of waves along a string subject to a tension T and having a mass per unit length σ, the wavelength is determined by

$$\lambda=\frac{2\pi v}{\omega}=\frac{2\pi}{\omega}\sqrt{\frac{T}{\sigma}}, \tag{16.23}$$

and is completely arbitrary as long as the angular frequency ω is also arbitrary.

Suppose now that we impose a *second* condition: that the point $x=L$, which may be the other end of the string, is also fixed. That condition means that $x=L$ must be a node and must satisfy the condition $kL=n\pi$. If we use Eq. (16.22), we have

$$L=\tfrac{1}{2}n\lambda \qquad \text{or} \qquad \lambda=\frac{2L}{n}=2L, \frac{2L}{2}, \frac{2L}{3}, \ldots. \tag{16.24}$$

This second condition automatically limits the wavelengths of the waves that can travel on this string to the values given by Eq. (16.24). In view of Eq. (16.23), the

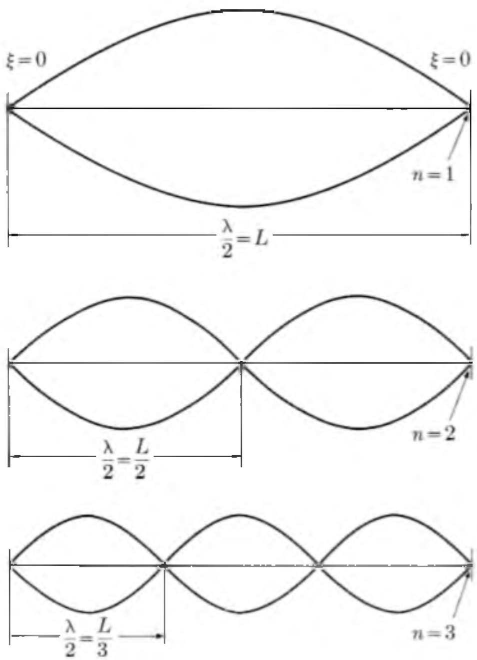


Fig. 16-16. Standing transverse wave on a string with both ends fixed.

frequencies of oscillations are also limited to the values

$$\nu_n = \frac{v}{\lambda} = \frac{\omega}{2\pi} = \frac{n}{2L}\sqrt{\frac{T}{\sigma}} = \nu_1, 2\nu_1, 3\nu_1, \ldots \tag{16.25}$$

where

$$\nu_1 = \frac{1}{2L}\sqrt{\frac{T}{\sigma}}$$

is called the *fundamental frequency*. Thus the possible frequencies of oscillation (called *harmonics*) are all multiples of the fundamental. We may say that the frequencies and wavelengths are *quantized*, and that the quantization is the result of the boundary conditions imposed at both ends of the string. This situation appears in many physical problems as we shall frequently have occasion to see later on.

Figure 16-16 indicates the amplitude distribution for the first three modes of vibration ($n = 1, 2, 3$). The nodes or points of zero amplitude are determined by means of Eq. (16.22). The points of maximum amplitude are the *antinodes*. The distance between successive antinodes is also $\frac{1}{2}\lambda$. Of course the separation between a node and an antinode is $\lambda/4$. Observe that while $\xi = 0$ at the nodes, $\partial\xi/\partial x = 0$ at the antinodes since the amplitude is maximum there.

Example 16.5. Tension in a musical string.

▼ A steel string has a length of 0.40 m and a diameter of 1 mm. Given that its fundamental vibration is 440 s^{-1}, corresponding to the musical tone A (or La) in the diatonic scale, key of C, find its tension. (Assume that the mass density of the string is $\rho = 7.86 \times 10^3$ kg m^{-3}.) Since the mass per unit length is $\sigma = \pi r^2 \rho$, using

$$r = 5 \times 10^{-4}\ \text{m} \qquad \text{and} \qquad \rho = 7.86 \times 10^3\ \text{kg m}^{-3},$$

we get $\sigma = 6.17 \times 10^{-3}$ kg m^{-1}. Solving Eq. (16.25) for the tension T, setting $n = 1$ since we want the fundamental tone, we obtain $T = 4L^2\sigma\nu_1^2$. Setting $L = 0.40$ m, $\nu_1 = 440$ s^{-1}, and introducing the value of σ that we calculated, we finally obtain

$$T = 764.9\ \text{N}.$$

Stringed musical instruments are usually tuned by adjusting the tensions or lengths of their strings. ▲

16.5 Standing Waves and the Wave Equation

In Chapter 10 we discussed the wave equation for the propagation of a wave; that is,

$$\frac{\partial^2 \xi}{\partial x^2} = \frac{1}{v^2}\frac{\partial^2 \xi}{\partial t^2}; \tag{16.26}$$

and proved that its general solution is of the form

$$\xi = f_1(x - vt) + f_2(x + vt). \tag{16.27}$$

When we discuss a wave propagating in only one direction, we use either $f_1(x - vt)$ or $f_2(x + vt)$, but not both. However, we have seen that when a wave is reflected at one point, the result is *two* waves traveling in opposite directions, and Eq. (16.27) must be used. This use of both forms is what we did in Eq. (16.19) for the case of a string with one end fixed; we then obtained Eq. (16.20) for the resultant wave motion. The important feature of Eq. (16.20) (that is, $\xi = 2\xi_0 \sin kx \cos \omega t$) is that the x and t variables are separated, resulting in a variable amplitude along the string, but a fixed amplitude for each point. This separation of time and space variables is the characteristic of standing waves. We must then explore the possibility of a more general formulation of a harmonic standing wave. Our requirement can be met by an expression of the form

$$\xi = f(x) \sin \omega t \tag{16.28}$$

where $f(x)$ is the amplitude of the wave at point x. Since ξ must be a solution of Eq. (16.26), we must substitute ξ as given in Eq. (16.28) into Eq. (16.26) to determine the requirements on the amplitude $f(x)$ for standing waves. Now by differentiation, we find that

$$\frac{\partial^2 \xi}{\partial x^2} = \frac{d^2 f}{dx^2} \sin \omega t \qquad \text{and} \qquad \frac{\partial^2 \xi}{\partial t^2} = -\omega^2 f(x) \sin \omega t.$$

Therefore, substituting these values in Eq. (16.26) and canceling the common factor $\sin \omega t$, we obtain

$$\frac{d^2 f}{dx^2} = -\frac{\omega^2}{v^2} f,$$

or since $k = \omega/v$,

$$\frac{d^2 f}{dx^2} + k^2 f = 0. \tag{16.29}$$

This is the differential equation that must be satisfied by the amplitude $f(x)$ if the standing wave given by Eq. (16.28) is to be a solution of the wave equation. The general solution of Eq. (16.29), as the student may verify by direct substitution, is

$$f(x) = A \sin kx + B \cos kx \tag{16.30}$$

where A and B are arbitrary constants. Therefore Eq. (16.28) becomes

$$\xi = (A \sin kx + B \cos kx) \sin \omega t. \tag{16.31}$$

Of course, we could have used $\cos \omega t$ instead of $\sin \omega t$, with the same result. In other words, the phase of the time factor is irrelevant for our discussion.

The constants in Eq. (16.31) are determined by the conditions imposed at the boundaries. Let us illustrate this determination with the problem of the string with fixed ends discussed in the preceding section. The conditions are that $\xi = 0$ for both

$x=0$ and $x=L$. Setting $x=0$ in Eq. (16.31), we have

$$\xi_{(x=0)}=B\sin\omega t=0.$$

Therefore $B=0$, and Eq. (16.31) reduces to

$$\xi=A\sin kx\sin\omega t. \tag{16.32}$$

Now if we set $x=L$, Eq. (16.32) yields

$$\xi_{(x=L)}=A\sin kL\sin\omega t=0.$$

But now we cannot set $A=0$ because that would make $\xi=0$ everywhere; i.e., we would not have any wave at all. Thus our only choice is to set $\sin kL=0$, which requires that

$$kL=n\pi \qquad \text{or} \qquad \lambda=\frac{2L}{n} \tag{16.33}$$

where n is an integer, in agreement with Eq. (16.24).

If, instead of imposing the condition that $\xi=0$ at the ends, we impose other conditions because the physical situation at the ends is different as in the strings illustrated in Fig. 16-17, we would end up with a solution different from Eqs. (16.32) and (16.33).

It is instructive to investigate two other simple examples, related to standing waves in the air inside a pipe, such as an organ pipe. Consider first a tube open at both ends (Fig. 16-18). Air is blown in at one end through the mouthpiece, and standing waves are produced because of the reflection occurring at the other end. The fundamental difference between this case and the previous one is that both ends are free, and therefore ξ has a maximum value at these ends because there cannot be a pressure difference between outside and inside; in other words, there is an antinode at each end. Our boundary conditions, corresponding to antinodes at both ends, are now $\xi=$ maximum, or $\partial\xi/\partial x=0$ at $x=0$ and $x=L$. From Eq. (16.31), we have

$$\frac{\partial\xi}{\partial x}=k(A\cos kx-B\sin kx)\sin\omega t. \tag{16.34}$$

Setting $x=0$, we obtain

$$\left(\frac{\partial\xi}{\partial x}\right)_{x=0}=kA\sin\omega t=0$$

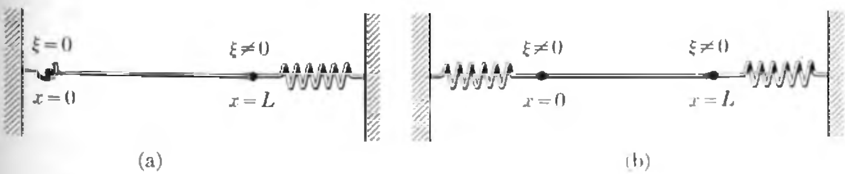


Figure 16-17

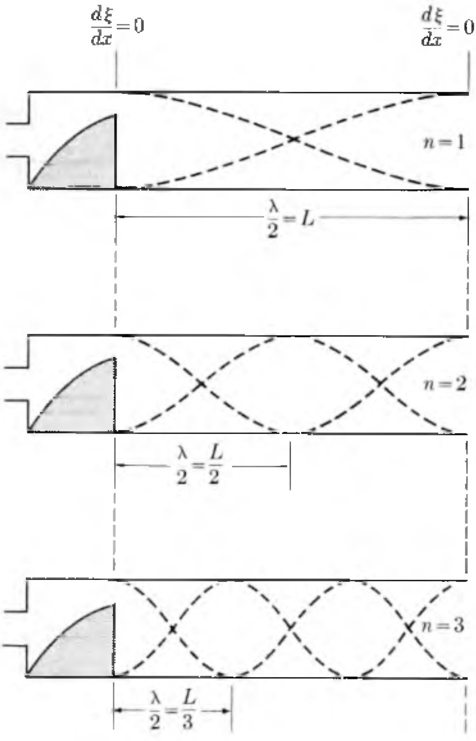


Fig. 16-18. Standing pressure wave in an air column with both ends open.

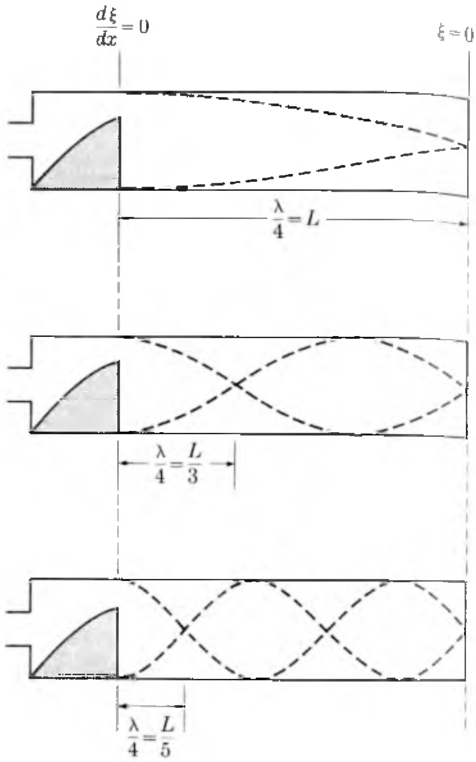


Fig. 16-19. Standing pressure wave in an air column with one end closed.

so that $A=0$. Then Eq. (16.34) becomes

$$\frac{\partial \xi}{\partial x}=-kB \sin kx \sin \omega t.$$

If we now set $x=L$, we have

$$\left(\frac{\partial \xi}{\partial x}\right)_{x=L}=-kB \sin kL \sin \omega t=0.$$

Now, just as in the case of the string, we cannot make $B=0$ because then we would have no wave at all, and our only choice is $\sin kL=0$, which again gives us

$$kL=n\pi \qquad \text{or} \qquad \lambda=\frac{2L}{n}. \tag{16.35}$$

This equation is identical to Eq. (16.33). The frequencies of the standing waves are

$$\nu_n=\frac{v}{\lambda}=n\left(\frac{v}{2L}\right)=\nu_1,\ 2\nu_1,\ 3\nu_1,\ \ldots, \tag{16.36}$$

with $n=1, 2, 3, \ldots$, and therefore the allowed frequencies comprise all the harmonics corresponding to a fundamental tone of frequency $\nu_1 = v/2L$. From Eq. (16.31) we see that in this case with $A=0$, $\xi = B \cos kx \sin \omega t$. In Fig. 16-18 the dashed lines indicate the amplitude distribution for the cases $n=1$, 2, and 3. We conclude then that the oscillations of an air column open at both ends are equivalent to those of a string with both ends fixed, but the positions of the nodes and antinodes are interchanged.

As our second example, consider a tube closed at the end opposite to the mouthpiece (Fig. 16-19). The physical conditions at that end have changed while at the mouthpiece they remain the same as in the preceding case. Therefore at the mouthpiece we must again have an antinode, or $\partial\xi/\partial x = 0$ at $x=0$; but at the closed end $(x=L)$ we must have a node, or $\xi=0$ at $x=L$. The first condition, at $x=0$, requires, as in the previous example, that $A=0$ so that Eq. (16.31) becomes

$$\xi = B \cos kx \sin \omega t.$$

Applying the boundary condition at the closed end, $x=L$, we obtain

$$\xi_{(x=L)} = B \cos kL \sin \omega t = 0.$$

This equation requires that $\cos kL = 0$. In other words,

$$kL = (2n'+1)\frac{\pi}{2} \qquad \text{or} \qquad \lambda = \frac{4L}{2n'+1} \tag{16.37}$$

with $n' = 0, 1, 2, 3 \ldots$ and with the corresponding frequencies

$$\nu = \frac{v}{\lambda} = (2n'+1)\frac{v}{4L} = \nu_1, 3\nu_1, 5\nu_1, \ldots. \tag{16.38}$$

The modes of vibration are now different from those given by Eqs. (16.35) and (16.36), corresponding to a tube open at both ends. Figure 16-19 shows the nodes and antinodes for the open-closed case for $n'=0$, 1, and 2. The most important feature is that a tube closed at one end can vibrate only with *odd* harmonics of the fundamental $\nu_1 = v/4L$. For equal lengths, the fundamental frequency of a closed tube is one-half that of an open tube.

A solution of the wave equation of type (16.23) corresponds to a standing harmonic wave of angular frequency ω. In general, however, the disturbance set up initially does not correspond to a particular frequency. In order to determine the wavelengths and frequencies involved, a Fourier analysis of the initial disturbance must be made. The disturbance at any later time is

$$\xi = \sum_{\omega} (A \sin kx + B \cos kx) \sin \omega t \tag{16.39}$$

where $k = \omega/c$ and the coefficients A and B are determined from the Fourier analysis. But Eq. (16.39) does not represent a standing wave in the sense defined before (i.e., a wave whose amplitude depends on the position) because as a result of the summation sign, the time and position variables are not completely separated.

Example 16.6. Determination of the frequency of oscillation of a mass suspended from a spring.

▼ Consider a spring of mass m_0, length L, and elastic constant κ. The spring is suspended from a fixed point and has a bob of mass M attached to the free end. The bob is displaced vertically from its equilibrium position and then released.

We would be tempted to say that the angular frequency is $\omega=\sqrt{\kappa/M}$ as we did in Volume I when first studying simple harmonic motion (here the elastic constant of the spring is given as κ to avoid confusion with the wave number, designated by k). However, now we shall find that this frequency is correct only when the mass of the spring is negligible compared with the mass of the bob. From Fig. 16-20, we see that when we hang the bob M, the spring stretches until the upward force it produces on M balances the weight of M. If M is now set into oscillation, waves are produced in the spring, traveling up and down, resulting in standing waves. The frequency of oscillation of M is determined by the frequency of the standing waves produced in the spring. Let us designate the displacement of each section of the spring by ξ. The boundary condition at the fixed end, $x=0$, is $\xi=0$. This requires that $B=0$ in Eq. (16.31) so that the displacement of every section of the spring is given by

$$\xi = A \sin kx \sin \omega t. \tag{16.40}$$

At the free end of the spring, M is accelerated by the force of the stretched spring, which, according to Example 10.6, is $-K(\partial\xi/\partial x)_{x=L}$ where $K=\kappa L$ is the elastic modulus of the spring as defined in that example. The negative sign appears because the positive direction for ξ is downward and the force is upward (or negative) when $\partial\xi/\partial x$ is positive. Therefore the equation of motion of M is

$$M\left(\frac{\partial^2\xi}{\partial t^2}\right)_{x=L} = -K\left(\frac{\partial\xi}{\partial x}\right)_{x=L}.$$

This equation gives the boundary condition at the free end of the spring. The equation above may be written (using the expression for ξ given before) as

$$-M\omega^2 \sin kL = -Kk \cos kL.$$

Setting $k=\omega/v$ and recalling that $\sigma=m_0/L$ is the mass of the spring per unit length and $v=\sqrt{K/\sigma}$ according to the result obtained in Example 10.6, we have

$$\frac{\omega L}{v}\tan\frac{\omega L}{v} = \frac{Mv^2}{Mv^2} = \frac{M}{M} = \frac{M}{M}. \tag{16.41}$$

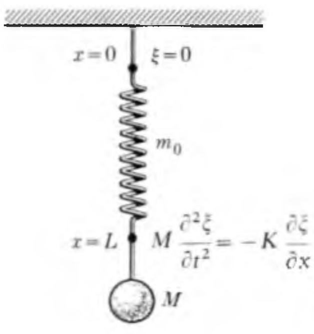


Figure 16-20

This is an equation of the type $\theta \tan \theta = \text{const}$ where $\theta = \omega L/v$. The solution of this equation gives the possible values of the angular frequency ω. It is a transcendental equation and cannot be solved by ordinary algebraic methods. However when the spring is very light so that v is very large and thus θ is very small, we may use the approximation $\tan \theta = \theta + \frac{1}{3}\theta^3 + \cdots$. Then the left-hand side of the equation becomes

$$\frac{\omega L}{v}\left[\frac{\omega L}{v} + \frac{1}{3}\left(\frac{\omega L}{v}\right)^3 + \cdots\right] = \left(\frac{\omega L}{v}\right)^2\left[1 + \frac{1}{3}\left(\frac{\omega L}{v}\right)^2 + \cdots\right].$$

Recalling that $K = \kappa L$ and thus $v = \sqrt{K/\sigma} = L\sqrt{\kappa/m_0}$, we have $\omega L/v = \omega\sqrt{m_0/\kappa}$ so that Eq. (16.41) may be written as

$$\frac{\omega^2 m_0}{\kappa}\left(1 + \frac{\omega^2 m_0}{3\kappa} + \cdots\right) = \frac{m_0}{M}$$

or

$$\omega^2\left(1 + \frac{\omega^2 m_0}{3\kappa} + \cdots\right) = \frac{\kappa}{M}.$$

As a first approximation, if m_0 is very small, we may neglect the second term inside the parentheses; the result is $\omega^2 = \kappa/M$ or $\omega = \sqrt{\kappa/M}$, which is the value we obtained in our original study of simple harmonic motion. As a second approximation, we introduce this value of ω in the second term in the parentheses; the result is

$$\omega^2\left(1 + \frac{1}{3}\frac{m_0}{M}\right) = \frac{\kappa}{M}$$

or

$$\omega = \sqrt{\frac{\kappa}{M + \frac{1}{3}m_0}}. \tag{16.42}$$

Therefore for small oscillations the effect of the spring on the angular frequency ω is equivalent to increasing the mass of the body by one-third the mass of the spring. This expression gives the fundamental frequency; but in addition, there are overtones that are not integral multiples of the fundamental. ▲

16.6 Standing Electromagnetic Waves

Interference and diffraction phenomena are so characteristic of waves that their presence has always been accepted by physicists as proof that a process can be interpreted as a wave motion. For that reason, when in the seventeenth century the Italian scientist Francesco Grimaldi (1618–1663) observed interference and diffraction in his research on light, the wave theory of light could no longer be ignored. However, it was not until early in the nineteenth century, when Young, Fresnel, and others performed the experiments mentioned earlier in this chapter and supported their experimentation with a strong mathematical foundation, that the wave theory of light became generally accepted. At that time electromagnetic waves

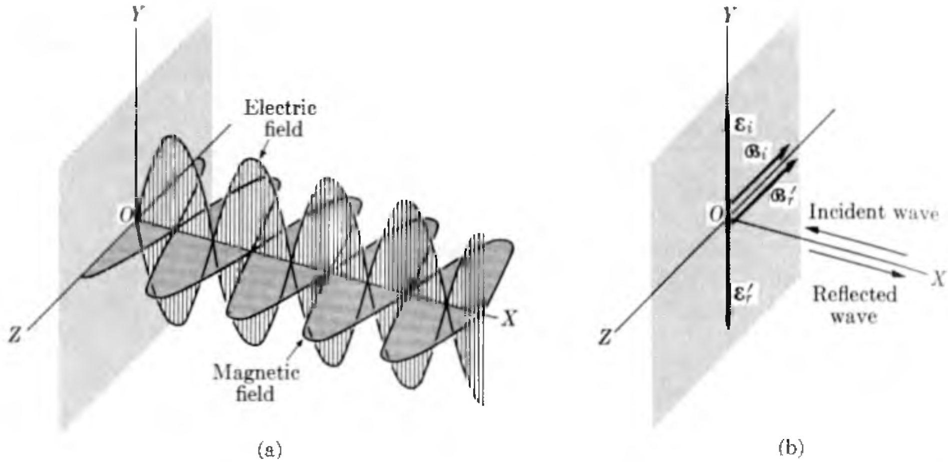


Fig. 16-21. Standing electromagnetic waves produced by reflection from a conducting surface.

were not known, and light was assumed to be an elastic wave in a subtle medium, called ether, that pervaded all matter. It was not until the end of the nineteenth century that the Scottish physicist James Clerk Maxwell (1831–1879) predicted the existence of electromagnetic waves, and the German scientist Heinrich R. Hertz (1857–1894), by means of interference experiments that gave rise to standing electromagnetic waves, experimentally verified the existence of electromagnetic waves in the radiofrequency range. Later their velocity was measured and found to be equal to that of light. The reflection, refraction, and polarization of electromagnetic waves were also found to be similar to those of light. The obvious conclusion was to identify light with electromagnetic waves of certain frequencies. At that time optics, to all intents and purposes, ceased to be an independent branch of physics and became simply a chapter of electromagnetic theory.

To understand the formation of standing electromagnetic waves, assume that the waves produced by an oscillating electric dipole are falling with perpendicular incidence on the plane surface of a perfect conductor (Fig. 16-21). Taking the X-axis as the direction of propagation and the Y- and Z-axes parallel to the electric and the magnetic fields, respectively, we have a wave that is plane polarized with the electric field oscillating in the XY-plane. The electric field of the incident wave is parallel to the surface of the conductor. But at the surface of a perfect conductor the electric field must be perpendicular to the conductor; that is, the electric field cannot have a tangential component (see Section 2.5 and Example 14.1). The only way to make this condition compatible with the orientation of the electric field in the incident wave is to require that the resultant electric field $\mathcal{E}$ be zero at the surface of the conductor. This requirement means that the electric field of the reflected wave at the surface must be equal and opposite to that of the incident wave so that $\mathcal{E}=0$ for $x=0$. This condition is mathematically equivalent to the condition for the reflection of waves in a string with one end fixed, a condition discussed in Section 16.5. Since the mathematics is the same, we may use an equation of the same form as Eq. (16.20) for the resultant electric field:

$$\mathscr{E} = 2\mathscr{E}_0 \sin kx \sin \omega t.$$

The magnetic field oscillates in the XZ-plane. Using Eq. (11.4), we find that the magnetic field is expressed by

$$\mathscr{B} = 2\mathscr{B}_0 \cos kx \cos \omega t,$$

with $\mathscr{B}_0 = \mathscr{E}_0 k/\omega = \mathscr{E}_0/c$. Therefore there is a phase difference of $\frac{1}{2}\lambda$ in the space variations and of $\frac{1}{2}P$ in the time variations of the two fields. From the mathematical expression for $\mathscr{B}$ given above, we see that the magnetic field has maximum amplitude at the surface. This can also be seen from the boundary condition at the surface: referring to Fig. 16-21b, we see that if the electric field of the incident wave is along the $+Y$-axis, the magnetic field must be along the $-Z$-axis, according to the relative orientation of the two fields with respect to the direction of propagation of the incident waves, which is along the $-X$-axis. For a zero resultant electric field to exist at the surface, the electric field of the reflected wave must be along the $-Y$-axis: and since the reflected wave propagates along the X-axis, the magnetic field must be along the $-Z$-axis. Thus, although the electric fields interfere destructively at the surface, the magnetic fields interfere constructively there.

The amplitudes of the electric and magnetic fields of the resulting wave at a distance x from the surface are $2\mathscr{E}_0 \sin kx$ and $2\mathscr{B}_0 \cos kx$. They are denoted by the shaded lines in Fig. 16-21a. At the points where $kx = n\pi$ or $x = \frac{1}{2}n\lambda$, the electric field is zero and the magnetic field is maximum. At the points where $kx = (n + \frac{1}{2})\pi$ or $x = (2n + 1)\lambda/4$, the electric field has a maximum value but the magnetic field is zero.

It is instructive to see how Hertz, in 1888 with his primitive equipment, verified these theoretical predictions. Hertz's oscillator is shown on the left in Fig. 16-22. The transformer T charges the metallic plates C and C'. These plates discharge through the gap G, which becomes the dipole oscillator. The directions of the $\mathscr{E}$- and $\mathscr{B}$-fields relative to the direction of propagation are also shown. To observe the waves, Hertz used a short wire, bent in circular shape, but with a small gap. This device is called a *resonator*. The diameter of the resonator used in this kind of experiment must be very small compared with the wavelength of the waves. If the resonator is placed with its plane perpendicular to the magnetic field of the wave (that is, in the XY-plane), the varying magnetic field induces an emf in the resonator, resulting in sparks at its gap when the induced emf is large enough. On the other hand, if the plane of the resonator is parallel to the magnetic field (so that the normal to the plane is perpendicular to the Z-axis), no emf is induced and no sparks are observed at the gap.

To produce standing electromagnetic waves, Hertz placed a reflecting surface (made of a good conductor) at Q. In such a case, when the resonator is at a node of the magnetic field, no matter what its orientation, it will show no induced emf (or

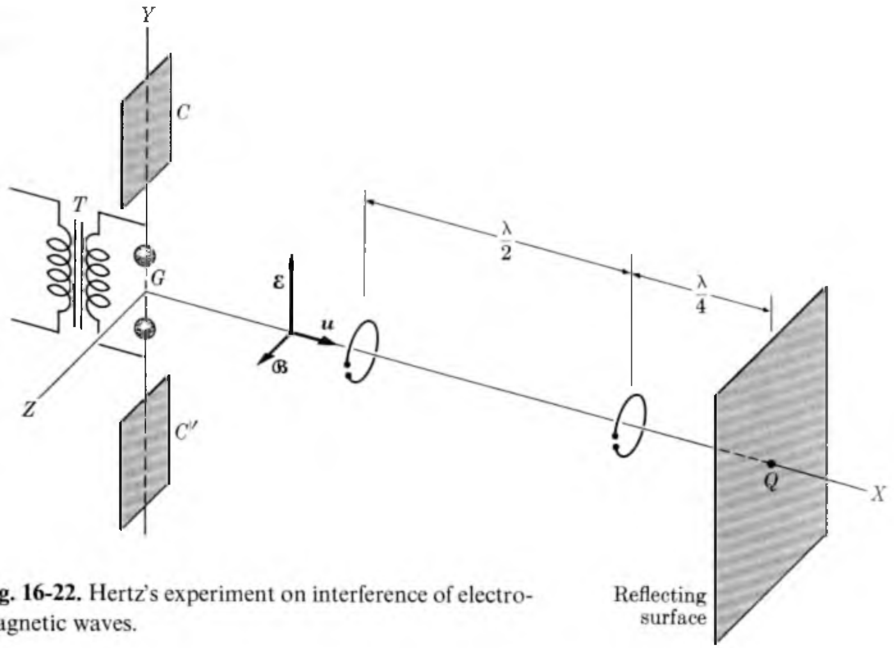


Fig. 16-22. Hertz's experiment on interference of electromagnetic waves.

Reflecting surface

sparks). At an antinode of the magnetic field, however, the sparking is greatest when the resonator is oriented perpendicular to the magnetic field. By moving the resonator along the line GQ, Hertz found the position of the nodes and antinodes and the direction of the magnetic field. The results obtained by Hertz coincided with the theoretical analysis we have given. By measuring the distance between two successive nodes, Hertz could calculate the wavelength λ; and since he knew the frequency ν of the oscillator, he could calculate the velocity of the electromagnetic waves by using the equation $c=\lambda\nu$. By this means Hertz obtained the first experimental value for the velocity of propagation of electromagnetic waves.

16.7 Standing Waves in Two Dimensions

Consider a rectangular membrane stretched over a frame so that the membrane edges are fixed. If the surface of the membrane is disturbed, waves are propagated in all directions, are reflected at the edges, and result in interference. Assume the special case in which plane waves of only one frequency are generated in the membrane. Assume further that these waves propagate parallel to either side as indicated in Fig. 16-23. Instead of nodes and antinodes we get nodal lines and antinodal lines, designated by N and A in Fig. 16-23. In Fig. 16-23a, the membrane is fixed at the left ($x=0$) and the right ($x=a$), but the other two sides are free. The waves propagate along the X-axis, both to the left and to the right, and result in a system of nodal and

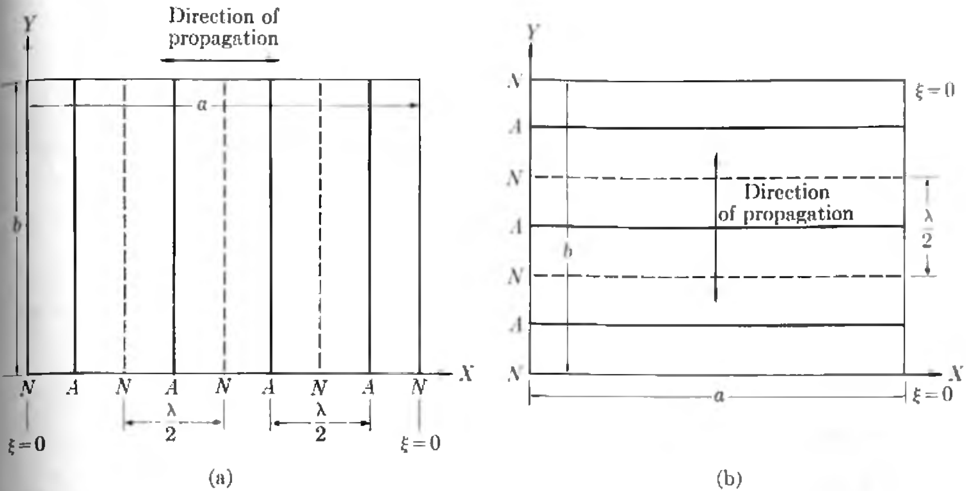


Fig. 16-23. Standing waves on a membrane.

antinodal lines parallel to the Y-axis. At $x=0$ and $x=a$ we must have nodal lines. Therefore the condition for standing waves is similar to Eq. (16.33); that is,

$$ka = n\pi \qquad \text{or} \qquad \lambda = \frac{2a}{n}. \tag{16.43}$$

The corresponding frequencies are

$$\nu = \frac{v}{\lambda} = n\left(\frac{v}{2a}\right) \tag{16.44}$$

where v is the velocity of propagation of the waves along the surface of the membrane as given in Example 10.15. These waves are described by an expression similar to Eq. (16.32),

$$\xi = A \sin kx \sin \omega t, \tag{16.45}$$

since the problem is mathematically the same. The addition of the second dimension has not changed our boundary conditions, which are still $\xi=0$ for $x=0$ and $x=a$. Symmetry suggests that the coordinate y plays no role so long as the membrane is not fixed along the sides parallel to the direction of propagation.

In Fig. 16-23b the membrane is fixed at the bottom ($y=0$) and at the top ($y=b$). For waves propagating parallel to the Y-axis the nodal and antinodal lines are parallel to the X-axis. The condition for standing waves is similar to Eq. (16.43), with a replaced by b, yielding

$$kb = n\pi \qquad \text{or} \qquad \lambda = \frac{2b}{n} \tag{16.46}$$

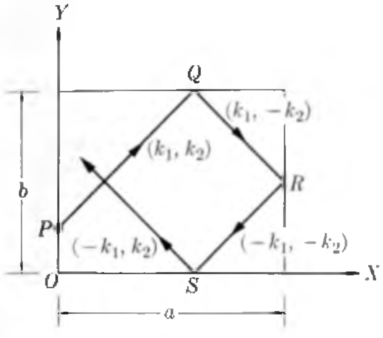


Fig. 16-24. Successive reflections of a wave in a rectangular membrane.

with frequencies

$$\nu = \frac{v}{\lambda} = n\left(\frac{v}{2b}\right), \tag{16.47}$$

which are different from those given in Eq. (16.44) for waves parallel to the X-axis. The equation of the standing waves is

$$\xi = A \sin ky \sin \omega t. \tag{16.48}$$

Next consider a membrane with all four sides fixed and plane waves traveling in an arbitrary direction along its surface. Recall that a plane wave in two dimensions [setting $z=0$ in Eq. (10.63)] is expressed by

$$\xi = \xi_0 \sin [\omega t - (k_1 x + k_2 y)]$$

where we have followed our present convention of writing the time factor first. The quantities k_1 and k_2 are the components of a vector $\boldsymbol{k}$ parallel to the direction of propagation in the XY-plane and of length $k = 2\pi/\lambda = \omega/v$. Then

$$k = \sqrt{k_1^2 + k_2^2}. \tag{16.49}$$

For an initial ray PQ (Fig. 16-24), characterized by the components k_1, k_2, there is a reflected ray QR characterized by k_1, $-k_2$. From R to S the ray is characterized by $-k_1$, $-k_2$. And from S on, the ray is characterized by the components $-k_1$, k_2. In the successive reflections of this ray, no new combinations of k_1 and k_2 appear. We conclude then that along the membrane there is a system of four waves, caused by reflection at the four sides. (This situation is different from one-dimensional problems, in which only two waves appear.) These four waves must interfere in such a way that at $x=0$ and a, and $y=0$ and b, the resultant value of ξ is zero. A direct algebraic procedure shows (as we shall see in Example 16.7) that the values of k_1 and k_2 satisfy the conditions

$$k_1 a = n_1 \pi \qquad \text{or} \qquad k_1 = \frac{n_1 \pi}{a}$$

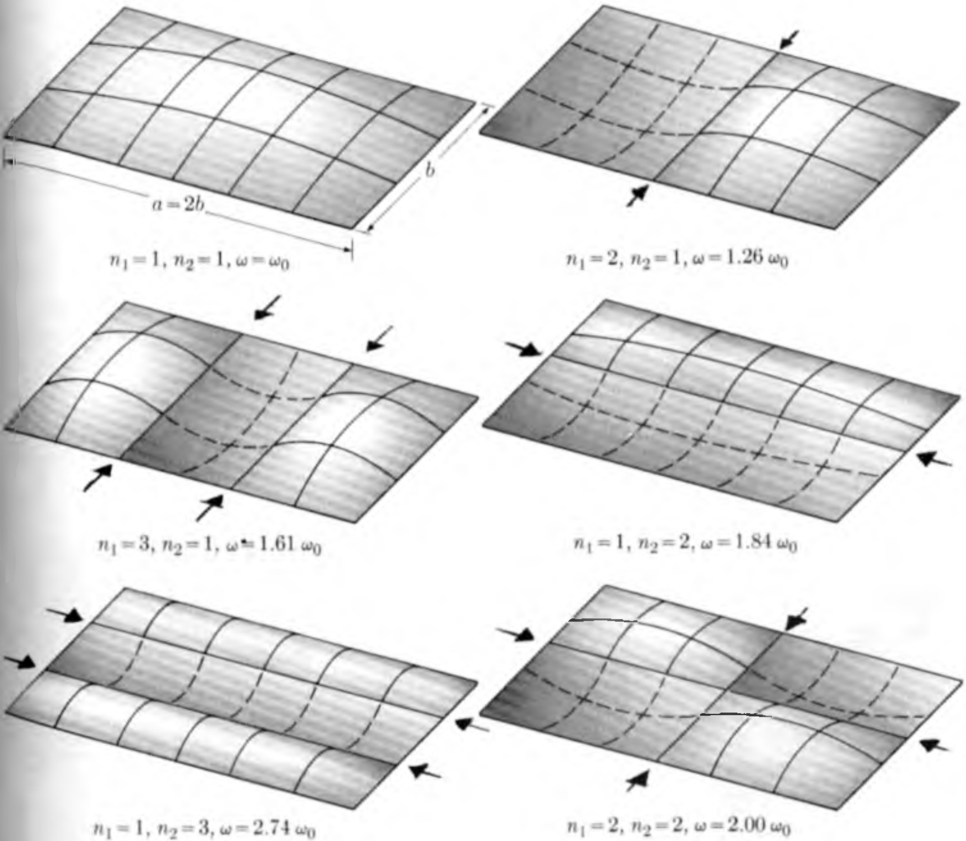


Fig. 16-25. First few modes of vibration of a rectangular membrane. Nodal lines are indicated by arrows. The frequency of each mode is given in terms of the fundamental frequency $\omega_0=\pi v/b$.

and

$$k_2 b = n_2\pi \qquad \text{or} \qquad k_2=\frac{n_2\pi}{b} \tag{16.50}$$

where n_1 and n_2 are integers. Then, by Eq. (16.49), we have

$$k=\pi\sqrt{\frac{n_1^2}{a^2}+\frac{n_2^2}{b^2}}. \tag{16.51}$$

The possible wavelengths are given by $\lambda=2\pi/k$ or

$$\frac{1}{\lambda}=\frac{1}{2}\sqrt{\frac{n_1^2}{a^2}+\frac{n_2^2}{b^2}}. \tag{16.52}$$

For the possible frequencies we use $\nu=v/\lambda$ so that

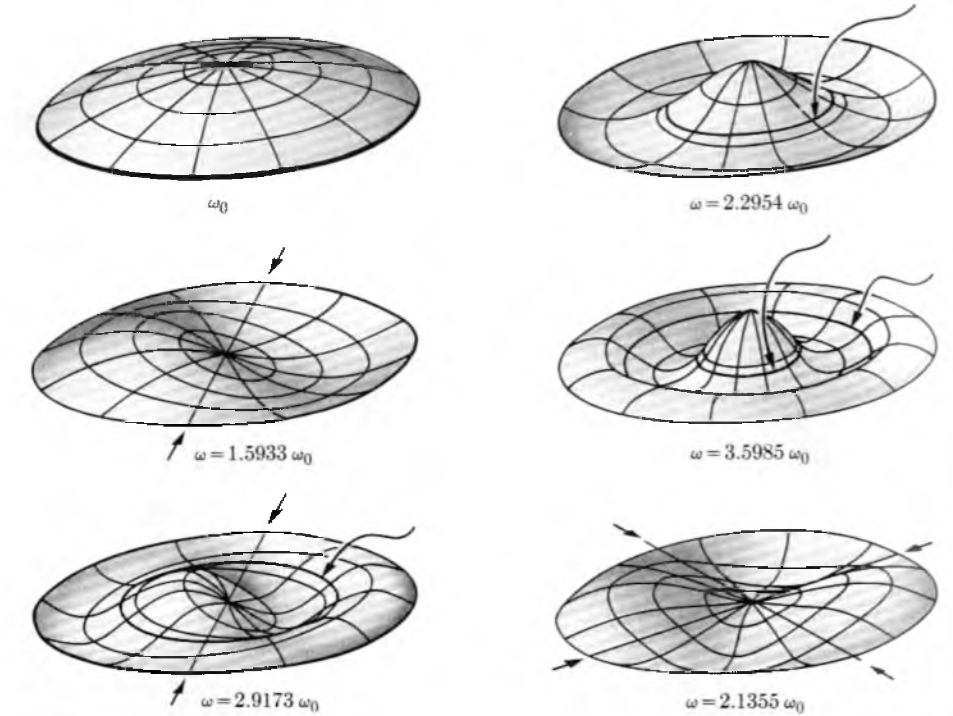


Fig. 16-26. Some possible modes of vibration of a circular membrane. Nodal lines are indicated by arrows. The frequency of each mode is given in terms of the fundamental frequency ω_0. (*Vibration and Sound*, by Philip M. Morse, McGraw-Hill Book Co., 1948.)

$$\nu=\frac{v}{2}\sqrt{\frac{n_1^2}{a^2}+\frac{n_2^2}{b^2}}\,. \tag{16.53}$$

Notice that the possible frequencies are no longer integers of a fundamental frequency, but follow a more irregular sequence.

The pattern of nodal lines obtained by use of Eq. (16.50) is given by $k_1x=n_1'\pi$ and $k_2y=n_2'\pi$ where n_1' and n_2' are integers equal to or smaller than n_1 and n_2, respectively, and form the rectangular patterns shown in Fig. 16-25.

The problem of a circular membrane is more complex mathematically; however, once more we find that only certain frequencies are possible. Symmetry suggests that the nodal lines are now circles and radii as indicated in Fig. 16-26 for some of the possible modes.

Example 16.7. Derivation of conditions (16.50).

▼ We have indicated that on the membrane of Fig. 16-24 there is a superposition of four waves corresponding to the four possible combinations of $\pm k_1$ and $\pm k_2$. The system of four waves gives a resultant ξ,

$$\begin{aligned}\xi=\xi_0 &\sin[\omega t-(k_1x+k_2y)]\\ &+\xi_0' \sin[\omega t-(k_1x-k_2y)]\\ &+\xi_0'' \sin[\omega t-(-k_1x-k_2y)]\\ &+\xi_0''' \sin[\omega t-(-k_1x+k_2y)],\end{aligned} \tag{16.54}$$

which is the equivalent of Eq. (16.19) for one dimension. Now at all points where $x=0$, we must have $\xi=0$. Setting $x=0$ in Eq. (16.54) and grouping equivalent terms, we may write

$$\xi=(\xi_0+\xi_0''')\sin(\omega t-k_2y)+(\xi_0'+\xi_0'')\sin(\omega t+k_2y)=0,$$

which requires that

$$\xi_0+\xi_0'''=0, \qquad \xi_0'+\xi_0''=0. \tag{16.55}$$

Similarly, at all points where $y=0$, we must have $\xi=0$ so that Eq. (16.54) yields

$$\xi=(\xi_0+\xi_0')\sin(\omega t-k_1x)+(\xi_0''+\xi_0''')\sin(\omega t+k_1x)=0,$$

which requires that

$$\xi_0+\xi_0'=0, \qquad \xi_0''+\xi_0'''=0. \tag{16.56}$$

Combining Eqs. (16.55) and (16.56), we find that

$$\xi_0=-\xi_0'=\xi_0''=-\xi_0''',$$

which gives the appropriate phase changes at each reflection, in agreement with the similar result in the case of a string. Therefore Eq. (16.54) becomes

$$\begin{aligned}\xi=\xi_0\{&\sin[\omega t-(k_1x+k_2y)]-\sin[\omega t-(k_1x-k_2y)]\\ &+\sin[\omega t-(-k_1x-k_2y)]-\sin[\omega t-(-k_1x+k_2y)]\}.\end{aligned}$$

Transforming each line of the formula above into a product, using Eq. (M.12), we have

$$\begin{aligned}\xi&=2\xi_0[-\sin k_2y\cos(\omega t-k_1x)+\sin k_2y\cos(\omega t+k_1x)]\\ &=2\xi_0\sin k_2y[-\cos(\omega t-k_1x)+\cos(\omega t+k_1x)].\end{aligned}$$

Again transforming the difference of the two cosines into a product, we obtain

$$\xi=-4\xi_0\sin k_1x\sin k_2y\sin\omega t, \tag{16.57}$$

which is the two-dimensional equivalent of Eq. (16.20). The minus sign in front is irrelevant and has no special meaning. We may check our first boundary condition by setting $x=0$ or $y=0$, and observing that we get $\xi=0$, which was our requirement.

We must now verify the second set of boundary conditions; that is, $\xi=0$ for $x=a$ or $y=b$. These conditions require that either $\sin k_1a=0$ or $\sin k_2b=0$, resulting in $k_1a=n_1\pi$ and $k_2b=n_2\pi$. These are the conditions (16.50). ▲

16.8 Standing Waves in Three Dimensions; Resonating Cavities

The problem of standing waves in three dimensions is a simple extension of the case for two dimensions. Consider a rectangular cavity of sides a, b, and c with perfectly reflecting walls (Fig. 16-27) so that $\xi=0$ at all six faces. A plane wave in space is characterized by a vector $\boldsymbol{k}$, perpendicular to the plane of the wave, with three components, k_1, k_2, and k_3, along the three axes. When a wave is produced inside the cavity, the wave is reflected successively at all faces, and a set of eight waves, resulting from all the different combinations possible among $\pm k_1$, $\pm k_2$, and $\pm k_3$, is established. The interference or superposition of these eight waves gives rise to standing waves if the k_1, k_2, and k_3 components of $\boldsymbol{k}$ have the appropriate values. By analogy with Eq. (16.50), these values are

$$k_1 a = n_1 \pi, \qquad k_2 b = n_2 \pi, \qquad k_3 c = n_3 \pi,$$

or (16.58)

$$k_1 = n_1 \frac{\pi}{a}, \qquad k_2 = n_2 \frac{\pi}{b}, \qquad k_3 = n_3 \frac{\pi}{c}$$

where n_1, n_2, and n_3 are integers. Since $k = \sqrt{k_1^2 + k_2^2 + k_3^2}$, we may write

$$k = \pi \sqrt{\frac{n_1^2}{a^2} + \frac{n_2^2}{b^2} + \frac{n_3^2}{c^2}}. \tag{16.59}$$

With $\nu = v/\lambda = vk/2\pi$, the possible frequencies of the standing waves in the cavity are

$$\nu = \frac{v}{2} \sqrt{\frac{n_1^2}{a^2} + \frac{n_2^2}{b^2} + \frac{n_3^2}{c^2}}. \tag{16.60}$$

A cavity such as the one shown in Fig. 16-27 will therefore resonate, sustaining standing waves for the frequencies given by Eq. (16.60).

In the case of a spherical or cylindrical cavity, the mathematical treatment is more complex, but again we find that only certain frequencies are allowed.

The results we have obtained for standing waves in cavities find many applications. In acoustics, for example, resonating cavities are used for sound analysis. Resonating cavities for electromagnetic waves have walls made of materials that are good electrical conductors so that the walls are the best possible reflectors. These cavities can sustain standing electromagnetic waves of definite frequencies with very little attenuation of the waves through energy loss by reflection. Thus the cavities serve as storage spaces for electromagnetic energy. The detailed theory of electromagnetic standing waves in cavities is slightly more complicated than our discussion here indicates because of the transverse character of the waves. But results such as Eq. (16.60) remain the same. Such cavities are used for frequency measurement or analysis (in the same manner as acoustical resonators), for frequency control in oscillating circuits, and for measuring the properties of the material filling the cavity.

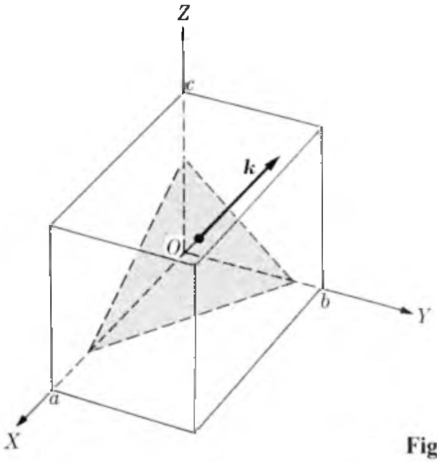


Figure 16-27

Example 16.8. Number of different modes of oscillation with a frequency equal to or smaller than ν in a cubical cavity of side a.

▼ If the cavity is cubical, $a=b=c$ in Eq. (16.60), and the possible frequencies are

$$\nu=\left(\frac{v}{2a}\right)\sqrt{n_1^2+n_2^2+n_3^2}$$

or

$$n_1^2+n_2^2+n_3^2=\frac{4\pi\nu^3a^3}{3v^3}. \tag{16.61}$$

In a coordinate system in which the coordinates are n_1, n_2, and n_3 (Fig. 16-28), Eq. (16.61) represents a sphere of radius $2\nu a/v$. Our problem is to find all the possible combinations of integers n_1, n_2,

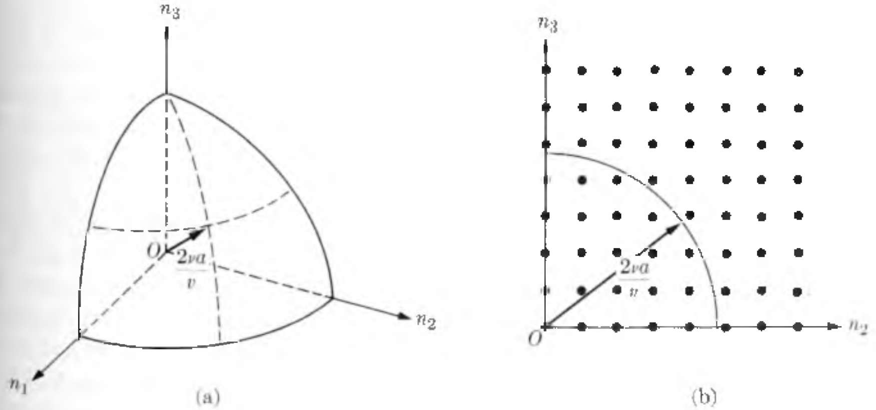


Figure 16-28

and n_3 that satisfy

$$n_1^2+n_2^2+n_3^2 \leqslant \frac{4\nu^2 a^2}{v^2}.$$

When the radius is very large, this number of possible combinations is equal to the volume of the octant of the sphere shown in Fig. 16-28a since to each set of integers n_1, n_2, and n_3, we may associate a cell of unit volume as indicated by the dots in Fig. 16-28b. Then the number of modes of oscillation of frequency equal to or smaller than ν is

$$N_\nu = \frac{1}{8}\frac{4\pi}{3}\left(\frac{2\nu a}{v}\right)^3 = \frac{4\pi\nu^3 a^3}{3v^3}.$$

Since a^3 is the volume of the cavity, the number of modes per unit volume is

$$n_{\nu,\mathrm{L}} = \frac{4\pi\nu^3}{3v^3}. \tag{16.62}$$

The subscript L is added because this result is valid only for longitudinal waves. If the waves are transverse, we have for each mode two independent and different states of polarization so that instead of Eq. (16.62) we must write

$$n_{\nu,\mathrm{T}} = \frac{8\pi\nu^3}{3v^3}. \tag{16.63}$$

Sometimes it is convenient to know the number of modes in a frequency range $d\nu$. This quantity may be written in the form $dn_\nu = g(\nu)\, d\nu$. Differentiating Eq. (16.62), we have

$$dn_{\nu,\mathrm{L}} = g_\mathrm{L}(\nu)\, d\nu = \frac{4\pi\nu^2}{v^3}\, d\nu; \tag{16.64}$$

for transverse waves, from Eq. (16.63) we get

$$dn_{\nu,\mathrm{T}} = g_\mathrm{T}(\nu)\, d\nu = \frac{8\pi\nu^2}{v^3}\, d\nu. \tag{16.65}$$

These results are very useful in several calculations, as, for example, in analyzing the modes of oscillation of radiation trapped in a cavity or in discussing atomic oscillations in solids. ▲

16.9 Wave Guides

The cavities discussed above allow only standing waves. There is also the possibility of producing traveling waves in certain enclosures called *wave guides*, which are long cavities open at both ends. Waves are fed in at one end and received at the other. We shall discuss in detail one simple type of guide for longitudinal waves: it consists of two parallel planes separated the distance a (Fig. 16-29). If a wave is set up inside the cavity at an angle with the planes as determined by the components k_1 and k_2 of the vector $\boldsymbol{k}$, parallel and perpendicular, respectively, to the planes, it will suffer successive reflections at both limiting surfaces and bounce back and forth between them. Since

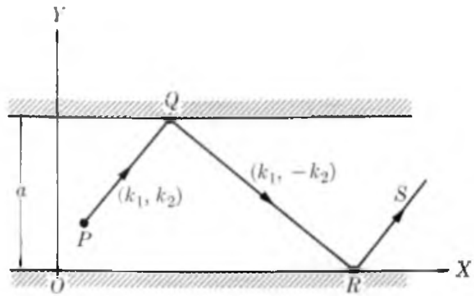


Fig. 16-29. Ray propagating between two parallel reflecting planes.

the space is not limited in the direction parallel to the planes (as happened with the cavities), the wave will keep progressing to the right. Let us choose the X-axis parallel to the reflecting planes and the Y-axis perpendicular to the planes so that the vector $\boldsymbol{k}$ is in the XY-plane. In Fig. 16-29 we have indicated the path of a particular ray. Along PQ the ray is characterized by the components k_1, k_2; from Q to R it is characterized by the components k_1, $-k_2$. From R on, it is again characterized by k_1 and k_2, and so on. We conclude then that in the space between the reflecting planes we have two sets of traveling waves, corresponding to k_1, k_2 and k_1, $-k_2$, respectively. (Remember that in the case of two-dimensional standing waves, such as those in a membrane, we had four waves because of the additional waves generated by reflections at the right and left ends.) These two waves interfere, giving rise to the resultant wave motion (as will be shown in Example 16.9) described by the expression

$$\xi = -2\xi_0 \sin k_2 y \cos(\omega t - k_1 x) \tag{16.66}$$

where

$$k_2 = \frac{n\pi}{a} \tag{16.67}$$

to satisfy the boundary condition $\xi = 0$ at $y = a$.

Equation (16.66) differs profoundly from our previous results with other kinds of standing waves in that the x-coordinate has not been separated from the time, but still appears in a term of the form $\cos(\omega t - k_1 x)$. This term corresponds to a wave traveling along the X-axis with a phase velocity

$$v_p = \frac{\omega}{k_1} = \left(\frac{k}{k_1}\right) v. \tag{16.68}$$

Since $k_1 \leqslant k$ because k_1 is a component of $\boldsymbol{k}$, Eq. (16.68) indicates that the phase velocity of the wave traveling along the cavity is larger than the phase velocity $v = \omega/k$ of the waves in free space. So for electromagnetic waves, the phase velocity would be greater than c. Now from $k^2 = k_1^2 + k_2^2$ and Eq. (16.67), we have that

$$k^2 = k_1^2 + \frac{n^2\pi^2}{a^2}.$$

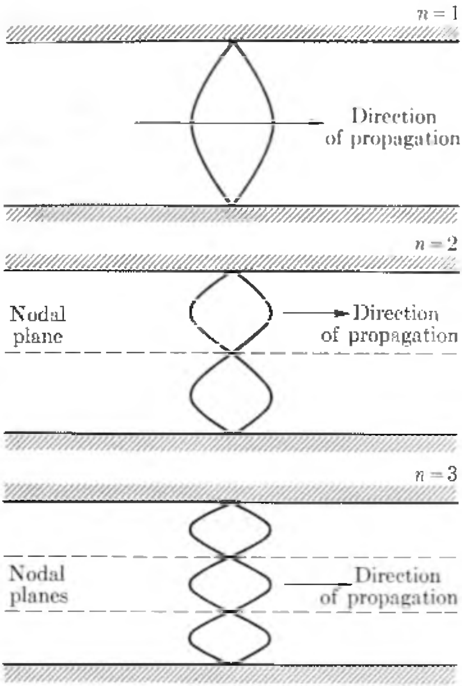


Fig. 16-30. First three modes of propagation of a wave between two parallel reflecting planes.

or

$$k_1 = \sqrt{k^2 - \frac{n^2\pi^2}{a^2}} = \sqrt{\frac{\omega^2}{v^2} - \frac{n^2\pi^2}{a^2}} \tag{16.69}$$

since $k = \omega/v$. The group velocity associated with the phase velocity given by Eq. (16.68) is, using Eqs. (10.55) and (16.69),

$$v_g = \frac{d\omega}{dk_1} = \frac{k_1}{\omega} v^2 = \left(\frac{k_1}{k}\right) v, \tag{16.70}$$

which is smaller than v since $k_1 \leqslant k$. Multiplying Eq. (16.68) by Eq. (16.70), we get $v_p v_g = v^2$; or for electromagnetic waves in vacuum ($v = c$), $v_p v_g = c^2$, a result found previously in Example 12.3 for a different situation. We see then that even if it is empty, an electromagnetic wave guide acts as a dispersive medium with an index of refraction less than one, and thus a phase velocity larger than c, but a group velocity smaller than c.

Equation (16.69) also indicates another important property of wave guides. Since k_1 must be a real number, in order for a wave to propagate along the wave guide, it is necessary that $\omega^2/v^2 \geqslant n^2\pi^2/a^2$, which yields

$$\omega \geqslant n\frac{\pi v}{a} \qquad \text{or} \qquad \nu \geqslant \frac{nv}{2a}. \tag{16.71}$$

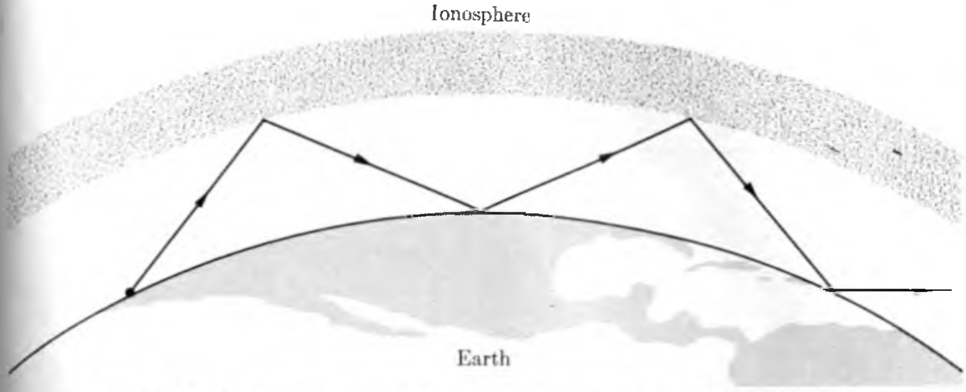


Fig. 16-31. The ionosphere and the earth act as a wave guide for radio waves.

In other words, only those waves with frequencies satisfying Eq. (16.71) are propagated along the guide. Each mode is determined by the value of n, and for each mode there is a *cutoff frequency*, equal to $nv/2a$, below which propagation is impossible. These wave guides act as frequency filters.

Although the wave propagates within the wave guide along the X-axis, the amplitude is modulated transversely in the Y-direction by the factor $\sin k_2 y$ in Eq. (16.66). The transverse variation of the amplitude is indicated in Fig. 16-30 for $n = 1, 2$, and 3. In practice, wave guides have either a rectangular or a circular cross section. The two shapes yield similar results with respect to phase velocity along the axis of the guide and cutoff frequency.

Although our discussion is valid for wave guides used for any kind of waves, the situation for electromagnetic waves has some peculiarities. Because of the transverse character of electromagnetic waves, for each $\boldsymbol{k}$ there are two possible modes, depending on the relative orientation of the electric field $\boldsymbol{\mathscr{E}}$ with respect to the sides of the wave guide. Electromagnetic wave guides are extensively used in the microwave region for the purpose of transmitting signals. These guides are made from materials that are excellent conductors.

It is interesting to note that the region between the earth's surface and the ionosphere, which is approximately 80 km above the earth, forms a wave guide that allows the propagation of radio waves around the curve of the earth, as shown in Fig. 16-31.

A simple example of parallel plane wave guides in the optical region is a pair of parallel mirrors, such as those found in some barber shops. Another type of optical wave guide consists of transparent fibers, called *optical fibers*, with a diameter of a few microns. These fibers are made of glass or quartz although other materials, such as nylon, are being tested. A ray entering at one end follows the axis of the fiber as a result of several reflections and emerges at the other end (Fig. 16-32). When the fibers are arranged in bundles, an image can be transmitted from one point to another.

Acoustical wave guides are also very common. The air ducts in the heating system of a house, for example, act as acoustical wave guides that are capable of transmitting

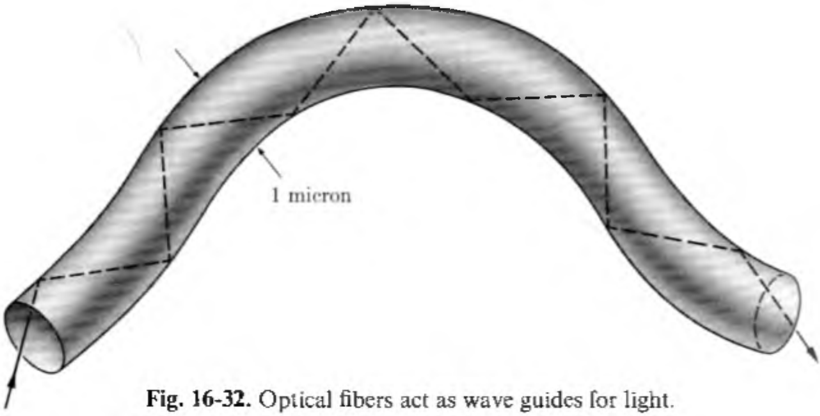


Fig. 16-32. Optical fibers act as wave guides for light.

the noises from the furnace or sounds from one room to another. The inner ear is essentially an acoustical wave guide.

Example 16.9. Derivation of Eqs. (16.66) and (16.67).

▼ The two waves that propagate along the guide of Fig. 16-29 give rise to a resultant wave described by the expression

$$\xi = \xi_0 \sin\left[\omega t - (k_1 x + k_2 y)\right] + \xi_0' \sin\left[\omega t - (k_1 x - k_2 y)\right], \tag{16.72}$$

that must be compared with Eq. (16.54). To determine ξ_0', we impose the condition that $\xi = 0$ at all points of the lower surface; that is, $y = 0$. Setting $y = 0$ in Eq. (16.72) yields

$$\xi = (\xi_0 + \xi_0') \sin(\omega t - k_1 x) = 0$$

so that $\xi_0 + \xi_0' = 0$ or $\xi_0' = -\xi_0$, a result to be expected from our experience with previous similar situations. Then Eq. (16.72) becomes

$$\xi = \xi_0 \{\sin\left[\omega t - (k_1 x + k_2 y)\right] - \sin\left[\omega t - (k_1 x - k_2 y)\right]\}.$$

Transforming the difference between the two sines into a product, we may write

$$\xi = -2\xi_0 \sin k_2 y \cos(\omega t - k_1 x). \tag{16.73}$$

By setting $y = 0$, we verify that our boundary condition at the lower plane is satisfied. The boundary condition at the upper plane ($y = a$) is also $\xi = 0$. This requires that $\sin k_2 a = 0$, resulting in $k_2 a = n\pi$ or $k_2 = n\pi/a$. However, there is no boundary condition for the X-coordinate. ▲

Example 16.10. Electromagnetic waves in a plane parallel wave guide.

▼ Electromagnetic waves in guides have certain peculiarities of their own that are due to their transverse character and to the boundary conditions at the surface of the conductor. These boundary conditions are (1) the electric field is normal, and (2) the magnetic field is tangent to the surface of a conductor. One possible solution of Maxwell's equations satisfying these condi-

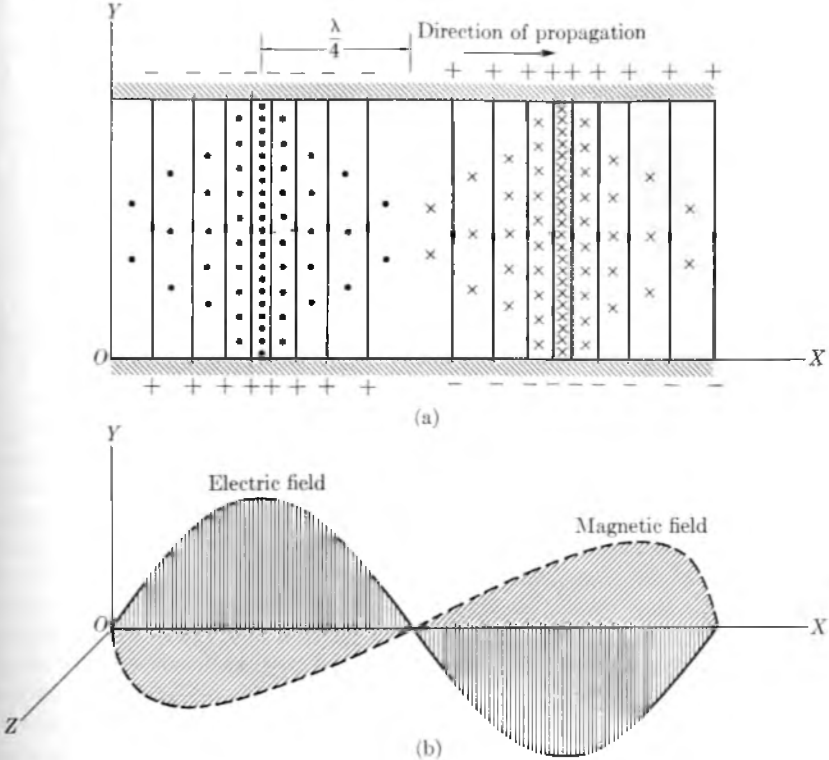


Fig. 16-33. (a) Electric lines of force (vertical lines) and magnetic lines of force (dots and crosses) in the XY-plane for an electromagnetic wave propagating parallel to two reflecting planes parallel to the XZ-plane. (b) Electric and magnetic fields in the wave depicted in part (a).

tions for a plane wave guide is that given by Eq. (11.12); that is, $\mathscr{E}_y = \mathscr{E}_0 \sin(\omega t - kx)$, $\mathscr{B}_z = \mathscr{B}_0 \sin(\omega t - kx)$ with $\mathscr{B}_0 = \mathscr{E}_0/c$. The lines of force of the electric field are indicated by lines in Fig. 16-33 and those of the magnetic field by dots and crosses. In this case the wave guide does not change the phase velocity of the wave, which propagates with the same phase velocity $c = \omega/k$, corresponding to propagation in free space; the wave guide limits only the wave front.

But Maxwell's equations admit other solutions that also satisfy our boundary conditions. One possible solution is

$$
\begin{aligned}
&\mathscr{E}_x = \mathscr{E}_y = 0,\\
&\mathscr{E}_z = \mathscr{E}_0 \sin k_2 y \cos(\omega t - k_1 x),\\
&\mathscr{B}_x = -\frac{k_2}{\omega}\mathscr{E}_0 \cos k_2 y \sin(\omega t - k_1 x), \qquad (16.74)\\
&\mathscr{B}_y = -\frac{k_1}{\omega}\mathscr{E}_0 \sin k_2 y \cos(\omega t - k_1 x),\\
&\mathscr{B}_z = 0.
\end{aligned}
$$

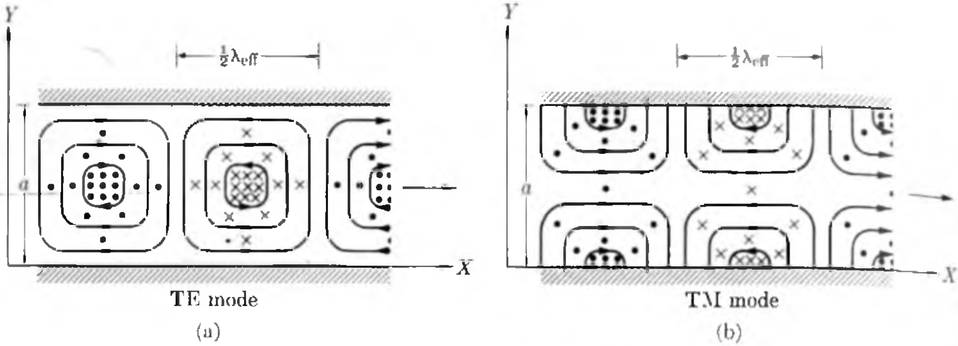


Fig. 16-34. Wave guides for electromagnetic waves. (a) Electric field perpendicular to the page, or TE mode. (b) Magnetic field perpendicular to the page, or TM mode.

This solution can be verified by direct substitution in Maxwell's equations. This solution is called the TE (*transverse electric*) solution because the electric field is transverse to the direction of propagation but the magnetic field has a component along the effective direction of propagation, or X-axis. The electric and magnetic fields are, however, perpendicular to each other. To satisfy the boundary conditions at both conducting planes, we must set $\mathscr{E}_z=0$ and $\mathscr{B}_y=0$ for $y=0$ and $y=a$. The first is automatically satisfied; the second requires that $\sin k_2a=0$ or $k_2a=n\pi$ so that the condition given by Eq. (16.67) is obtained. Figure 16-34a shows the lines of force for the lowest TE mode, $n=1$. The lines of force of the electric field are straight lines parallel to the planes (perpendicular to the page) and thus are indicated by dots or crosses; the lines of force of the magnetic field are the closed curves. Each pattern occupies one-half of the effective wavelength $2\pi/k_1$, and successive patterns have a phase difference of π. The patterns travel along the guide with the phase velocity $v_p=\omega/k_1$.

Another possible solution of Maxwell's equations is

$$
\begin{aligned}
&\mathscr{E}_x=-\frac{k_2}{k_1}\mathscr{E}_0 \sin k_2y \sin(\omega t-k_1x),\\
&\mathscr{E}_y=\mathscr{E}_0 \cos k_2y \cos(\omega t-k_1x),\\
&\mathscr{E}_z=0, \qquad\qquad (16.75)\\
&\mathscr{B}_x=\mathscr{B}_y=0,\\
&\mathscr{B}_z=\frac{\omega}{k_1c^2}\mathscr{E}_0 \cos k_2y \cos(\omega t-k_1x).
\end{aligned}
$$

This solution again can be verified by direct substitution in Maxwell's equations. This second solution, denoted TM (for *transverse magnetic*), is so called because the magnetic field is transverse to the direction of propagation. The electric field, however, has a component along the effective direction of propagation. Both fields remain perpendicular to each other. To satisfy the boundary conditions at the conducting planes, we must make $\mathscr{E}_x=0$ at $y=0$ and $y=a$. The first is automatically satisfied, and the second once more requires that $\sin k_2a=0$ or $k_2a=n\pi$ so that again the condition given by Eq. (16.67) is obtained. Therefore both modes have the same cutoff frequency.

Figure 16-34b shows the lines of force for the lowest TM mode, $n=1$. However, the lines of force of the *magnetic* field are now straight lines parallel to the planes (perpendicular to the page) and are indicated by dots or crosses; the lines of force of the electric field correspond to the patterns shown. As in the TE case, each pattern occupies one-half of the effective wavelength $2\pi/k_1$, and the patterns travel along the guide with the phase velocity $v_p=\omega/k_1$. ▲

Problems

16.1 Two slits, separated a distance of 10^{-3} m, are illuminated with red light of wavelength 6.5×10^{-7} m. The interference fringes are observed on a screen placed 1 m from the slits. (a) Find the distance between two bright fringes and between two dark fringes. (b) Determine the distance of the third dark fringe and the fifth bright fringe from the central fringe.

16.2 By means of a *Fresnel biprism* (Fig. 16-4), interference fringes are produced on a screen 0.80 m away from the biprism, using light of wavelength equal to 6.0×10^{-7} m. Find the distance between the two images produced by the biprism if 21 fringes cover a distance of 2.4×10^{-3} m on the screen.

16.3 Figure 16-35 shows an arrangement, called *Lloyd's mirror*, which produces interference patterns. The coherent light sources are the source S_1 and its image, S_2, which is due to reflection on the upper surface of the glass plate. Therefore the interfering rays are those coming directly from the source and those reflected from the glass. (a) What would you conclude about the phase change by reflection if the fringe corresponding to zero path difference (i.e., $\delta=0$) is (i) bright, and (ii) dark? (b) In the actual experiment, result (ii) is obtained. Explain why this result was to be expected in view of the discussion of Section 14.2.

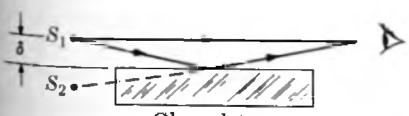


Figure 16-35

16.4 In Lloyd's mirror, the source slit S_1 and its virtual image S_2 lie in a plane 0.20 m behind the left edge of the mirror (see Fig. 16-35). The mirror is 0.3 m long, and a screen is placed at the right edge. Calculate the distance from this edge to the first light maximum if the perpendicular distance from S_1 to the mirror's plane is 2×10^{-3} m and if $\lambda=7.2\times10^{-7}$ m.

16.5 Discuss the interference pattern on a screen when the sources S_1 and S_2, separated the small distance a, are placed along a line perpendicular to the screen (Fig. 16-36a). Experimentally the two sources could be the two images of a light source produced by reflection in the two faces of a thin mica sheet (Fig. 16-36b). This arrangement is called *Pohl's interferometer.*

16.6 Two synchronized sources of sound waves send out waves of equal intensity at a frequency of 680 Hz. The sources are 0.75 m apart. The velocity of sound is 340 m s^{-1}. Find the posi-

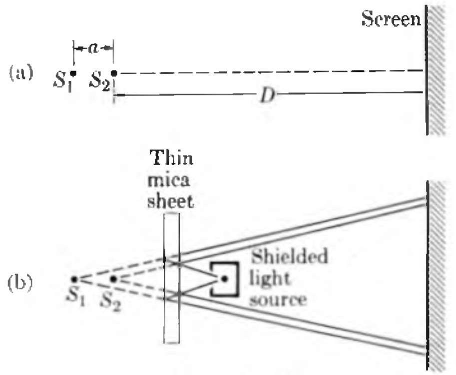


Figure 16-36

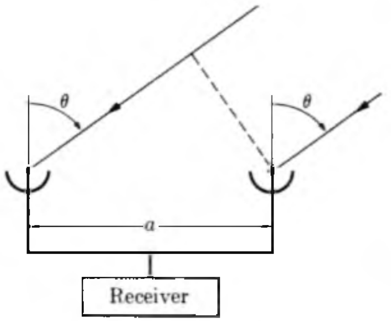


Fig. 16-37. Radio interferometer arrangement.

Fig. 16-38. Two-element radio interferometer at Green Bank, West Virginia. (Photograph courtesy National Radio Astronomy Observatory.)

tions of minimum intensity: (a) on a line that passes through the sources, (b) in a plane that is the perpendicular bisector of the line between the sources, and (c) in a plane that contains the two sources. (d) Is the intensity zero at any of the minima?

16.7 An interferometric arrangement used in radio astronomy consists of two radio telescopes separated a certain distance. The antennas of these telescopes can be oriented in different directions but are always kept parallel. The signals received by the antennas are transmitted to a receiving station where they are mixed (Fig. 16-37). (a) Show that the directions of incidence for which the resultant signal is maximum are given by Eq. (16.11). (*Hint:* Note that the situation is just the reverse of that for two sources, illustrated in Fig. 16-3.) (b) Make a polar plot of the intensity of the signal as a function of the angle θ.

16.8 Figure 16-38 shows a radio interferometer at Green Bank, West Virginia that operates at a wavelength of 0.11 m. The distance a between the two radio telescopes can be adjusted up to 2.7×10^3 m. Find the angle subtended by the central intensity maximum at maximum separation of the two telescopes.

16.9 Suppose that we have, instead of two parallel slits as in a Young's experiment, three parallel slits equally spaced a distance a. Discuss the intensity distribution of the interference pattern observed on a distant screen.

16.10 Find the intensities of the secondary maxima in the source arrangement of Fig. 16-12 relative to the principal maxima. (*Hint:* The first secondary maximum may be shown to occur at $\theta \sim 48°$.)

16.11 (a) Find the spacing between the sources in the array of Fig. 16-12 to produce an "end-fire" pattern; i.e., one with principal maxima at $\theta = \pm\pi/2$. (b) Determine the position of the secondary maxima. (c) Make a plot of the angular distribution of the intensity.

16.12 The first multiple radio interferometer, built in 1951 in Australia by Prof. W. N. Christiansen and shown in Fig. 16-39, consists of 32 antennas, 7 m apart, with their corresponding parabolic reflectors. The system is tuned to a wavelength of 0.21 m. The signals received by the antennas are superposed at the observing station to give a resultant signal. The system is thus equivalent to 32 equally spaced sources. Find (a) the angular width of the central maximum, and (b) the angular separation between successive principal maxima.

16.13 Two rectangular pieces of plane glass are laid one upon the other. A thin strip of paper is placed between them at one edge so that a very thin wedge of air is formed. The plates are illuminated by a beam of sodium light ($\lambda = 5.9 \times 10^{-7}$ m) at normal incidence. Ten inter-

Fig. 16-39. Grating radio interferometer at the University of Sydney, Australia. (Photograph courtesy Prof. W. N. Christiansen.)

ference fringes are formed per centimeter length of wedge. Find the angle of the wedge.

16.14 A square piece of cellophane film with index of refraction 1.5 has a wedge-shaped section so that its thickness at two opposite sides is a_1 and a_2 (Fig. 16-40). If the film is illuminated with monochromatic light of wavelength 6.0×10^{-7} m at normal incidence, the number of fringes appearing by reflection on the film is 10. What is the difference $a_2 - a_1$?

16.15 Light of wavelength 5.0×10^{-7} m is incident perpendicularly on a film that has an index of refraction equal to 1.4 and is 10^{-6} m thick. Part of the light enters the film and is reflected back at the second face. (a) How many wavelengths are contained along the path of this light in the film from the point of incidence to the point of emergence? (b) What is the phase difference between these waves as they leave the film and as they enter it? (c) Repeat the problem for light whose angle of incidence is 30°.

16.16 Two glass plates having a length of 5×10^{-2} m are placed in contact at one end and separated at the other by a thin paper sheet, thus forming an air prism. When the prism is illuminated by light of wavelength 5.9×10^{-7} m at normal incidence, 42 dark fringes are observed. Find the thickness of the paper sheet.

Figure 16-40

16.17 Newton's rings are observed with a plano-convex lens resting on a plane glass surface (see Fig. 16-14). The radius of curvature of the lens is 10 m. (a) Find the radii of the dark interference rings of the various orders observed by reflection under nearly perpendicular incidence, using light of wavelength 4.8×10^{-7} m. (b) How many rings are seen if the diameter of the lens is 4×10^{-2} m.

16.18 The radius of curvature of the convex surface of a plano-convex lens is 1.20 m. The lens is placed on a plane glass plate with the convex side down, and illuminated from above with red light of wavelength 6.5×10^{-7} m. Find the diameter of the third bright ring of the interference pattern.

16.19 A copper wire having a radius of 10^{-3} m and a length of 1 m is fixed at both ends and is subject to a tension of 10^4 N. Find (a) the fundamental frequency and the first two overtones, and (b) the corresponding wavelengths. (c) Plot the vibrational state of the wire in each case. (d) Write the equation describing the standing waves for each frequency. (The density of copper is 8.92×10^3 kg/m^{-3}.)

16.20 (a) How is the fundamental frequency of a string changed if one doubles (i) its tension, (ii) its mass per unit length, (iii) its radius, and (iv) its length? (b) Repeat the problem for a case in which the quantities listed are halved.

16.21 A tube whose length is 0.60 m is (a) open at both ends, and (b) closed at one end and open at the other. Find its fundamental frequency and the first overtone if the temperature of the air is 300 K. Plot the amplitude distribution along the tube corresponding to (c) the fundamental frequency and (d) the first overtone.

16.22 Estimate the percentage change in the fundamental frequency of an air column per degree change in temperature at a temperature

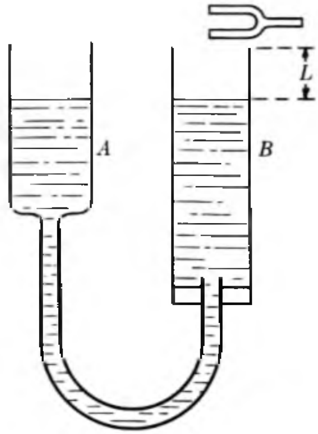


Figure 16-41

of 300 K. (See Example 10.7.)

16.23 A string vibrating with a frequency of 256 Hz is in resonance with a tuning fork. Determine the frequency of the beats produced if the tension of the string is increased by 20 percent.

16.24 A tuning fork with a frequency of 256 Hz is placed in front of the open end of a tube as shown in Fig. 16-41. The length of the air column can be changed by displacing the level of the water surface, moving container A up or down. (a) Find the lengths, L, of the first three air columns that are in resonance with the tuning fork. (b) Make a sketch in each case showing the amplitude distribution along the tube and the position of the nodes and antinodes. Assume a temperature of 300 K.

16.25 Two surface waves, $A \sin k(x-vt)$ and $A \sin k(y-vt)$, propagate along a membrane. (a) Discuss the resulting motion; show that these waves are equivalent to a modulated wave propagating in a direction making an angle of 45° with the X-axis and with a phase velocity equal to $\sqrt{2}v$. (b) Verify that the wavelength is reduced by the factor $\sqrt{2}$. (c) Show that the amplitude is zero over the lines

$$x-y=(2n+1)\pi/k.$$

16.26 (a) Show that for a square membrane of side a, if $\nu_0 = v/2a$ is the fundamental frequency, the successive frequencies are $\nu=\sqrt{2}\nu_0, 2\nu_0, \sqrt{5}\nu_0, 2\sqrt{2}\nu_0, 3\nu_0, \sqrt{10}\nu_0, \sqrt{13}\nu_0, \ldots$. (b) Determine the number of different combinations of n_1 and n_2 needed to obtain the fundamental and successive modes of vibration. The number of different combinations gives the *degeneracy* of the vibrating mode.

16.27 Repeat the preceding problem for a cubical cavity of side a.

16.28 Estimate the number of transverse vibrational modes per unit volume, in the frequency range between 1.0×10^{15} Hz and 1.2×10^{15} Hz, for electromagnetic radiation trapped in a cavity.

CHALLENGING PROBLEMS

16.29 The lowest frequency standing wave that can be set up in a certain piano string has a frequency of 512 Hz. The string is 0.80 m long. (a) What is the speed of propagation of waves in the string? (b) Sketch the standing wave patterns of the next two possible higher frequency standing waves and specify their frequencies. [AP-B: 1969]

16.30 Two speakers S_1 and S_2 are transmitting sound waves in phase and of the same frequency f. A microphone is used to detect the resultant sound by moving it along the line PP'. An intensity pattern for the sound is shown in Fig. 16-42. The speed of the sound is 340 m per second. (a) How much farther from point M is speaker S_1 than speaker S_2? (b) Find the wavelength and frequency of the sound waves. (c) The frequency of the sound waves is gradually lowered until a minimum of intensity is observed at point M. What is the frequency of the sound waves now? [AP-B: 1969]

16.31 Two glass plates are in contact at one

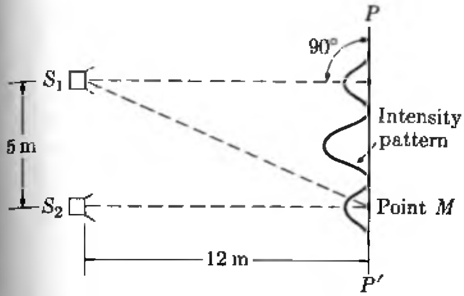


Figure 16-42

edge and are separated by a piece of aluminum foil at the opposite edge as shown in Fig. 16-43a. Light of wavelength 6,000 Å is approximately normally incident from above, illuminates the entire plate area, and is reflected from surfaces A and B. As an observer moves from left to right across the plates, the observer counts a total of 30 bright interference bands. The center of bright band 30 coincides with the edge of the foil. (a) What is the thickness of the foil? (b) Is the region near the line of contact of the plates a dark band or a bright band? Explain. The top plate is broken and the remaining piece is only two-thirds as long as before as shown in Fig. 16-43b. The same piece of foil is used again. (c) Describe any change in the number and position of the interference bands now observed. [AP-B; 1970]

16.32 A rectangular, transparent glass plate having an index of refraction of 1.5 is coated with a thin film of transparent material designed to reduce the intensity of the reflected light. Yellow light of a single frequency is directed at normal incidence onto the film as shown

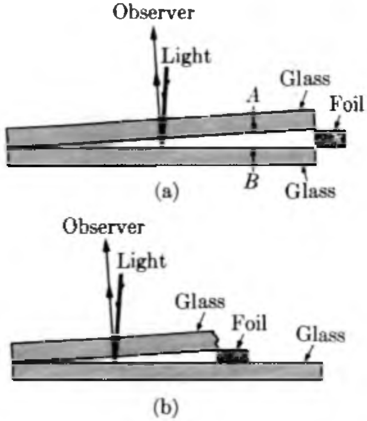


Figure 16-43

in Fig. 16-44. (a) If the index of refraction of the film is 1.3, determine the speed of light in the film. (b) If the frequency of yellow light in air is 5.0×10^{14} Hz, what is the frequency of yellow light in the film? (c) Determine the wavelength of yellow light in the film. (d) Determine the smallest film thickness that will reduce the intensity of the reflected yellow light to a minimum. Explain your reasoning. [AP-B; 1971]

16.33 Consider an organ pipe 8 ft long that is open at both ends. Assume the speed of sound is 1,200 ft per second. (a) Determine the lowest two frequencies of sound produced by the pipe. (b) If the temperature drops, what will be the effect on the frequencies? Explain the reason for your answer. [AP-B; 1973]

16.34 Two loudspeakers, S_1 and S_2, a distance d apart as shown in Fig. 16-45, vibrate in phase and emit sound waves of equal amplitude and

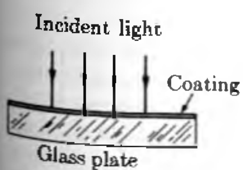


Figure 16-44

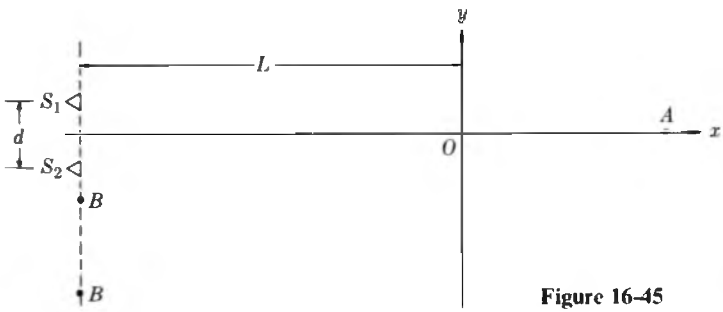


Figure 16-45

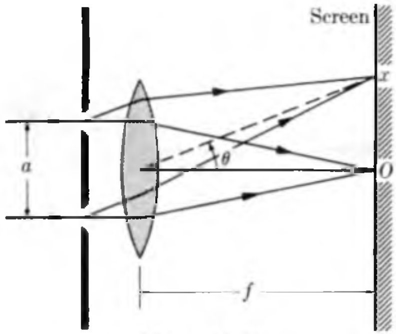


Figure 16-46

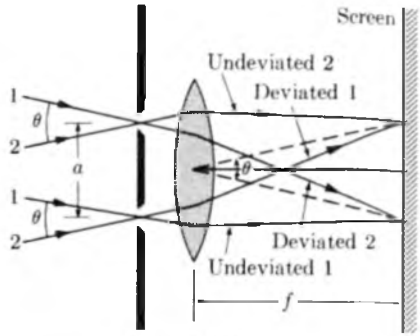


Figure 16-47

wavelength λ. Assume $d \ll L$. (a) Describe how sound intensity I varies as a function of position x along the line segment OA. Sketch a graph of this function on a set of axes. (b) Assume $\lambda \ll d$. On a set of axes, sketch a graph of the sound intensity I as a function of position y along the y-axis. (c) Assume that $d = 2$ m and that the speed of sound is 360 m per second. Find the lowest speaker frequency that will yield the minimum sound intensity along the line BB'. [AP-B; 1977]

16.35 (a) Show that if a source is placed at a distance d from a Fresnel biprism having an index n and a very small angle A, the distance between the two images is

$$a = 2(n-1)Ad$$

where A is expressed in radians. (b) Calculate the spacing of the fringes of green light of wavelength 5×10^{-7} m produced by a source placed 5×10^{-2} m from a biprism, having an index equal to 1.5 and an angle of 2'. The screen is 1 m from the biprism.

16.36 One technique for observing an interference pattern produced by two slits is to illuminate them with parallel rays of light, place a convergent lens behind the plane of the slits, and observe the interference pattern on a screen placed at the focal plane of the lens (Fig. 16-46). Show that the position of the bright fringes relative to the central fringe is given by

$$x = n\left(\frac{f\lambda}{a}\right)$$

and the dark fringes correspond to

$$x = (2n+1)\left(\frac{f\lambda}{2a}\right)$$

where n is an integer, f the focal length of the lens, and a is the separation of the slits.

16.37 Two parallel beams of monochromatic light of the same wavelength, which form a small angle θ with each other, fall on two slits separated the distance a, in front of a convergent lens (Fig. 16-47). Because of the angular displacement of the beam, the two sets of fringes observed on a screen placed at the focal plane of the lens (see the preceding problem) are not coincident. (a) Show that if

$$\theta = \frac{\lambda}{2a},$$

the bright fringes of one beam fall on the dark fringes of the other beam, and the interference pattern disappears. (In 1868 Fizeau proposed this method for measuring the angular separation of two distant objects by varying the distance a until the interference pattern disappears. It has been used, for example, to measure the angular separation of stars by placing a screen with two slits in front of the objective of a telescope and varying the separation of the slits until the diffraction pattern disappears.) (b) Find the minimum angular separation that can be detected with the Mount Wilson refracting telescope, whose objective has a diameter

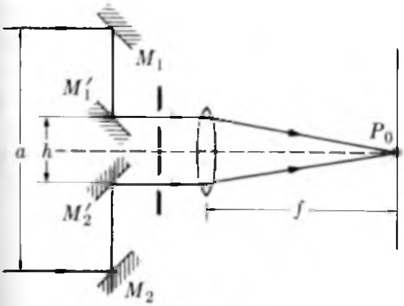


Figure 16-48

of 2.54 m. Assume that the wavelength is 5.7×10^{-7} m.

16.38 To increase the resolving power of a telescope, Michelson in 1921 built an interferometric arrangement (such as that shown in Fig. 16-48) in which the M's are four mirrors placed in front of the objective of a telescope. Therefore what is observed through the telescope is the interference pattern of rays received by mirrors M_1 and M_2. This arrangement is essentially the Fizeau arrangement described in the previous problem. It can be shown that when one observes an extended circular source of light, the interference pattern disappears if the angle subtended by the source is related to the separation of the mirrors by the relation

$$\theta = \frac{1.22\lambda}{a}.$$

(See Section 17.3.) By using $a = 3.073$ m and a wavelength of 5.75×10^{-7} m, Michelson found that the fringes corresponding to the star α-Orionis (Betelgeuse) disappeared. (a) Show that the angular diameter of this star is 0.047″. This method provided the first measurement of a star's diameter. (b) What would the diameter of the objective of a telescope have to be to produce the same resolving power? (c) If the distance to the star is 1.80×10^{18} m, find the linear diameter of α-Orionis. (d) Compare this value with the diameter of the sun and of the earth's orbit.

16.39 Discuss the angular distribution of intensity for (a) three and (b) five identical sources of waves equally spaced the distance a along a straight line. Assume that $a = \lambda/2$.

16.40 Considering the interference of wave motions produced by N sources as discussed in Section 16.3, show that the initial phase of the resulting motion is given by

$$\delta_N = \tfrac{1}{2}(N-1)\delta$$

where δ is given by Eq. (16.12). Note that δ_N is the angle that the vector OP makes with the X-axis in Fig. 16-9.

16.41 Using the result of the preceding problem and the law of vector addition, prove the following trigonometric relations:

(a) $1 + \cos\delta + \cos 2\delta + \cdots + \cos(N-1)\delta$

$$= \frac{\sin \frac{1}{2}N\delta}{\sin \frac{1}{2}\delta} \cos \tfrac{1}{2}(N-1)\delta.$$

(b) $\sin\delta + \sin 2\delta + \cdots + \sin(N-1)\delta$

$$= \frac{\sin \frac{1}{2}N\delta}{\sin \frac{1}{2}\delta} \sin \tfrac{1}{2}(N-1)\delta.$$

(*Hint:* Note that in Fig. 16-9 the components of the resultant vector along the X- and Y-axes are equal to the sum of the components of the individual vectors.)

16.42 A thin film having a thickness of 2.4×10^{-6} m and an index of refraction of 1.4 is illuminated with monochromatic light of wavelength 6.2×10^{-7} m. Find the smallest angles of incidence for which there is maximum (a) constructive and (b) destructive interference by reflection. Repeat the problem for light that is transmitted.

16.43 Show that if a glass plate (index n_g) is covered by a thin film having an index $n_a = \sqrt{n_g}$ (see Problem 14.23) and having a thickness equal to one-quarter of the wavelength of light in the film, complete destructive interference between light reflected at both surfaces results, for normal incidence. This method is effective for decreasing the intensity of reflection from lenses and plates in optical instruments. Such a thin film is called an *antireflection coating*.

16.44 Show that if the thin film of the previous problem has an index much larger than that of

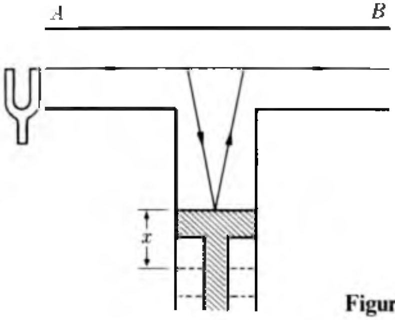


Figure 16-49

glass and a thickness of one-quarter of the wavelength of light (in the film), the intensity of reflected light for that particular wavelength is increased.

16.45 Prove that if R is the radius of the convex side of a plano-convex lens used to produce Newton's rings, the radii of the bright rings are given by

$$r^2 = N\lambda R$$

and the radii of the dark rings by

$$r^2 = (2N+1)\left(\frac{\lambda R}{2}\right)$$

where N is a positive integer. The index of refraction of air has been taken as one.

16.46 A T-tube has one of the branches closed by a movable piston, as shown in Fig. 16-49. A tuning fork is placed at one of the open ends, A. Show that the separation between successive positions of the piston for which maximum intensity of sound is received at the other open end, B, is $x = \frac{1}{2}\lambda$.

16.47 A wave guide consists of a long tube of rectangular cross section with sides a and b. Show that the resultant wave is described by

$$\xi = 4\xi_0 \sin k_2 y \sin k_3 z \cos(\omega t - k_1 x)$$

and that the only frequencies transmitted along the wave guide are those satisfying $\nu \geqslant \frac{1}{2}v\sqrt{n_1^2/a^2 + n_2^2/b^2}$ where n_1 and n_2 are integers. Discuss the nodal planes in the wave guide for $n_1 = 2$ and $n_2 = 3$.

16.48 Given the wave equation in two dimensions

$$\frac{\partial^2 \xi}{\partial x^2} + \frac{\partial^2 \xi}{\partial y^2} = \left(\frac{1}{v^2}\right)\frac{\partial^2 \xi}{\partial t^2}.$$

(a) Try a solution corresponding to standing waves of the form $\xi = f(x, y) \sin \omega t$. (b) Show that $f(x, y)$ satisfies the differential equation

$$\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + k^2 f = 0$$

where $k = \omega/v$. (c) Determine the constants k_1 and k_2 in order that

$$f(x, y) = A \sin k_1 x \sin k_2 y$$

be a solution of the preceding equation. (d) Compare your results with those of Section 16.7.

16.49 Extend the discussion of the preceding problem to the case of the three-dimensional wave equation. In this case the trial solution is

$$f(x, y, z) = A \sin k_1 x \sin k_2 y \sin k_3 z.$$

16.50 Discuss graphically the solutions of the transcendental equation $x \tan x = C$ where C is a constant. You can proceed in either one of two ways. Draw the curves $y = \tan x$ and $y = C/x$ and determine their intersections, or draw the curves $y = x$ and $y = C \cot x$ and determine their intersections. In any case, with the exception of the first intersection, all the other intersections fall closely after $x = n\pi$ where n is an integer. Apply your results to the discussion of Example 16.6.

CHAPTER SEVENTEEN

DIFFRACTION

17.1 Introduction

Another phenomenon characteristic of wave motion is known by the generic name of *diffraction*. Diffraction is noticeable when a wave is distorted by an obstacle that has dimensions comparable to the wavelength of the wave. The obstacle may be a screen with a small opening or slit that allows a small portion of the incident wave front to pass. The obstacle may also be a small object, such as a wire or a disk, that blocks the passage of a small portion of the wave front. Whereas interference is the result of individual sources interacting with each other, diffraction is the interference of a finite wave with itself.

If a stream of particles falls on a screen with a small opening, only those particles falling on the opening will be transmitted and allowed to continue their motion undisturbed (Fig. 17-1). The other particles either will be stopped or will bounce back. Conversely, if an object is placed in a stream of particles, the object will block those particles falling on it, but the remaining particles will continue their motion undisturbed. However, we know from common experience, especially for the case of sound waves and surface waves in water, that waves behave in a different way, and that they extend around the obstacles interposed in their path as illustrated in Fig. 17-2. This effect becomes more and more noticeable as the dimensions of the slits or the size of the obstacles approach the wavelength of the waves. One cannot usually observe diffraction of light with the naked eye since most of the objects interposed in a beam of light are much larger than the wavelength of light waves, whose magnitude is of the order of 5×10^{-7} m.

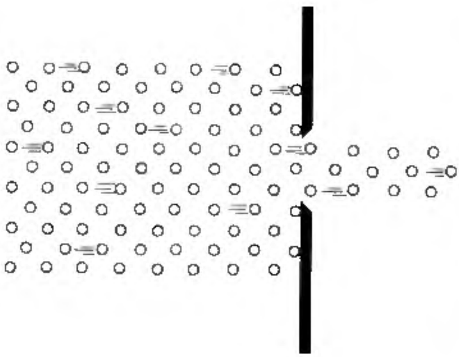


Fig. 17-1. Behavior of a stream of particles impinging on a screen with a small opening.

Fig. 17-2. Behavior of a wave impinging on a screen with a small opening. (*Ripple Tank Studies of Wave Motion*, by permission of W. Llowarch. The Clarendon Press, Oxford, England.)

In this chapter we shall discuss diffraction produced by certain apertures and screens of simple geometry under two special circumstances. In *Fraunhofer diffraction* we assume that the incident rays are parallel, and that we observe the diffraction pattern at a distance sufficiently large so that we effectively receive only diffracted rays that are parallel. This condition can also be accomplished by using a lens that focuses rays diffracted in the same direction to the same position on a screen. In *Fresnel diffraction*, either the incident rays originate from a point source, or the diffracted rays are observed at a particular point of space, or both.

Closely related to diffraction is *scattering*, which takes place when the obstacles interposed in the wave may be considered as sources of new waves. We discussed the scattering of electromagnetic waves by individual electrons in Chapter 12 when we described the Compton effect. In this chapter we shall briefly consider scattering from a more general point of view.

17.2 Fraunhofer Diffraction by a Rectangular Aperture

As our first example, consider diffraction of a plane wave by a rectangular aperture or slit, very narrow and very long, so that at the beginning we may ignore the effects at the ends (Fig. 17-3a). We also assume that the incident waves are normal to the plane of the slit. This simplifies the mathematics without changing the physical situation. According to Huygens's principle, when the incident wave falls on the slit, all points of the plane of the slit become secondary sources of waves, emitting new waves, called, in our case, *diffracted* waves, whose resultant amplitude is computed by using Eq. (13.2). The diffracted waves may be observed on a screen placed at a very large distance from the slit. At different angles θ with respect to the direction of incidence (Fig. 17-3b), we find that their intensity is zero. These angles are given by the relation

$$b \sin \theta = n\lambda, \qquad n \neq 0 \tag{17.1}$$

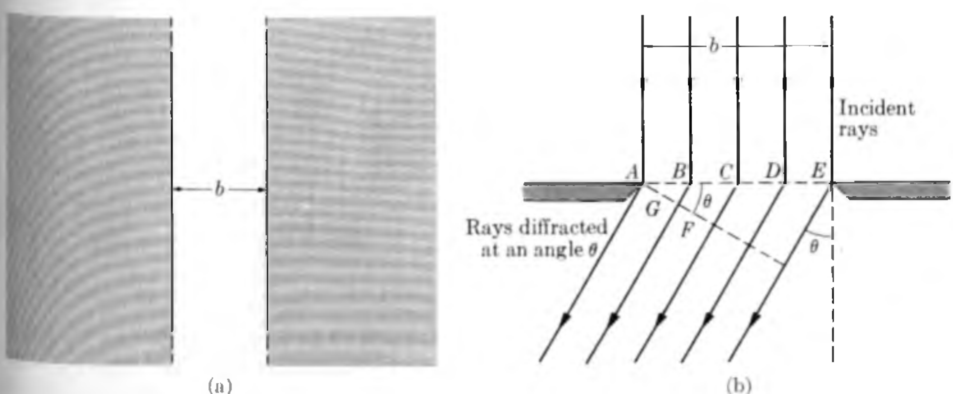


Fig. 17-3. Diffraction by a long narrow slit.

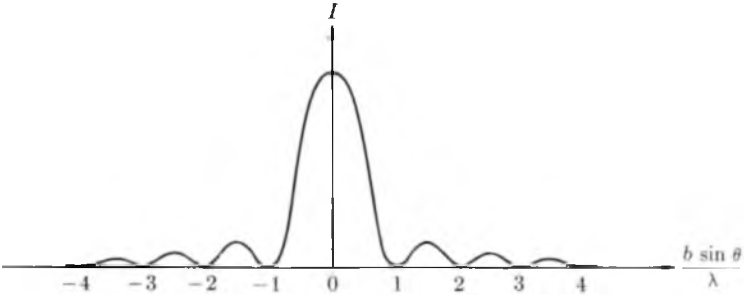


Fig. 17-4. Intensity distribution of the diffraction pattern of a long narrow slit.

Fig. 17-5. Fraunhofer diffraction pattern produced by a long narrow slit.

where n is a positive or negative integer, b is the width of the slit, and λ the wavelength of the incident waves. The value $n=0$ is excluded because it corresponds to observation along the direction of incidence, which obviously implies a maximum of illumination.

From Eq. (17.1), we have

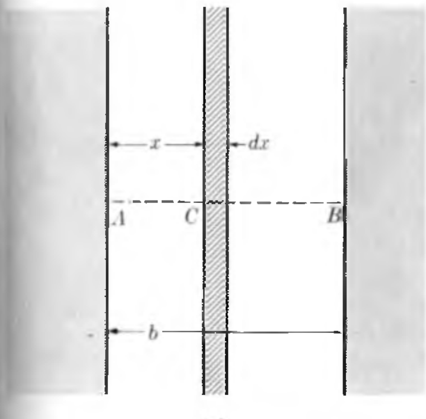
$$\sin\theta = \frac{n\lambda}{b} \tag{17.2}$$

so that the intensity is zero for $\sin\theta = \pm\lambda/b,\ \pm 2\lambda/b,\ \pm 3\lambda/b, \ldots$.

To justify Eq. (17.1), we recall from Eq. (16.8) that when the difference in path length for two rays is $r_1 - r_2 = \text{odd integer} \times (\lambda/2)$, destructive interference results. From Fig. 17-3b we see that for rays coming from A and the midpoint C we have $r_1 - r_2 = CF = \frac{1}{2}b\sin\theta = n(\lambda/2)$. Thus for $n = 1, 3, 5, \ldots$, these two rays, as well as all other pairs of rays originating at points separated by $\frac{1}{2}b$, interfere destructively; and no intensity is observed at the angle θ. For even n, consider points A and B separated by $b/4$. Then

$$r_1 - r_2 = BG = \tfrac{1}{4}b\sin\theta = \left(\frac{n}{2}\right)\left(\frac{\lambda}{2}\right).$$

So when $n/2$ is an odd integer, or $n = 2, 6, 10, \ldots$, these two rays, as well as all other pairs of rays originating at points separated by $b/4$, interfere destructively; and again no intensity is observed in the direction corresponding to the angle θ. The procedure can be extended until all integers are included. For $\theta = 0$, however, there



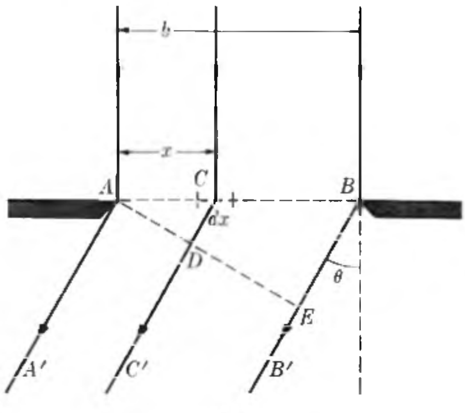


Figure 17-6

is no phase difference for the rays coming from different points; and the interference is constructive, resulting in a pronounced maximum.

Between each zero of intensity given by Eq. (17.1) there is a maximum; but these maxima gradually decrease in intensity, a situation different from that for interference. The intensity of the diffracted waves as a function of θ is represented in Fig. 17-4. Note that the central maximum has twice the width of the others. Figure 17-5 shows the actual diffraction pattern of a long narrow slit.

It is easy, as well as instructive, to compute the intensity distribution shown in Fig. 17-4. If we divide the slit into very narrow strips of width dx as shown in Fig. 17-6a, we may consider each strip as a secondary source of waves of very small (and identical) amplitude $d\xi_0$. When we consider the rays emitted in the direction corresponding to the angle θ (Fig. 17-6b), the phase difference between rays CC' and AA', taken as reference, is

$$\delta=\frac{2\pi}{\lambda}CD=\frac{2\pi x\sin\theta}{\lambda}. \qquad (17.3)$$

and therefore increases gradually with x. To obtain the amplitude in the direction corresponding to the angle θ, we must plot the rotating vectors corresponding to the waves from all the strips from A to B. (Recall this technique was used in Section 16.3.) Since all the waves are of infinitesimal amplitude and since the phase angle δ increases proportionately with x, the vectors fall on an arc of circle OP whose center is at C and whose radius is ρ (Fig. 17-7). The resultant amplitude A is the chord OP. The slope at any point of the arc from O to P is just the angle δ given by Eq. (17.3). At P, which corresponds to $x=b$, the inclination of the tangent is the angle

$$\alpha=\frac{2\pi b\sin\theta}{\lambda}. \qquad (17.4)$$

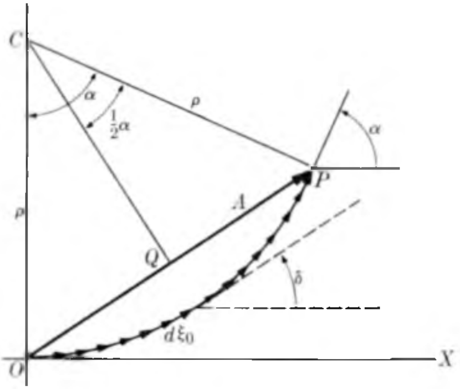


Figure 17-7

This is also the angle formed by the two radii CO and CP. Therefore the resultant amplitude is

$$A = \text{chord } OP = 2QP$$
$$= 2\rho \sin \tfrac{1}{2}\alpha = 2\rho \sin\left(\frac{\pi b \sin\theta}{\lambda}\right). \tag{17.5}$$

For observation in a direction perpendicular to the slit (i.e., $\theta = 0°$), all vectors $d\xi_0$ are parallel; and their resultant is just the sum of their lengths, which is equal to the length of the arc from O to P. Designating the resultant amplitude for normal observation by A_0, we then have

$$A_0 = \text{arc } OP = \rho\alpha = \rho\left(\frac{2\pi b \sin\theta}{\lambda}\right). \tag{17.6}$$

Dividing Eq. (17.5) by Eq. (17.6), we get

$$A = A_0\left[\frac{\sin(\pi b \sin\theta/\lambda)}{\pi b \sin\theta/\lambda}\right]; \tag{17.7}$$

and since the intensities are proportional to the squares of the amplitudes, we obtain

$$I = I_0\left[\frac{\sin(\pi b \sin\theta/\lambda)}{\pi b \sin\theta/\lambda}\right]^2 = I_0\left(\frac{\sin u}{u}\right)^2 \tag{17.8}$$

where $u = \pi b \sin\theta/\lambda$. We verify then that the zeros of the intensity occur when $u = n\pi$, or $b \sin\theta = n\lambda$, in agreement with Eq. (17.1), except for $n = 0$ because then $(\sin u/u)_{u=0} = 1$. To obtain the maxima of intensity, we find the values of u satisfying $dI/du = 0$ (see Example 17.1). But because these maxima of intensity correspond to successively larger values of u, they grow smaller and smaller, resulting in the pattern that was shown in Fig. 17-4. For λ very small compared with b, the first zeros of intensity on

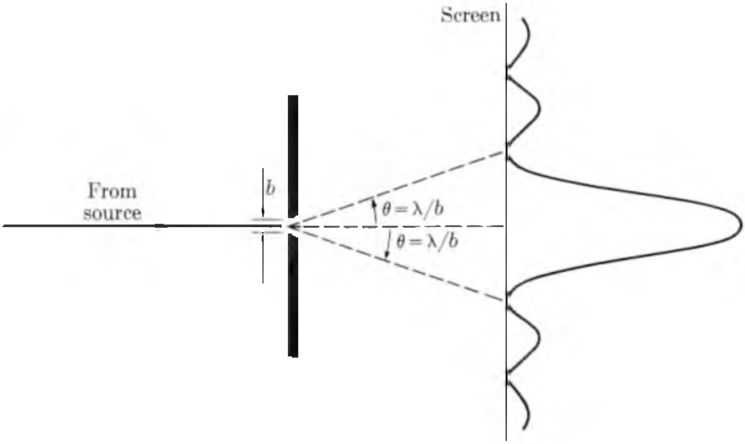


Fig. 17-8. Angle subtended by the central intensity peak of the diffraction pattern of a single slit ($\lambda \ll b$).

either side of the central maximum (Fig. 17-8) correspond to an angle

$$\theta \approx \sin\theta = \pm\frac{\lambda}{b}, \qquad (17.9)$$

obtained by setting $n = \pm 1$ in Eq. (17.1).

A useful concept is the *resolving power* of a slit, defined by the English physicist Lord Rayleigh (1842–1919) as the minimum angle subtended by two incident waves coming from two distant point sources that permit their respective diffraction patterns to be distinguished. When waves coming from two distant sources S_1 and S_2 pass through the same slit in two directions, making an angle θ (Fig. 17-9), the diffraction

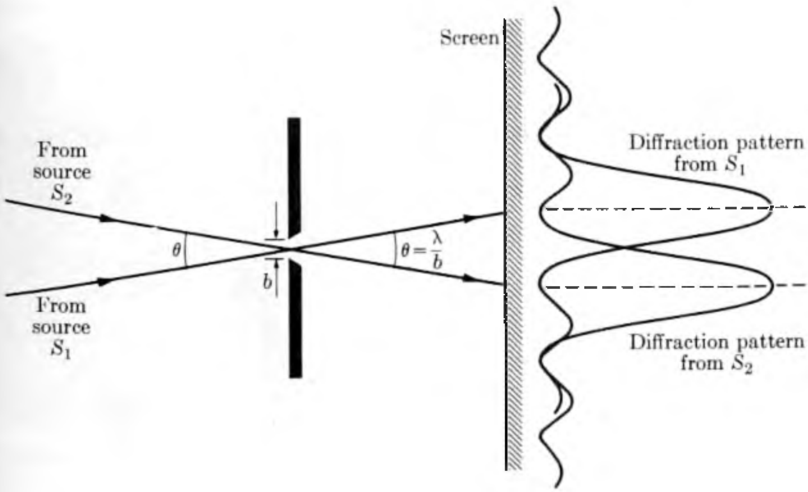


Fig. 17-9. Rayleigh's rule for the resolving power of a slit.

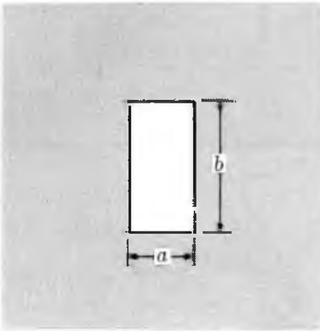


Fig. 17-10. Rectangular slit.

Fig. 17-11. Fraunhofer diffraction pattern of a rectangular slit whose height is twice its width.

patterns of the two sets of waves are superposed. They begin to be distinguishable when the central maximum of one falls on the first zero on either side of the central maximum of the other as indicated on the right in Fig. 17-9. But then, in view of Eq. (17.9) and Fig. 17-8, the angle θ must be

$$\theta=\frac{\lambda}{b}, \tag{17.10}$$

which gives the resolving power of the slit according to Rayleigh's definition. Assuming that S_1 and S_2 are two points on a distant object, Eq. (17.10) gives the minimum angular separation between them in order for the two points to be recognizable as different when the object is observed through the slit. If the light passing through the slit forms an image on a screen, and that image is observed with a microscope, for example, it is not possible, no matter what the magnification of the microscope, to observe more detail in the image than that allowed by the resolving power of the slit. These considerations must be taken into account in the design of optical instruments.

If the slit is rectangular with sides a and b that are of comparable size (Fig. 17-10), the diffraction pattern is the combination of the two patterns due to each pair of sides. Instead of the series of bands shown in Fig. 17-5, we get a series of rectangles arranged in a crosswise form as in the photograph of Fig. 17-11.

In our calculation we have not taken into account the directionality factor, $g=\frac{1}{2}(1+\cos\theta)$, mentioned in Section 13.2 when we were discussing Huygens's principle. This factor tends to decrease further the amplitude of the maxima of higher order.

Example 17.1. Estimation of the magnitude of the successive maxima in the diffraction pattern of a slit.

▼ Successive maxima occur at the maxima of the fraction $\sin u/u$, according to Eq. (17.8). Therefore we must find

$$\frac{d}{du}\left(\frac{\sin u}{u}\right)=0 \qquad \text{or} \qquad \tan u = u.$$

This is a transcendental equation of a kind similar to that of Example 16.6. Its solutions are found by plotting $y=\tan u$ and $y=u$ and finding the points of intersection of the curves; it is left to the student to verify that these intersections occur at $u \approx 4.49, 7.73, 10.9 \ldots$ and correspond to relative intensities of 0.047, 0.017, 0.0088, ... These points are estimated by assuming that the maxima of $\sin u/u$ occur very close to the maxima of $\sin u$; that is, when $u=(n+\frac{1}{2})\pi$ when $n=1, 2, 3 \ldots$. The actual values of u are always slightly *less* than the estimate. Neglecting this small difference, we find that the values of $\sin u/u$ at the maxima are $1/[(n+\frac{1}{2})\pi]$; and the corresponding intensities are

$$I=\frac{I_0}{(n+\frac{1}{2})^2\pi^2}=0.045I_0, 0.016I_0, 0.008I_0. \ ▲$$

17.3 Fraunhofer Diffraction by a Circular Aperture

The diffraction pattern produced by a circular aperture exhibits many of the features already seen in the case of the rectangular slit. But instead of a rectangular pattern like the one shown in Fig. 17-11, the diffraction pattern consists of a bright disk surrounded by alternate dark and bright rings as shown in Fig. 17-12. The radii of the central disk and successive rings do not follow a simple sequence as in the case of the square slit. We shall omit the mathematical analysis of the problem, which is much more involved than in the case of the rectangular slit because of the geometrical arrangement. Assuming that R is the radius of the aperture (Fig. 17-13), the angle cor-

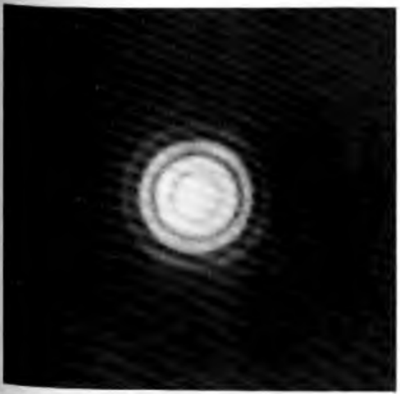


Fig. 17-12. Fraunhofer diffraction pattern of a circular slit.

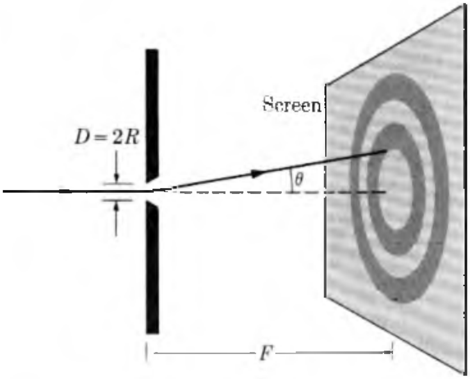


Figure 17-13

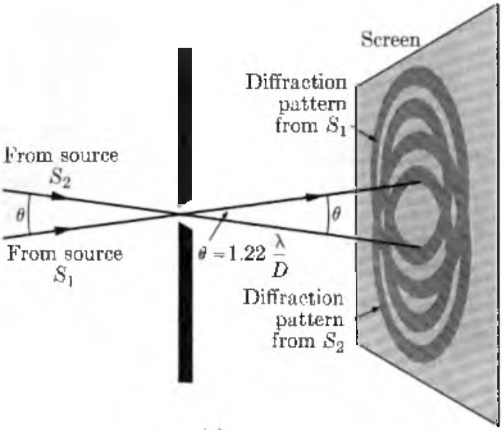


Fig. 17-14. (a) Rayleigh's rule for the resolving power of a circular slit. Part (b) shows two distant point sources imaged through a lens and just resolved.

responding to the first dark ring is given by the condition

$$\frac{2\pi R \sin\theta}{\lambda} = 3.8317 \tag{17.11}$$

or

$$\theta \approx \sin\theta = 1.22\,\frac{\lambda}{2R} = 1.22\,\frac{\lambda}{D} \tag{17.12}$$

where $D = 2R$ is the diameter of the aperture and θ is expressed in radians. This expression also gives the resolving power for a circular aperture, defined again according to Rayleigh as the minimum angle between the directions of incidence of two plane waves coming from two distant point sources such that their respective diffraction patterns may be distinguished. This situation occurs when the center of the bright disk of the diffraction pattern of one source falls on the first dark ring of the diffraction pattern of the second (Fig. 17-14). The angular separation is given by Eq. (17.12); that is, $\theta = 1.22\lambda/D$. This expression appeared in Section 15.5 when we were discussing the magnification of a telescope.

A lens is actually a circular aperture; and therefore the image of a point, which in Chapter 15 was assumed to be another point, is in fact a diffraction pattern. However, the radius of a lens is in general so large compared with the wavelength of light that for most practical purposes, diffraction effects may be ignored.

Example 17.2. A lens with a diameter of 2×10^{-2} m has a focal length of 0.40 m. It is illuminated with a beam of parallel monochromatic light of wavelength 5.9×10^{-7} m. Find the radius of the central disk of the diffraction pattern observed in a plane at the focus. Also determine the resolving power of the lens for this wavelength.

▼ When we use Eq. (17.12), the angle subtended by the central disk in the diffraction pattern is

$$\theta = 1.22 \times \frac{5.9 \times 10^{-7}\ \text{m}}{2 \times 10^{-2}\ \text{m}} = 3.60 \times 10^{-5}\ \text{rad} = 7.42''$$

This is also the resolving power of the lens. The radius of the central disk is

$$r = f\theta = 0.40\ \text{m} \times 3.60 \times 10^{-5}\ \text{rad} = 1.44 \times 10^{-5}\ \text{m},$$

and thus for practical purposes we may say that the image at the focal plane is a point. ▲

17.4 Fraunhofer Diffraction by Two Equal, Parallel Slits

Consider two slits, each of width b, displaced the distance a (Fig. 17-15a). For a direction corresponding to the angle θ, we now have two sets of diffracted waves, and what we observe is the result of the interference of these waves at the screen. In other words, we now have a combination of diffraction and interference. To determine the intensity of the resultant waves in terms of the angle θ, we must compute the resultant amplitude from each slit and combine the two amplitudes to obtain a final resultant. This determination is shown in Fig. 17-16, in which the different rotating vectors are drawn. The angle α has the value given by Eq. (17.4). The magnitude of the vector $\overrightarrow{OP}$ gives the resultant amplitude A_1 from slit 1. The value of this amplitude as given by Eq. (17.7) is

$$A_1 = A_0 \frac{\sin(\pi b \sin\theta/\lambda)}{\pi b \sin\theta/\lambda}. \tag{17.13}$$

Since the two slits have the same width, the resultant amplitude for slit 2 has the same value, A_1, but its phase is different. From Fig. 17-15b we note that between

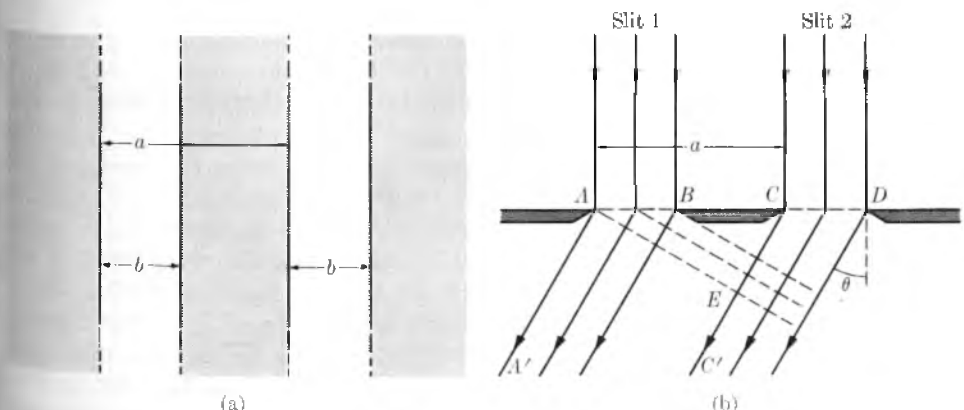


Fig. 17-15. Front view and cross section of two parallel long narrow slits.

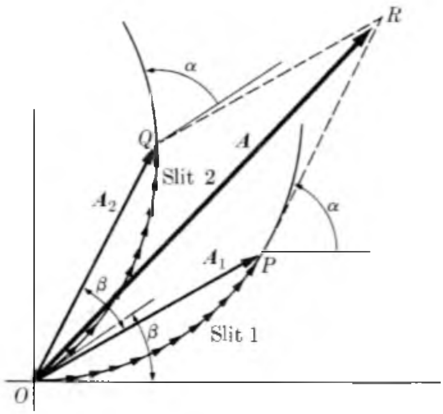


Figure 17-16

corresponding rays from slits 1 and 2 such as AA' and CC', there is a constant phase difference given by

$$\beta = \frac{2\pi}{\lambda} CE = \frac{2\pi a \sin\ \theta}{\lambda}. \tag{17.14}$$

The corresponding amplitudes or vectors from the two slits thus make an angle equal to β. Accordingly, in Fig. 17-16 the line $OQ = A_2$ for slit 2 is obtained by rotating line $OP = A_1$ for slit 1 through the angle β. Their resultant amplitude A is then

$$A = \sqrt{A_1^2 + A_2^2 + 2A_1 A_2 \cos \beta}.$$

Setting $A_1 = A_2$, and using standard trigonometric identities, we may write,

$$A = A_1\sqrt{2(1 + \cos \beta)} = 2A_1 \cos \tfrac{1}{2}\beta.$$

Therefore, using Eqs. (17.13) and (17.14), we obtain

$$A = 2A_0 \frac{\sin (\pi b \sin \theta/\lambda)}{\pi b \sin \theta/\lambda} \cos \frac{\pi a \sin \theta}{\lambda}.$$

The intensity distribution, which we know is proportional to the square of the amplitude, is then

$$I = I_0 \left[\frac{\sin (\pi b \sin \theta/\lambda)}{\pi b \sin \theta/\lambda}\right]^2 \cos^2 \frac{\pi a \sin \theta}{\lambda}. \tag{17.15}$$

When we compare this equation with Eq. (17.8) for a single slit, we realize that we now have the additional factor $\cos^2 [(\pi a \sin \theta)/\lambda]$. But if we recall Eq. (16.10), which gives the intensity distribution for the interference pattern of two synchronous sources, we see that Eq. (16.10) and Eq. (17.15) coincide insofar as the interference factor is concerned since in Eq. (17.15) a is the separation of the two slits and in Eq. (16.10) a is the separation of the two sources. Therefore the equation that describes the overall diffraction pattern for two slits is the equation for the interference pattern

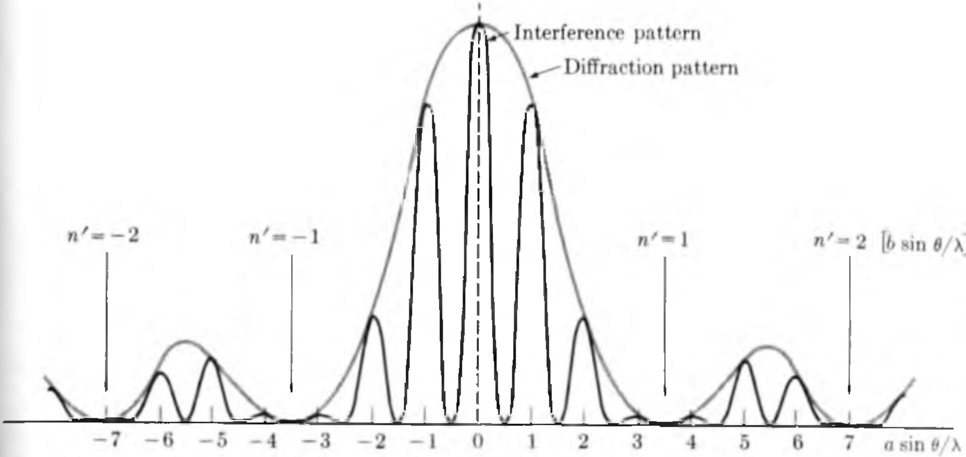


Fig. 17-17. Intensity distribution (along a plane set normal to the incident light) resulting from two parallel long narrow slits. ($a/b = 3.5$.)

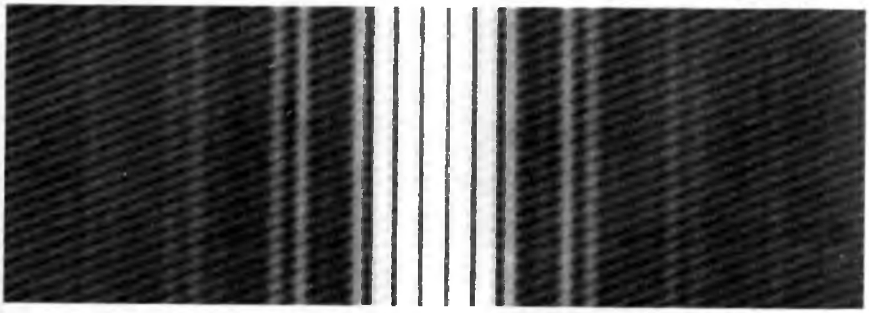


Fig. 17-18. Fraunhofer diffraction pattern caused by two parallel long narrow slits.

of two synchronous sources modulated by the expression for the diffraction pattern of a single slit as is shown in Fig. 17-17 and in the photograph of Fig. 17-18.

Note that the maxima of the interference pattern occur for $\pi a \sin \theta/\lambda = n\pi$, or $\sin \theta = n(\lambda/a)$; the zeros of the diffraction pattern are given by Eq. (17.2), or $\sin \theta = n'(\lambda/b)$. Since $a > b$, the zeros of the diffraction pattern are more widely spaced than the maxima of the interference pattern. Therefore, when there are two slits, the bright fringes are much narrower and more closely spaced than those produced by a single slit.

17.5 Diffraction Gratings

The next step is to consider the diffraction pattern produced by several parallel slits of equal width b, equally spaced the distance a. Let N be the number of slits. From

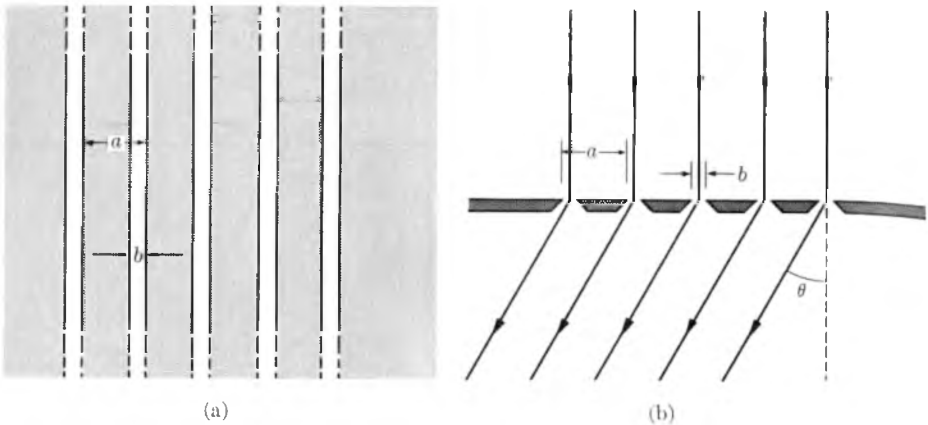


Fig. 17-19. Front view and cross section of a diffraction grating.

Fig. 17-19 we see, by similarity with the problem of the double slit, that in the direction corresponding to the angle θ we will observe the interference caused by N synchronous sources (one per slit) modulated by the diffraction pattern of one slit. Since the separation between successive sources is a, the interference factor for the intensity is the same as that found in Eq. (16.14); that is,

$$\left[\frac{\sin(N\pi a\sin\theta/\lambda)}{\sin(\pi a\sin\theta/\lambda)}\right]^2;$$

the intensity is modified by the diffraction factor, which according to Eq. (17.8) is

$$\left[\frac{\sin(\pi b\sin\theta/\lambda)}{\pi b\sin\theta/\lambda}\right]^2.$$

The intensity distribution is then

$$I=I_0\left[\frac{\sin(\pi b\sin\theta/\lambda)}{\pi b\sin\theta/\lambda}\right]^2\left[\frac{\sin(N\pi a\sin\theta/\lambda)}{\sin(\pi a\sin\theta/\lambda)}\right]^2. \tag{17.16}$$

If the number of slits N is large, the pattern will consist of a series of narrow bright fringes corresponding to the main maxima of the interference pattern that are given by

$$a\sin\theta=n\lambda \qquad \text{or} \qquad \sin\theta=n\left(\frac{\lambda}{a}\right) \tag{17.17}$$

where $n=0,\ \pm1,\ \pm2,\ldots$; but their intensities are modulated by the diffraction pattern. Figure 17-20 shows the case for eight slits ($N=8$). According to the value of n, the principal maxima are called the first, second, third, etc., order of diffraction.

A system such as the one we have just discussed is called a *transmission diffraction grating*. For purposes of analyzing near-infrared, visible, or ultraviolet light, transmission diffraction gratings consist of several thousands of slits per centimeter, obtained by etching a series of parallel lines on a transparent film. The lines then act

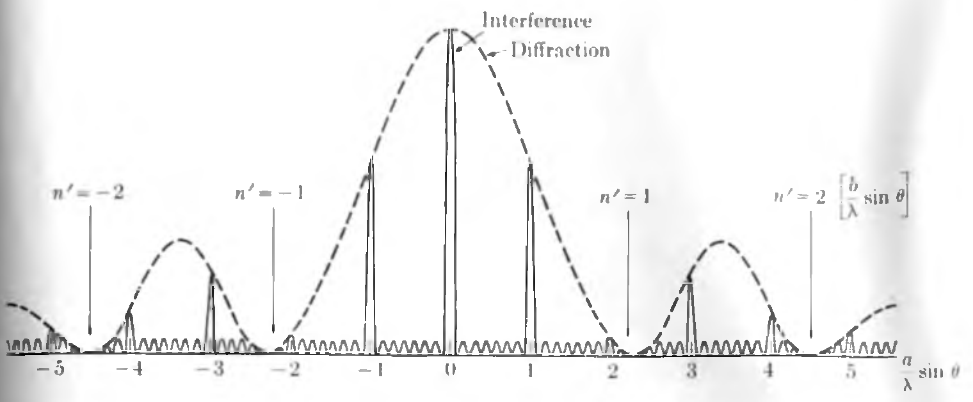


Fig. 17-20. Intensity distribution produced by a diffraction grating on a plane placed normal to the incident light and parallel to the grating. ($a/b \approx 2.33$).

as the opaque spaces between the slits. A diffraction grating can also work by reflection: a series of parallel lines is etched on a metallic surface. The narrow strips between the etchings reflect light and produce a diffraction pattern (see Problem 17.36). Sometimes the surface is made concave to improve focusing (see Problem 17.37).

When light of several wavelengths falls on a grating, the different wavelengths produce diffraction maxima at different angles except for the zero order, which is the same for all. The set of maxima of a given order for all wavelengths constitutes a *spectrum*. So we have spectra of first, second, third, etc., orders. Note that the longer the wavelength, the larger the deviation for a given order of spectrum. Therefore red is deviated more than violet; this effect is the opposite of what happens when light is dispersed by a prism. The *dispersion* of a grating is defined by $D = d\theta/d\lambda$; that is, the rate of charge of the angle of deviation with respect to wavelength. From Eq. (17.17) we have $\cos\theta(d\theta/d\lambda) = n/a$ and thus

$$D = \frac{d\theta}{d\lambda} = \frac{n}{a\cos\theta}, \tag{17.18}$$

indicating that the higher the order of diffraction, the greater the dispersion.

Diffraction gratings are very important in spectrum analysis in a wide range of regions of the electromagnetic spectrum, and have several distinct advantages over prisms. One advantage is that diffraction gratings do not depend on the dispersive properties of the material, but only on the geometry of the grating. Figure 17-21 shows the basic elements of a grating spectroscope.

Example 17.3. Angular separation of the whole visible spectrum for first-order and second-order diffraction spectra.

▼ Assume that the wavelength of visible light extends from 3.90×10^{-7} m up to 7.70×10^{-7} m. For a grating containing 20,000 lines and a length of 4 cm, we have that $a = 4 \times 10^{-2}$ m/20,000 =

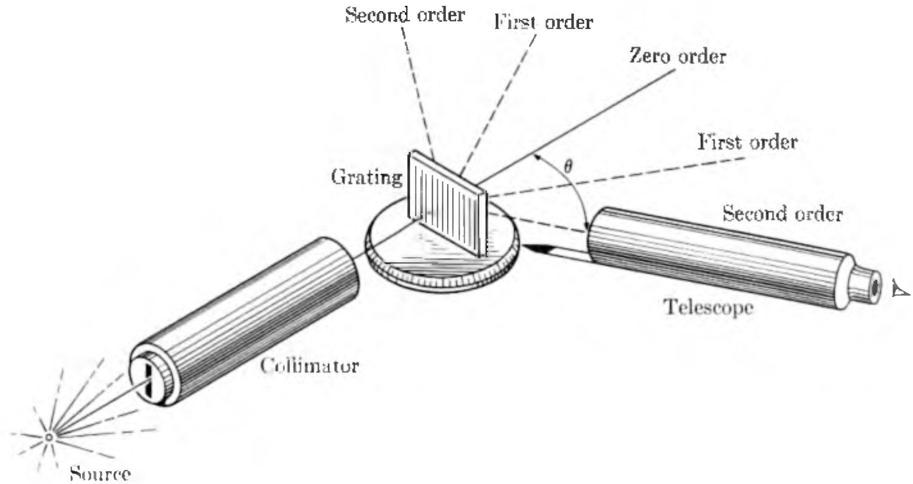


Fig. 17-21. Grating spectroscope. The source is placed in front of the slit on the collimator. The diffraction grating is placed perpendicular to the collimator's axis, and the spectra of different orders are investigated by moving the telescope.

2×10^{-6} m. Therefore, using Eq. (17.17), we have for $n=1$

$$\sin \theta_{\text{red}} = \frac{7.70 \times 10^{-7}}{2 \times 10^{-6}} = 0.385 \qquad \text{or} \qquad \theta_{\text{red}} = 22^\circ 39',$$

$$\sin \theta_{\text{violet}} = \frac{3.90 \times 10^{-7}}{2 \times 10^{-6}} = 0.195 \qquad \text{or} \qquad \theta_{\text{violet}} = 11^\circ 15'.$$

Therefore the first-order spectrum covers an angle of $11^\circ 24'$. Similarly, for the second-order spectrum, the angle is $27^\circ 24'$ as the student may calculate. Is a full third-order spectrum possible? ▲

Example 17.4. Position of the principal maxima of a diffraction grating when the angle of incidence of plane monochromatic waves is not zero.

▼ The principal maxima are determined by the interference pattern, and this in turn is determined by the phase difference between corresponding rays in successive slits. Figure 17-22, in which i is the angle of incidence and θ is an angle of diffraction, shows that such a phase difference is given by

$$\delta = \frac{2\pi}{\lambda}(AB + BC) = \frac{2\pi a(\sin i + \sin \theta)}{\lambda}. \tag{17.19}$$

In order that Eq. (17.19) have general validity, the signs indicated in the figure must be observed: otherwise, Eq. (17.19) must be rewritten for each of the possible orientations of i and θ relative to the normal. The condition for a maximum then becomes

$$a(\sin i + \sin \theta) = n\lambda. \tag{17.20}$$

For $n=0$, we have $\sin \theta = -\sin i$ or $\theta = -i$, corresponding to the continuation of the incident ray.

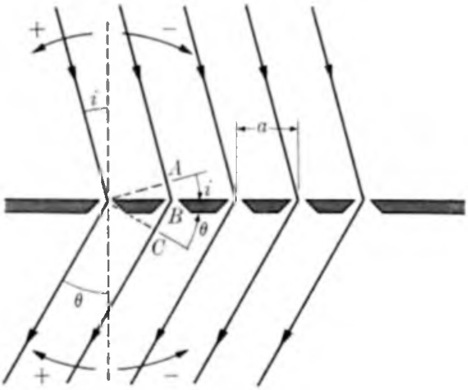


Fig. 17-22. Diffraction grating with oblique incidence.

If we transform this condition for a maximum into a product by means of trigonometric formulas, we have

$$2a \sin \tfrac{1}{2}(i+\theta) \cos \tfrac{1}{2}(i-\theta) = n\lambda.$$

Therefore the deviation $D = i + \theta$ for the maximum of order n may be found from

$$\sin \tfrac{1}{2}D = \frac{n\lambda}{2a} \sec \tfrac{1}{2}(i-\theta); \tag{17.21}$$

and thus the deviation is a minimum when $\theta = i$; and the angle of incidence for the minimum deviation for order n is found from

$$\sin i = \frac{n\lambda}{2a}. \ \blacktriangle$$

Example 17.5. *Resolving power* of a diffraction grating.

▼ When two plane waves of slightly different wavelength fall on a diffraction grating, the principal maxima of the same order for each wavelength may fall so close to each other that it is impossible to distinguish whether the original beam was monochromatic or not. In order that the two wavelengths may be distinguished (or resolved) in a given order, it is necessary that the principal maximum for one of the wavelengths fall on the first zero on either side of the principal maximum of the other wavelength. Given that $\Delta\lambda$ is the minimum wavelength difference for which the condition above is met at a wavelength λ, the resolving power of the grating is

$$R = \frac{\lambda}{\Delta\lambda}. \tag{17.22}$$

Consider, as an example, a wavelength λ such that Eq. (17.17) holds. The maxima of intensity correspond to the angle given by $\sin \theta = n\lambda/a$. Then by differentiating,

$$\cos \theta \, \Delta\theta = n \frac{\Delta\lambda}{a}.$$

In accordance with Eq. (17.16), the zeros on either side of a maximum of order n are given by

$$\frac{N\pi a \sin\theta}{\lambda} = (Nn \pm 1)\pi \qquad \text{or} \qquad \sin\theta = \frac{Nn \pm 1}{N}\frac{\lambda}{a}.$$

Calling θ' and θ'' the two minimum angles given by this equation such that $\theta' - \theta'' = 2\Delta\theta$, we can write

$$\sin\theta' - \sin\theta'' = \frac{2\lambda}{Na},$$

or using the trigonometric identity,

$$\sin\tfrac{1}{2}(\theta' - \theta'')\cos\tfrac{1}{2}(\theta' + \theta'') = \frac{\lambda}{Na}$$

Since θ' is almost equal to θ'', we may replace $\sin\frac{1}{2}(\theta' - \theta'')$ by $\frac{1}{2}(\theta' - \theta'')$ and $\cos\frac{1}{2}(\theta' + \theta'')$ by $\cos\theta$ and write $\frac{1}{2}(\theta' - \theta'')\cos\theta = \Delta\theta\cos\theta = \lambda/Na$. But from the equation above, $\cos\theta\,\Delta\theta = n\Delta\lambda/a$. Therefore we finally have $\lambda/N = n\Delta\lambda$ or

$$R = \frac{\lambda}{\Delta\lambda} = Nn. \tag{17.23}$$

This equation means that the greater the total number of lines of the grating and the higher the order of the spectrum, the smaller $\Delta\lambda$, and so the greater the resolving power of the grating. On the other hand, Eq. (17.23) shows that the resolving power is independent of the size and spacing of the ruling in the grating. ▲

Example 17.6. Resolution of the yellow doublet of sodium.

▼ We wish to determine whether the grating of Example 17.3 can resolve the two yellow lines of sodium, whose wavelengths are 5.890×10^{-7} m and 5.896×10^{-7} m. The average wavelength of the two lines is 5.893×10^{-7} m, and their separation is 6×10^{-10} m. From the results of Example 17.6, we have that the resolving power of the grating is $R = Nn = 2 \times 10^4 n$. At the given wavelength, the minimum wavelength separation in the first-order spectrum ($n = 1$) is

$$\Delta\lambda = \frac{\lambda}{R} = \frac{5.893 \times 10^{-7}}{2 \times 10^4 \times 1\ \text{m}} = 2.947 \times 10^{-11}\ \text{m},$$

which is one-twentieth of the separation of the two sodium lines. Hence the two D-lines in the first-order spectrum produced by this grating could be easily separated (or seen distinctly). ▲

17.6 Fresnel Diffraction

As mentioned in Section 17.1, *Fresnel diffraction* takes place when either the point source of incident waves or the observation point from which they are seen (or both) are at a finite distance from the aperture or obstacle responsible for the diffraction. The mathematical calculations for Fresnel diffraction are more involved than the calculations for Fraunhofer diffraction, but the physical ideas remain the same. Therefore we shall discuss only the fundamental aspects, and for simplicity we shall

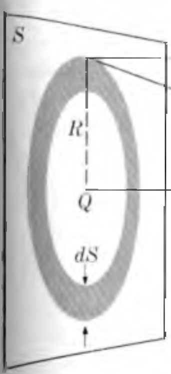


Figure 17-23

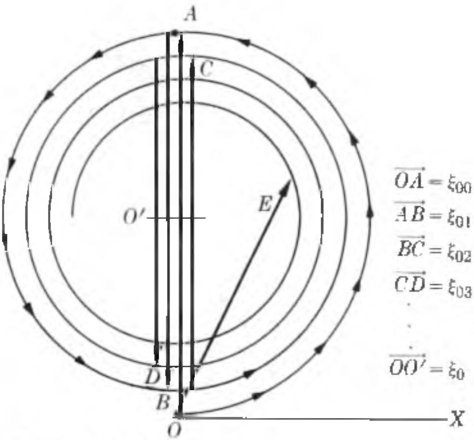


Figure 17-24

assume that the source of the waves is so far away from the screen that the incident waves are plane, and that they are perpendicular to the aperture or obstacle.

Suppose that we want to compute the wave motion to be expected at point P when we know the wave motion at a certain plane wave front S (Fig. 17-23). According to the Huygens-Kirchhoff principle as formulated in Section 13.2, we may divide the wave front into surface elements. Symmetry suggests that they be chosen as circular rings concentric at the projection, Q, of P on the plane S. Then the contribution of the surface element of area dS to the wave motion at P, according to Eq. (13.2), has an amplitude proportional to

$$\frac{dS}{r} g(\theta) \tag{17.24}$$

where dS is the area of the ring. The phase at P of the wave produced by dS will be

$$\delta = \frac{2\pi r}{\lambda}. \tag{17.25}$$

By adding the rotating vectors of successive rings whose amplitudes are characterized by Eqs. (17.24) and whose phases are given by Eq. (17.25), we can obtain the resultant amplitude and phase at P for all points of the plane S. Because of the $1/r$ and $g(\theta)$ factors, the vectors become smaller and smaller in size as we go to greater radii R, and they result in a spiral instead of a circle as shown in Fig. 17-24.

To simplify the calculation, assuming that λ is much smaller than r_0 and P is reasonably close to the plane S, we divide the surface into rings called *Fresnel zones* (Fig. 17-25), whose outer edge distances to P differ successively by $\frac{1}{2}\lambda$. That is, $r_1 = r_0 + \frac{1}{2}\lambda$, $r_2 = r_1 + \frac{1}{2}\lambda$, $r_3 = r_2 + \frac{1}{2}\lambda$, etc. This arrangement has the property that corresponding rays from successive zones arriving at P have a phase difference of π and interfere destructively; that is,

$$\delta_{n+1} - \delta_n = \frac{2\pi}{\lambda}(r_{n+1} - r_n) = \frac{2\pi}{\lambda}\left(\frac{\lambda}{2}\right) = \pi.$$

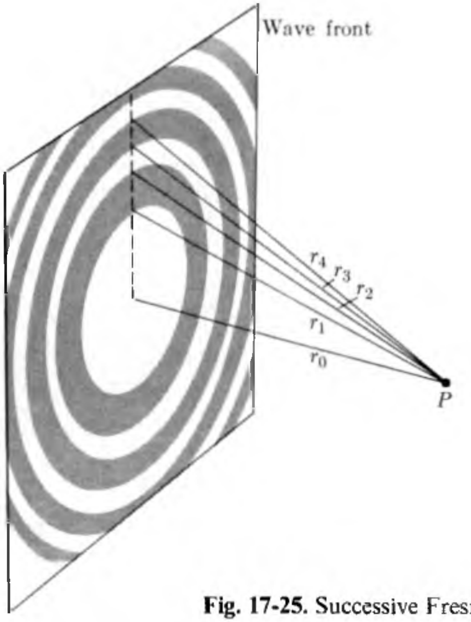


Fig. 17-25. Successive Fresnel zones.

If ξ_{0n} is the amplitude produced at P by the nth zone, which is proportional to the value given in Eq. (17.24), the resultant amplitude at P is

$$\xi_0 = \xi_{00} - \xi_{01} + \xi_{02} - \xi_{03} + \cdots. \tag{17.26}$$

We may also write this in the form

$$\xi_0 = \tfrac{1}{2}\xi_{00} + (\tfrac{1}{2}\xi_{00} - \xi_{01} + \tfrac{1}{2}\xi_{02}) + (\tfrac{1}{2}\xi_{02} - \xi_{03} + \tfrac{1}{2}\xi_{04}) + \cdots.$$

The amplitudes from neighboring zones are almost equal in magnitude although they decrease as n increases; that is, $\xi_{00} > \xi_{01} > \xi_{02} > \ldots$. So we may assume as a good approximation that $\frac{1}{2}\xi_{00} - \xi_{01} + \frac{1}{2}\xi_{02} \approx 0$, and in general $\frac{1}{2}\xi_{0(n-1)} - \xi_{0n} + \frac{1}{2}\xi_{0(n+1)} \approx 0$. Therefore the summation in Eq. (17.26) for an infinite plane effectively reduces to

$$\xi_0 = \tfrac{1}{2}\xi_{00}, \tag{17.27}$$

and the wave motion at P results from the part of the wave front directly in line with P, and is equal in amplitude to one-half the contribution of the first Fresnel zone only.

Note that each Fresnel zone is composed of many of the circular surface elements illustrated in Fig. 17-23. To understand the situation in terms of an amplitude vector diagram such as that of Fig. 17-24, note that for the first zone the distance goes from r_0 to $r_0 + \frac{1}{2}\lambda$, or the phase from $2\pi r_0/\lambda$ to $(2\pi r_0/\lambda) + \pi$. This fact means that when we draw all the amplitude vectors from all secondary sources within this zone, their phase difference changes gradually from zero to π. These vectors constitute the arc from O to A in Fig. 17-24, and the amplitude ξ_{00} of the first zone is the vector $\overrightarrow{OA}$.

For the second zone, the distances go from $r_0+\frac{1}{2}\lambda$ to $r_0+\lambda$, or the phases from $(2\pi r_0/\lambda)+\pi$ to $(2\pi r_0/\lambda)+2\pi$, and again result in a phase difference of π between extremes so that the second zone corresponds to the arc from A to B with its amplitude ξ_{01} equal to vector $\vec{AB}$. This procedure is repeated until all zones are covered. The spiral converges at a point O' so that the resultant amplitude is $\vec{OO'}$, which is approximately $\frac{1}{2}\vec{OA}$ as in Eq. (17.27).

Since $r_n=r_0+\frac{1}{2}n\lambda$, the radius of zone n, from Fig. 17-25, is $R_n^2=r_n^2-r_0^2=(r_0+\frac{1}{2}n\lambda)^2-r_0^2=n\lambda r_0+\frac{1}{4}n^2\lambda^2$. If n is not very large, the last term can be neglected (since $\lambda \ll r_0$) so that

$$R_n^2=n\lambda r_0. \tag{17.28}$$

Incidentally, this equation shows that all Fresnel zones have the same area, equal to $\pi\lambda r_0$.

When the wave front is blocked by a screen, the situation is different from Eq. (17.27) because some zones now contribute only partially (or not at all) to the wave motion at P since they are blocked by the screen. Suppose that an incident wave falls at normal incidence on a screen having a circular aperture of radius a. The observation point is on a line perpendicular to the screen through the center of the aperture so that the Fresnel zones are concentric with the aperture. When the point of observation is a distance r_0 such that $a^2=\lambda r_0$, only one zone passes through the aperture and produces at P an amplitude ξ_{00}. The amplitude at P is twice the value obtained in Eq. (17.27) for the whole wave front; the result is that the illumination at P is four times as great as when no screen is present and the whole wave front is exposed! If the aperture is larger or the point closer so that $a^2=2\lambda r_0$, the first two zones pass through the aperture and result in an amplitude of $\xi_{00}-\xi_{01}$, which is practically zero and results in darkness at P! In general, as long as our approximation remains valid, we shall have maximum brightness or darkness at the center of the diffraction pattern, depending on whether n is odd or even, where n is the number of Fresnel zones falling within the aperture relative to the point at which the diffraction is observed. The situation for a few different values of n is shown in Fig. 17-26.

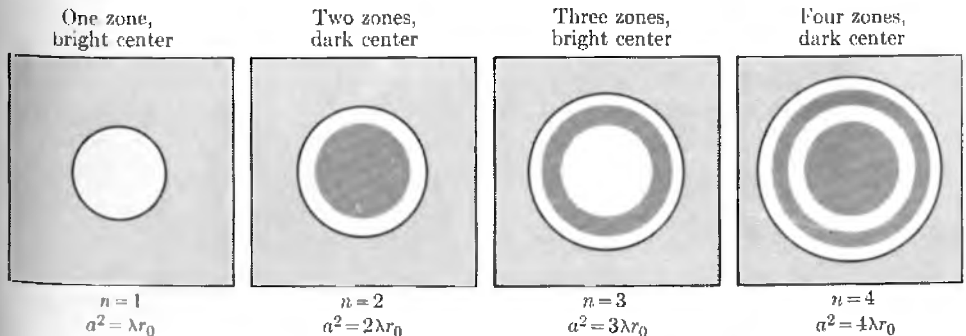


Fig. 17-26. Change in Fresnel zones for a fixed point resulting from a change in the size of the aperture.

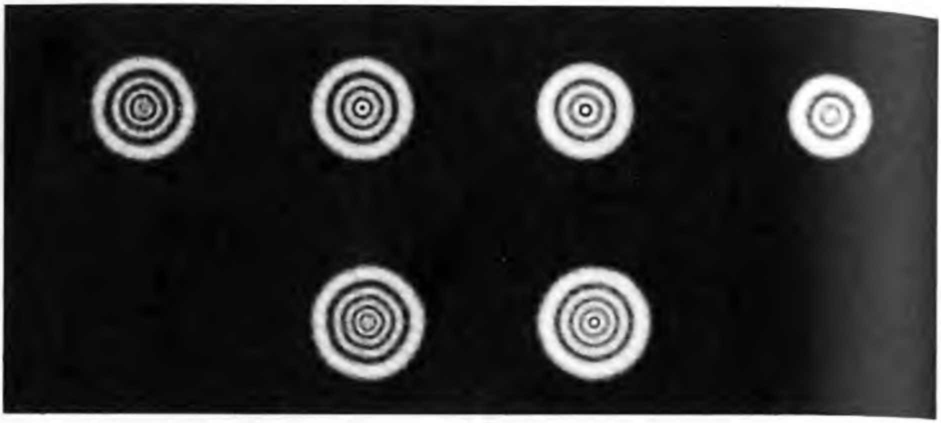


Fig. 17-27. Fresnel diffraction patterns of circular apertures of different radii.

Using the diagram of Fig. 17-24, we see that when only one zone is exposed, the resultant amplitude is $\vec{OA}=\xi_{00}$. When two zones are exposed, the resultant amplitude is $\vec{OB}=\vec{OA}+\vec{AB}=\xi_{00}-\xi_{01}$. For three zones, it is

$$\vec{OC}=\vec{OA}+\vec{AB}+\vec{BC}=\xi_{00}-\xi_{01}+\xi_{02}\approx\tfrac{1}{2}(\xi_{00}+\xi_{02}),$$

and so on. In general when a certain number of complete Fresnel zones plus a fraction are exposed, one can obtain the resultant amplitude by drawing, on a diagram such as Fig. 17-24, the vector $\vec{OE}$ that goes from O to the point E, which corresponds to the exact number of zones plus the fraction of the last. In the case shown in Fig. 17-24, E corresponds to six zones plus a fraction of a seventh.

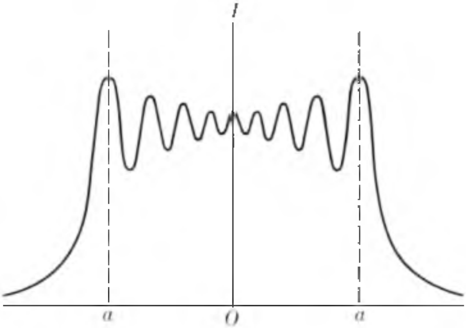


Fig. 17-28. Intensity distribution of Fresnel diffraction by a circular aperture.

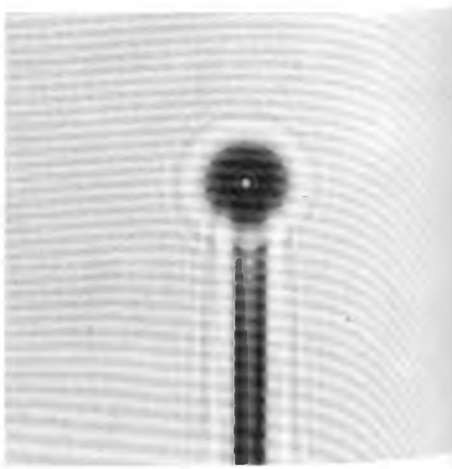


Fig. 17-29. Fresnel diffraction by a small circular disk supported by a thin rod.

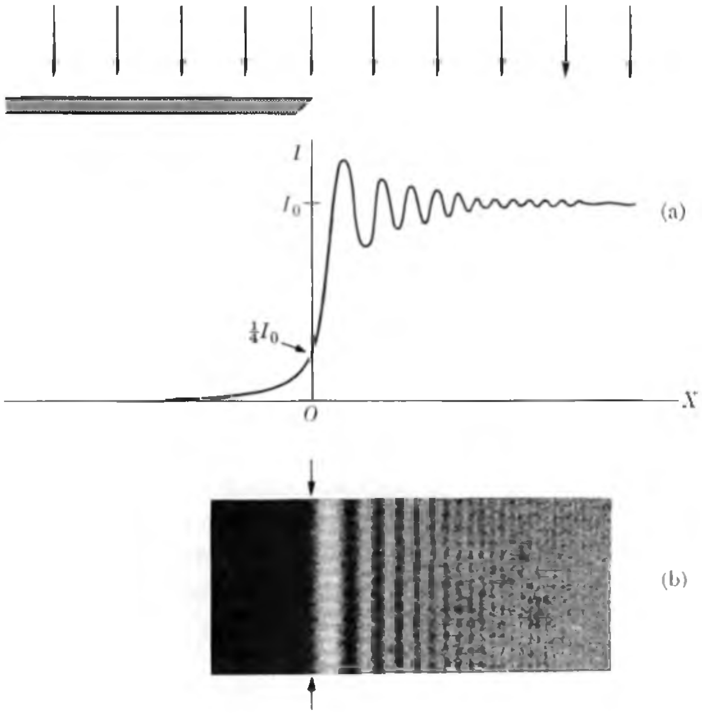


Fig. 17-30. (a) Intensity distribution for Fresnel diffraction by a straight edge. (b) Photograph of the Fresnel diffraction by a straight edge.

When the size of the aperture is changed but the distance to the screen is held fixed. the different active zones contribute differently to the resulting wave motion and give rise to a diffraction pattern composed of a series of rings concentric with P. These rings alternate in brightness as shown in Fig. 17-27. Figure 17-28 shows the intensity distribution as a function of the distance from the axis of the aperture for a circular aperture of radius a, comprising several zones.

If instead of a circular aperture we have a circular disk, the diffraction pattern is similar except that at the center there is always brightness (Fig. 17-29). The cause of this brightness is that the first unexposed Fresnel zone always gives a positive contribution at the center, for the same reason that a completely exposed plane wave front always results in a bright spot.

For a rectangular slit the situation is very similar to that for the circular aperture except that instead of rings the Fresnel zones are strips parallel to the slit.

At an edge the diffraction pattern has the intensity distribution shown in Fig. 17-30 with the intensity falling off gradually to zero within the geometrical shadow and fluctuating during the first few wavelengths within the geometrical region of illumination.

Example 17.7. A screen with a small hole 1 mm in diameter is illuminated with light of wavelength 5.9×10^{-7} m. Calculate the distance along the perpendicular from the screen to the farthest point of darkness.

▼ In this case the radius of the hole is $a = 0.5$ mm $= 5 \times 10^{-4}$ m and the wavelength is $\lambda = 5.9 \times 10^{-7}$ m. The farthest point of darkness is that point at which only two Fresnel zones are within the aperture. Thus according to Eq. (17.28) with $n = 2$ and R_n replaced by a, we have $a^2 = 2\lambda r_0$ or $r_0 = a^2/2\lambda = 0.212$ m, which means that the farthest point of darkness is about 21.2 cm from the screen. At about .42 m from the screen, there should be a very bright point corresponding to the case in which only the first Fresnel zone is within the aperture. In general, successive closer points of darkness are at distances of $a^2/2k\lambda$ (where k is an integer greater than one) from the screen. ▲

17.7 Scattering

So far in our discussion of diffraction we have implicitly assumed that the objects interposed in the path of a wave play a *passive* role. That is, we have assumed that their only role is to interrupt a part of the wave front without themselves adding any new wave. With such an assumption, the diffraction effects observed are exclusively due to the distorted incident wave motion.

However, in many instances this is not a realistic picture. Suppose, for example, that a sphere of elastic material is suspended in the air and that an acoustic or compressional wave is produced nearby. When the wave passes around the sphere, the wave first of all suffers a diffraction of the type discussed previously. But in addition the elastic sphere undergoes oscillatory deformations caused by the pressure fluctuations accompanying the wave. The oscillations of the surface of the sphere in turn produce new perturbations or waves in the surrounding air; these are superposed on the initial wave. The new waves produced by the oscillating sphere are the *scattered* waves, and the process is called *scattering.*

Similarly if a conducting sphere is placed in the path of an electromagnetic wave, the electric and magnetic fields of the wave induce oscillations in the free charges on the sphere. These oscillating charges emit electromagnetic radiation, thus producing a new or scattered electromagnetic wave.

In Chapter 12 we discussed scattering by a single electron, a purely dynamical problem at the atomic order of magnitude. The scattering we are describing here has a more macroscopic nature since this scattering involves bodies composed of many atoms or containing many electrons. We can compute the magnitude of this macroscopic scattering by applying certain boundary conditions at the surface of the body. These conditions determine the nature of the scattered wave. For example in the case of a perfectly conducting sphere, we must require that at the surface of the sphere the tangential component of the resultant electric field (i.e., the sum of the field of the incident wave and the scattered wave) be zero.

Scattering processes are extremely important to all wave phenomena. However, a more thorough discussion of scattering requires a mathematical treatment beyond the scope of this text.

17.8 X-Ray Scattering by Crystals

Electromagnetic waves, such as X-rays and γ-rays, with wavelengths shorter than the ultraviolet are not noticeably affected by objects of the dimensions used for the optical region. However, a crystal lattice with atoms or molecules regularly spaced at distances of the order of 10^{-10} m provides an excellent medium for producing diffraction of X-rays. This problem is somewhat more complicated than those discussed previously in this chapter, for two reasons. In the first place, since such a crystal is a three-dimensional array, the diffraction centers are distributed in space rather than in one direction as indicated in Fig. 17-31 for a NaCl crystal. (The dark and light spheres correspond to the Na^+ and Cl^- ions.) Second, under the action of the electric field of an electromagnetic wave, the atoms or molecules in a crystal become secondary sources of radiation as previously explained. Therefore we actually have more a scattering than a diffraction phenomenon.

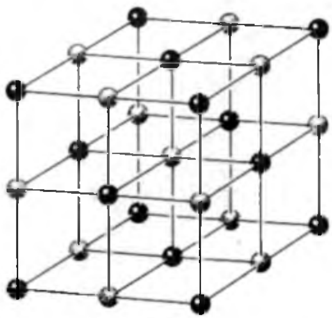


Fig. 17-31. Simplified representation of a sodium chloride crystal shows the regular arrangement of atoms forming a cubic lattice.

When X-rays pass through the crystal, the intensity of the scattered rays is the result of the interference (along the direction of observation) of the waves emitted by each atom or molecule. When the crystal is composed of more than one class of atoms, each kind of atom contributes in a different way to the scattering of the X-rays. Thus, to simplify our calculation, we shall assume that we have only one class of atoms and only one atom per unit cell in the crystal.

Consider two atoms A and B, separated the distance r (Fig. 17-32). Let $\boldsymbol{u}_i$ be a unit vector along the direction of propagation of the incident waves, and $\boldsymbol{u}_s$ a similar unit vector along the direction of the scattered waves. The path length difference for the incident and scattered waves for those two atoms is $AD - BC$, and the phase shift is given by

$$\delta = \frac{2\pi}{\lambda}(AD - BC).$$

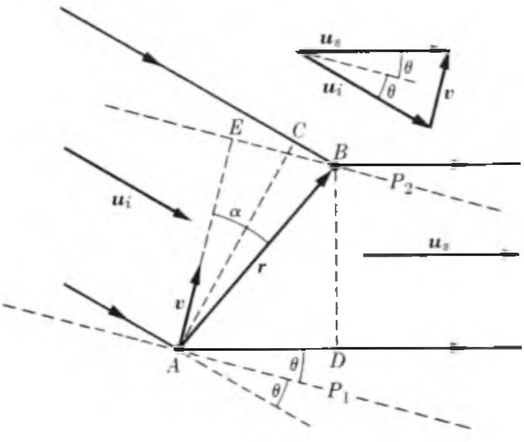


Fig. 17-32. X-ray scattering by two atoms A and B.

But $AD = \boldsymbol{u}_s \cdot \boldsymbol{r}$ and $BC = \boldsymbol{u}_i \cdot \boldsymbol{r}$. Therefore

$$\delta = \frac{2\pi}{\lambda}(\boldsymbol{u}_s - \boldsymbol{u}_i) \cdot \boldsymbol{r} = \frac{2\pi}{\lambda}\, \boldsymbol{v} \cdot \boldsymbol{r} \tag{17.29}$$

where $\boldsymbol{v} = \boldsymbol{u}_s - \boldsymbol{u}_i$. Designating the angle between $\boldsymbol{u}_s$ and $\boldsymbol{u}_i$ by 2θ, we see from the insert in Fig. 17-32 that

$$v = 2 \sin \theta. \tag{17.30}$$

The condition for constructive interference in the direction $\boldsymbol{u}_s$ is $\delta = 2n\pi$ or, in view of Eq. (17.29),

$$\boldsymbol{v} \cdot \boldsymbol{r} = n\lambda \tag{17.31}$$

where as before, n is a positive or a negative integer. Equation (17.31) is the equation of a plane perpendicular to the vector $\boldsymbol{v}$. Therefore for a given wavelength λ and a given direction of incidence, Eq. (17.31) gives a series of parallel planes, one for each value of n. Figure 17-32 shows two such planes, P_1 and P_2. For all atoms located on these planes, condition (17.31) holds and they all contribute to a maximum of intensity in the direction $\boldsymbol{u}_s$. In Eq. (17.31), $n = 0$ corresponds to the plane passing through A, $n = \pm 1$ to the next closest plane on either side, $n = \pm 2$ for the next pair of planes, and so on.

From Fig. 17-32 and using Eq. (17.30), we see that $\boldsymbol{v} \cdot \boldsymbol{r} = vr \cos \alpha = 2d \sin \theta$ where $d = AE = r \cos \alpha$ is the distance between planes P_1 and P_2. Then Eq. (17.31) becomes

$$2d \sin \theta = n\lambda, \tag{17.32}$$

an expression known as *Bragg's equation* after the English physicists W. H. Bragg (1862–1942) and his son W. L. Bragg (1890–1971), who first obtained it. The values of n are limited by the condition that $\sin \theta$ is always smaller than one. The geometry involved in this equation is shown in Fig. 17-33. For rays such as 1 and 2 that are

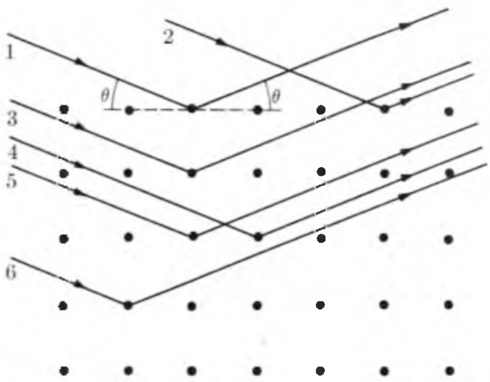


Fig. 17-33. Parallel scattering planes in a crystal.

scattered by atoms in the same plane, the phase difference is zero ($n=0$) and they interfere constructively. This situation however, holds for any angle of incidence. The important point of Bragg's condition is that rays such as 3, 4, 5, . . . , coming from successive planes also interfere constructively, giving rise to a very intense maximum. Therefore Bragg's condition expresses a sort of collective effect, in which the rays scattered by all atoms in certain parallel planes interfere constructively. For fixed planes (or fixed d) and wavelength λ, changing the angle θ alternately produces positions of maximum and minimum intensity, corresponding to constructive (as given by Eq. 17.32), or destructive interference. Note that Eq. (17.32) can be used to measure the plane separation d if the wavelength λ is known, and conversely.

A schematic drawing of the experimental arrangement for observing Bragg's scattering of X-rays, a device called a *crystal spectrometer*, is shown in Fig. 17-34. For a given direction of incidence $\boldsymbol{u}_i$, Eq. (17.31) defines a series of possible families of parallel planes, producing a maximum for scattering in the directions $\boldsymbol{u}_s$ characteristic of each family. The intensity depends on the number of atoms in each family

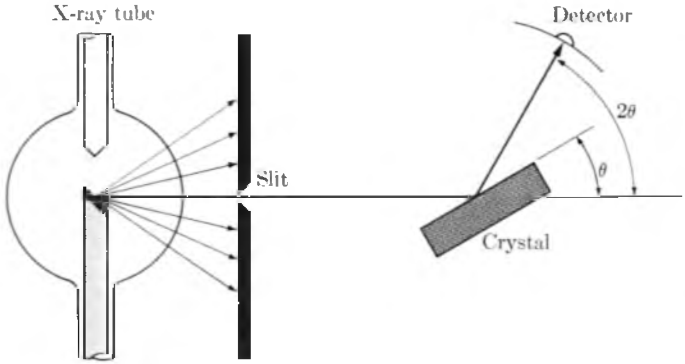


Fig. 17-34. Crystal spectrometer for X-ray diffraction. X-rays generated by the tube at left and collimated by a slit in a lead block are diffracted by the crystal. The diffracted X-rays are observed by a movable detector, usually an ion chamber.

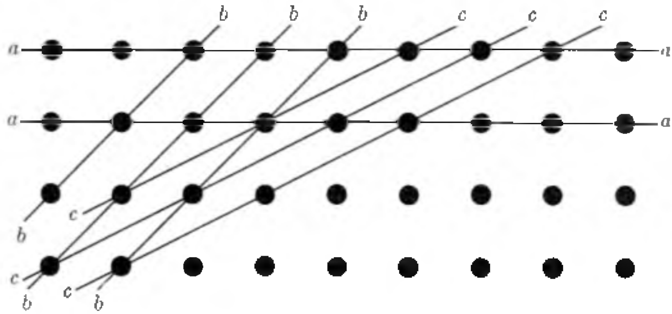


Fig. 17-35. Several possible parallel scattering planes in a crystal.

of planes. Some of the possible families of planes are shown in Fig. 17-35. Each plane corresponds to a difference density of scattering centers and a different spacing. If a screen is interposed in the path of the rays scattered by a single crystal (see Fig. 17-36), a regular pattern, which is characteristic of the crystal structure, appears. The pattern is called the *Laue pattern* after the German physicist Max von Laue (1879–1960). Each dot in the pattern corresponds to the direction of $\boldsymbol{u}_s$ related to the different families of planes illustrated in Fig. 17-35. The photograph of Fig. 17-37 shows one such Laue pattern.

If the scatterer, instead of being a single crystal, is a powder containing a large number of small crystals, all randomly oriented, the corresponding $\boldsymbol{u}_s$ vectors are distributed on conical surfaces about the direction of incidence as shown in Fig. 17-38. On a photographic film, each conical surface produces a bright ring as shown in Fig. 17-39, and the result is the so-called *Debye-Scherrer patterns*, named after the Dutch physicist Peter Debye (1884–1966) and his student, the Swiss Paul Scherrer (1890–1979). By analysis of patterns such as those of Figs. 17-37 and 17-39, the internal structure of a crystal may be deduced, or conversely the wavelength of the X-rays may be found.

It is interesting to note that when Roentgen observed X-rays for the first time at the end of the nineteenth century, a great argument arose about their nature. Were

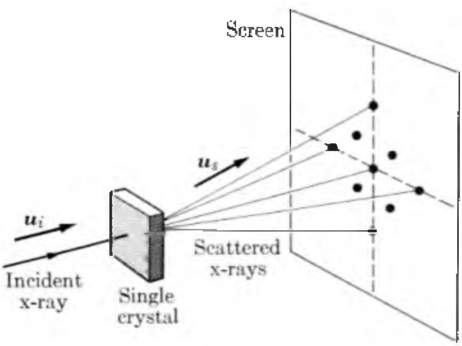


Fig. 17-36. Laue X-ray diffraction by a single crystal.

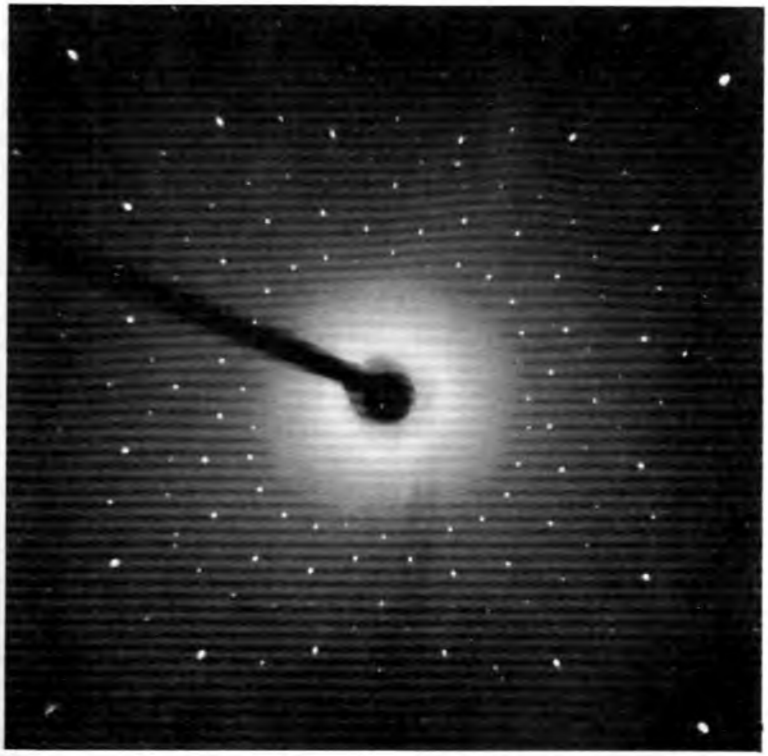


Fig. 17-37. Laue diffraction pattern for a quartz crystal. An attempt has been made to mask the effect of the undeviated incident beam.

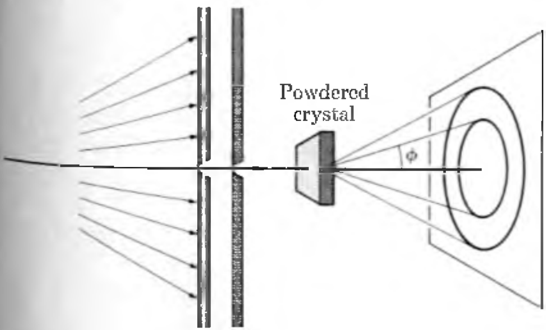


Fig. 17-38. Powder X-ray diffraction.

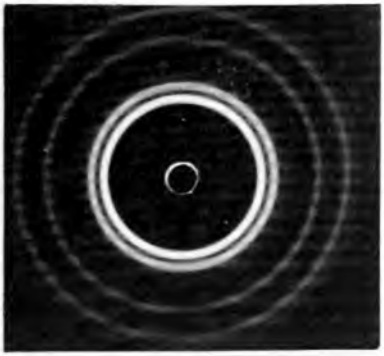


Fig. 17-39. X-ray diffraction pattern for powdered aluminum.

they waves or particles? To answer this question, physicists performed interference and scattering experiments with equipment similar to that used for experiments dealing with light. However, the results were either negative or unconvincing. The trend was to discard any wave interpretation until von Laue, the Braggs, and others studied the passage of X-rays through crystals, and obtained the results we have discussed. These results offered proof of the wave character of X-radiation.

Example 17.8. A beam of X-rays is diffracted by a rock-salt crystal. The first-order spectrum corresponds to an angle of $6°50'$ and the distance between the planes is 2.81×10^{-10} m. Determine the wavelength of the X-rays and the position of the second-order spectrum.

▼ Using Bragg's relation (17.32) with $d = 2.81 \times 10^{-10}$ m, $\theta = 6°50'$, and $n = 1$, we find that

$$\lambda = 2d \sin \theta = 6.69 \times 10^{-11} \text{ m}.$$

To find the position of the second-order spectrum, we set $n = 2$. Thus $\sin \theta = n\lambda/2d = 0.238$ or $\theta = 13°46'$. Note that the maximum diffraction order is limited by the condition $n\lambda/2d < 1$, which in our case amounts to $n < 8.4$ or $n_{max} = 8$. ▲

Problems

17.1 Parallel rays of green mercury light of wavelength 5.6×10^{-7} m pass through a slit of width 4×10^{-4} m covering a lens of focal length 0.40 m. What is the distance from the central maximum to the first minimum on a screen at the focal plane of the lens?

17.2 The Fraunhofer diffraction pattern of a single slit, reproduced at twice its size in Fig. 17-5, was formed on a photographic film in the focal plane of a lens of focal length 0.60 m. The wavelength of the light used was 5.9×10^{-7} m. Compute the width of the slit. (*Hint:* Measure (on the photograph) the distance between corresponding minima on the right and the left of the central maximum.)

17.3 A telescope is used to observe two distant point sources 0.30 m apart. The objective of the telescope is covered with a screen in which there is a slit of width 10^{-3} m. What is the maximum distance at which the two sources will be distinguishable? Assume $\lambda = 5.0 \times 10^{-7}$ m.

17.4 The Fraunhofer diffraction pattern of a single slit is observed in the focal plane of a lens of focal length 1 m. The width of the slit is 4×10^{-4} m. The incident light contains two wavelengths, λ_1 and λ_2. The fourth minimum corresponding to λ_1 and the fifth minimum corresponding to λ_2 occur at the same point, 5×10^{-3} m from the central maximum. Compute λ_1 and λ_2.

17.5 A plane monochromatic wave of wavelength 6.0×10^{-7} m is incident perpendicularly on an opaque screen that has a rectangular aperture of 5×10^{-4} m $\times$ 10^{-3} m. (a) Describe the diffraction pattern observed in the focal plane of a converging lens of focal length 2 m placed directly behind the aperture. (b) Compute the sides of the rectangle formed by the dark lines surrounding the central maximum.

17.6 Compute the radius of the central disk of the Fraunhofer diffraction pattern of the image of a star formed by (a) a camera lens 2.5×10^{-2}

m in diameter and focal length 7.5×10^{-2} m, (b) a telescope objective 0.15 m in diameter, with a 1.5-m focal length. Assume light of wavelength 5.6×10^{-7} m.

17.7 The headlights of an approaching automobile are 1.30 m apart. Estimate the distance at which the two headlights can be resolved by the naked eye if the resolution of the eye is determined by diffraction alone. Assume a mean wavelength of 5.5×10^{-7} m and assume that the diameter of the pupil of the eye is 5×10^{-3} m. Compare with the result obtained by the resolving power of the eye as given in Section 15.5.

17.8 In Fig. 17-40, two point sources of light, S_1 and S_2, 6×10^{-3} m apart and both at a distance of 50 m from lens L, produce images that are just resolved by Rayleigh's criterion. The focal length of the lens is 0.20 m. What is the diameter of the first diffraction circles?

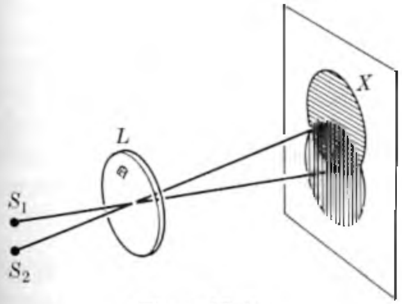


Figure 17-40

17.9 In a double-slit diffraction pattern the third principal maximum is missing because that interference maximum coincides with the first diffraction zero. (a) Find the ratio a/b. (b) Plot the intensity distribution over several maxima on either side of the central maximum. (c) Make a half-tone sketch of the fringes as they would appear on a screen.

17.10 Two pinholes 1.5×10^{-3} m apart are placed in front of a bright light source and viewed through a lens covered by a screen that has a central circular hole (aperture) with a diameter of 4×10^{-3} m. What is the maximum distance at which the pinholes can be resolved? Assume a wavelength of 5.5×10^{-7} m.

17.11 The Fraunhofer diffraction of a double slit is observed in the focal plane of a lens of focal length 0.50 m. The incident monochromatic light has a wavelength of 5.0×10^{-7} m. It is found that the distance between the two minima adjacent to the maximum of order zero is 5×10^{-3} m, and the maximum of the fourth order is missing. Compute the width of the slits and the distance between their centers.

17.12 Plane monochromatic waves of wavelength 6.0×10^{-7} m are incident normally on a plane transmission grating having 5×10^5 lines per m. Determine the angles of deviation for the (a) first-order, (b) second-order, and (c) third-order spectra.

17.13 A plane transmission grating is ruled with 4×10^5 lines per m. Compute in degrees the angular separation in the second-order spectrum between the α and β lines of atomic hydrogen, whose wavelengths are, respectively 6.56×10^{-7} m and 4.10×10^{-7} m. Assume normal incidence.

17.14 (a) What is the wavelength of light deviated in the first order through an angle of 20° by a transmission grating having 6×10^5 lines per m? (b) What is the second-order deviation of this wavelength? Assume normal incidence.

17.15 What is the longest wavelength that can be observed in the fourth order for a transmission grating having 5×10^5 lines per m? Assume normal incidence.

17.16 Assuming that the limits of the visible spectrum are at wavelengths of 4×10^{-7} m and 7×10^{-7} m, find the angles subtended by the first- and second-order spectra produced by a plane grating having 6×10^5 lines per m. Assume normal incidence.

17.17 A transmission grating 4×10^{-2} m long has 4×10^5 lines per m. (a) Compute the resolving power of the grating for a wavelength of 5.9×10^{-7} m in the first-order spectrum. (b) Will the grating separate the two lines of wavelength 5.890×10^{-7} m and 5.896×10^{-7} m that constitute the sodium yellow doublet? Also compute (c) the minimum deviation and (d) the corresponding dispersion for the wavelength considered.

17.18 Show that, no matter what the grating spacing, the violet of the third-order spectrum overlaps the red of the second-order spectrum. Assume normal incidence.

17.19 Monochromatic light of wavelength 6.0×10^{-7} m, originating at a distant point source, passes through a circular opening. The Fresnel diffraction pattern is observed on a screen 1 m beyond the opening. Determine the diameter of the circular aperture if it exposes (a) the central Fresnel zone only, (b) the first four Fresnel zones.

17.20 A point is placed 0.10 m from a circular aperture illuminated by light of wavelength 5.0×10^{-7} m. If the aperture corresponds to 10 Fresnel zones, determine its radius.

17.21 Light of wavelength 5.0×10^{-7} m falls on a circular aperture of 10^{-5} m radius. At what distance from the aperture should a point be located so that the aperture corresponds to (a) three Fresnel zones, (b) four Fresnel zones? (c) Estimate in each case whether there will be brightness or darkness at that point.

17.22 Parallel light waves of wavelength 5.6×10^{-7} m pass through a circular aperture 2.60×10^{-3} m in diameter. The Fresnel diffraction pattern is observed on a screen 1 m from the aperture. (a) Will the center of the diffraction pattern appear bright or dark? (b) What minimum distance should the screen be moved in order to reverse the condition found in (a)?

17.23 A plane monochromatic light wave of wavelength $\lambda = 5.0 \times 10^{-7}$ m is incident perpendicularly on a screen that has a circular aperture 4×10^{-3} m in diameter. (a) Determine the positions of the points of minimum and maximum intensity along the axis of the screen. (b) How far from the screen does the last minimum occur?

17.24 A screen having a circular aperture of radius 4×10^{-3} m is illuminated with plane light waves falling perpendicularly. Assume that the incident light is a mixture of two monochromatic light beams of wavelengths $\lambda_1 = 6.0 \times 10^{-7}$ m and $\lambda_2 = 4.0 \times 10^{-7}$ m, respectively. Determine the points on the line perpendicular to the aperture and through its center, where only (a) λ_1 and (b) λ_2 is observed.

17.25 The spacing between the principal planes in a NaCl crystal is 2.82×10^{-10} m. It is found that a first-order Bragg reflection of a monochromatic X-ray beam occurs at an angle of 10°. (a) Compute the wavelength of the X-rays. (b) What angle corresponds to the second-order spectrum?

17.26 Potassium iodide, KI, is a cubic crystal having a density of 3.13×10^3 kg m^{-3}. (a) Find the smallest interplanar distance; i.e., the length of a unit cell. (b) Determine the angles corresponding to the first two Bragg reflections for X-rays of wavelength 3.0×10^{-10} m.

17.27 An X-ray tube accelerates electrons through a potential difference of 10^5 V. The X-rays produced are examined by means of the crystal described in Problem 17.25. Find the angle at which the first-order spectrum of the shortest wavelength produced by the tube occurs. (*Hint:* See Example 12.2.)

17.28 A beam of X-rays, of wavelength 5×10^{-11} m, falls on a powder composed of microscopic crystals of KCl oriented at random. The lattice spacing in the crystal is 3.14×10^{-10} m. A photographic film is placed 0.1 m from the powder target. Find the radii of the circles corresponding to the first- and second-order spectra from planes having the same spacing as the lattice spacing.

CHALLENGING PROBLEMS

17.29 Coherent light passes through two parallel slits and then falls on a screen 10 m distant. The light has a wavelength of 6.0×10^{-7} m; each slit is 7.5×10^{-5} m wide; the center lines of the slits are 1.5×10^{-4} m apart. (a) First one of the slits is covered, so that the light passes

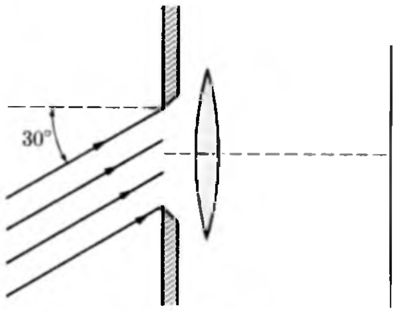


Figure 17-41

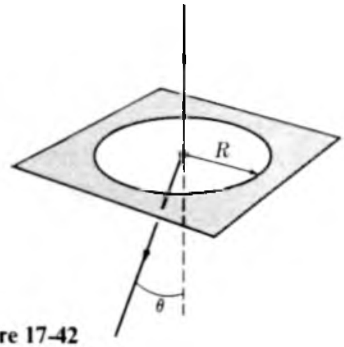


Figure 17-42

through a single slit. Describe as quantitatively as you can the appearance of the pattern of light on the screen. (b) Now both slits are uncovered, so that the light passes through both. Describe as quantitatively as you can the appearance of the pattern of light on the screen. (c) Suppose the wavelength of the light passing through the slits is decreased. How will the pattern observed in part (b) change? [AP-B; 1972]

17.30 (a) Light of a single wavelength is incident on a single slit of width w. (w is a few wavelengths.) Sketch a graph of the intensity as a function of position for the pattern formed on a distant screen. (b) Repeat (a) for the case in which there are two slits. The slits are of width w and are separated by a distance $d(d \gg w)$. Sketch a graph of the intensity as a function of position for the pattern formed on a distant screen. [AP-B; 1975]

17.31 A plane monochromatic wave of wavelength λ is incident, at an angle of 30°, on a plane opaque screen that has a long narrow slit of width a (Fig. 17-41). Behind the screen is a converging lens whose principal axis is perpendicular to the plane of the screen. Describe the diffraction pattern observed in the focal plane of this lens.

17.32 Discuss the intensity distribution of the Fraunhofer diffraction by three identical, equally spaced slits. Assume normal incidence on the slits.

17.33 Two equally bright stars subtend an angle of one second. Assuming a wavelength of 5.5×10^{-7} m: (a) What is the smallest diameter of a telescope objective lens that will permit these stars to be resolved? (b) What should be the magnifying power of the telescope? (c) Compute the focal length of the eyepiece to be used if the focal length of the objective is 1.80 m.

17.34 It can be shown that, in the case of Fraunhofer diffraction, the amplitude of the waves diffracted by a circular aperture of radius R is proportional to the *Bessel function* $J_1(x)$ (see Chemical Rubber Company's Standard Mathematical Tables, 25th edition, page 416), where for normal incidence

$$x = \left(\frac{2\pi R}{\lambda}\right)\sin\theta,$$

and θ is the angle the diffracted rays make with the axis of the aperture (Fig. 17-42). (a) Show that the directions for which the diffracted waves have zero amplitude correspond to the roots of the equation $J_1(x)=0$. (b) By looking at a table of roots of $J_1(x)=0$ (*loc. cit.*, page 417), obtain the values of $\sin\theta$ for the first three directions of zero amplitude, checking Eq. (17.11). (c) Assuming that the diffracted rays are focused by a convergent lens of focal length f on a screen at the focal plane of the lens, express the radii of the first three dark rings formed. (Note that in this problem $\sin\theta$ can be replaced by θ.) (d) Obtain the values of θ and of the radii of the rings, given that $R=10^{-4}$ m, $\lambda=5.9 \times 10^{-7}$ m, and $f=0.20$ m.

17.35 Show that, in a grating having a large number of lines, the intensity of the first secondary maximum on either side of the first

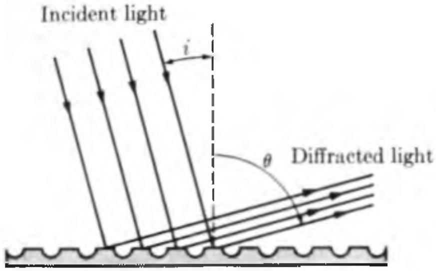


Figure 17-43

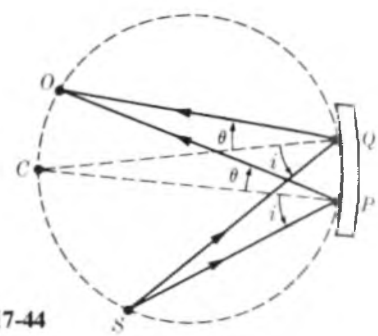


Figure 17-44

principal maximum is equal to about 4% of the intensity of the principal maximum.

17.36 A *reflection grating* is made by using a diamond point to etch fine lines on a polished metal surface (Fig. 17-43). The polished spaces left between adjacent rulings are the equivalent of the slits in a transmission grating. Show that the principal maxima are obtained by the condition

$$a(\sin i - \sin \theta) = n\lambda,$$

where a is the separation between consecutive lines.

17.37 To assure proper focusing by a diffraction grating, the American physicist H. A. Rowland constructed concave gratings of large radii. Assume that C in Fig. 17-44 is the center of curvature of the grating, and that the dashed circle has a diameter equal to the radius of the grating. Show that for any source S placed on that circle, (a) all rays fall on the grating with the same angle of incidence, (b) all rays diffracted by the grating through the same angle converge on some point O on the dashed circle. Thus if a photographic plate is placed at O, tangent to the circle, the diffraction spectrum corresponding to that diffraction angle can be recorded. This arrangement is called *Rowland's mounting*, and is used widely in physics laboratories for spectroscopic research. (*Hint:* Note that the normal to the grating at the point of incidence of a ray passes through C, and that the grating's surface departs very little from that of the dashed circle.)

17.38 A plane monochromatic light wave of wavelength $\lambda = 5.0 \times 10^{-7}$ m is incident perpendicularly on an opaque screen that has an aperture of the shape shown in Fig. 17-45. The radius of the inner circle is 10^{-3} m and that of the outer circle is 1.41×10^{-3} m. (a) Compute the amplitude and the intensity of the optical disturbance at a point P on the axis of the circles, 2 m from the screen, relative to the values one would obtain in the absence of the screen. (b) Determine the phase of this disturbance relative to that of the disturbance one would observe at P without the screen.

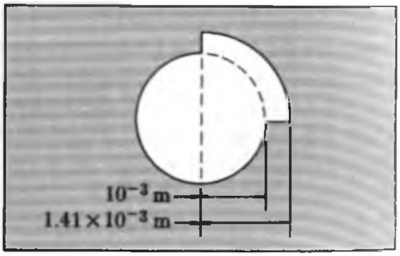


Figure 17-45

17.39 A crystal lattice may be characterized by three fundamental vectors, $\boldsymbol{a}_1$, $\boldsymbol{a}_2$, $\boldsymbol{a}_3$, so that the crystal structure is periodic for displacements that are linear combinations of integral multiples of the three vectors (Fig. 17-46). (a) Show that the relative position vectors of two points occupying similar positions in two different cells are given by $\boldsymbol{r} = \gamma_1\boldsymbol{a}_1 + \gamma_2\boldsymbol{a}_2 + \gamma_3\boldsymbol{a}_3$, where γ_1, γ_2, and γ_3 are positive or negative integers. (b) Show that the atoms participating in the X-ray diffraction spectrum of order n are

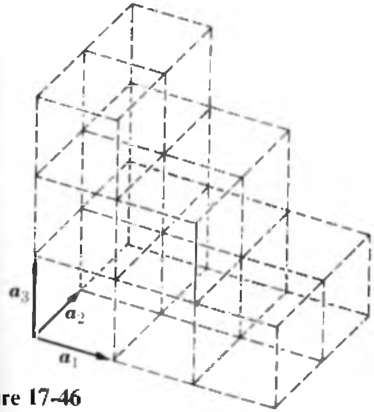


Figure 17-46

given by the integers satisfying the equation $\boldsymbol{v}\cdot(\gamma_1\boldsymbol{a}_1+\gamma_2\boldsymbol{a}_2+\gamma_3\boldsymbol{a}_3)=n\lambda$, where $\boldsymbol{v}$ is as defined in Eq. (17.29). (c) Show that the intensity of the radiation scattered in the direction associated with $\boldsymbol{v}$ is proportional $(A_1A_2A_3)^2$, where $A_i = \sin(N_i\pi\boldsymbol{v}\cdot\boldsymbol{a}_i/\lambda)/\sin \pi\boldsymbol{v}\cdot\boldsymbol{a}_i/\lambda$ and N_i is the number of cells of the crystal in the direction of $\boldsymbol{a}_i$. (d) From the result derived in (c), show that the principal maxima occur in a direction satisying the relations $\boldsymbol{v}\cdot\boldsymbol{a}_1=n_1\lambda$, $\boldsymbol{v}\cdot\boldsymbol{a}_2=n_2\lambda$, $\boldsymbol{v}\cdot\boldsymbol{a}_3=n_3\lambda$, where n_1, n_2, and n_3 are integers. These relations are called the *Laue equations*. (e) Using the reciprocal vectors $\boldsymbol{a}^1$, $\boldsymbol{a}^2$, $\boldsymbol{a}^3$ (see Problem 3.34 in Volume I), show that

$$\boldsymbol{u}_s=\boldsymbol{u}_i+(n_1\boldsymbol{a}^1+n_2\boldsymbol{a}^2+n_3\boldsymbol{a}^3)\lambda.$$

This equation determines the position of the bright dots in a Laue pattern, as shown in Fig. 17-37.

CHAPTER EIGHTEEN

QUANTUM MECHANICS

18.1 Introduction

The motion of the bodies we observe around us can be described (independently of the interactions among them) in terms of general rules based on experimental evidence. These principles or laws are:

1. the conservation of momentum,
2. the conservation of angular momentum, and
3. the conservation of energy.

Based on these conservation laws, a formalism, called *classical mechanics*, was developed for describing the detailed motion of particles under the assumptions that (1) the particles are localized in space, and (2) we can observe them without appreciably disturbing their motion. These assumptions are, in general, made implicitly rather than precisely and explicitly. The formalism of classical mechanics has been used to describe and analyze the motions of various bodies, ranging in size from planets at one extreme down to electrons at the other. However when applied to the motion of the basic constituents of matter, classical mechanics gives only approximate results; in some instances classical mechanics is entirely inadequate.

At the end of the first quarter of the twentieth century an important revolution in physical concepts radically changed our approach to the description of the motion of atomic and subatomic particles. Although the laws of conservation of momentum, angular momentum, and energy remain valid, there is now agreement that a detailed description of the motion of atomic particles, in the sense of classical mechanics, is not possible. For example, the experimental fact of *quantization* of energy and other physical quantities, discussed in Chapter 7, is a novel idea that does not appear in classical mechanics. A satisfactory theory must contain information on how to calculate allowed energy levels as well as the space distribution of the particles involved. The interaction of radiation and matter by means of the absorption or emission of photons (Chapter 12) is yet another concept that must be incorporated into this new theory.

For these reasons a new formalism, called *quantum mechanics*, has been developed. Quantum mechanics in its present form is the result of the work of many people. Among those who developed the new mechanics in the late 1920s, the Germans Max Born (1882–1969), Werner Heisenberg (1901–1976), and Erwin Schrödinger (1897–1961), the Englishman Paul Dirac (1902–), and the Frenchman Prince Louis de Broglie (1892–) stand out as notable contributors. The theoretical framework of quantum mechanics is mathematically elaborate, but its basic ideas are relatively simple.

18.2 Particles and Fields

Our sensory experience tells us that the objects we touch and see have well-defined shapes and sizes and therefore are localized in space. We thus tend to think of the fundamental particles (i.e., electrons, protons, neutrons, etc.) as having shape and size, and imagine them as small spheres with a characteristic radius as well as mass and charge. This conceptualization, however, is an extrapolation beyond our direct sensory experience; and we should analyze this picture carefully before we accept it.

Experimentation shows that our extrapolated picture of the basic constituents of matter is erroneous. The dynamical behavior of elementary particles requires that we associate with each particle a field—a *matter field*—in the same way that in the reverse manner we associate a photon (which is equivalent to a particle) with an electromagnetic field. This matter field describes the dynamical condition of a particle in the same sense that the electromagnetic field describes photons that have precise momentum and energy. In discussing the connection between the matter field and the dynamical properties of the particle (i.e., momentum and energy), we may be guided by the relations $E=h\nu$ and $p=h/\lambda$, previously found for the photon in Chapter 12. The quantity h is Planck's constant, equal to 6.626×10^{-34} Js. Writing these relations in reverse, we may assume that the wavelength λ and the frequency ν of the monochromatic field associated with a particle of momentum p and energy E are given by

$$\lambda=\frac{h}{p} \qquad \nu=\frac{E}{h}. \tag{18.1}$$

These relations were first proposed in 1924 by de Broglie; therefore $\lambda=h/p$ is sometimes called the *de Broglie wavelength* of a particle. Introducing the wave number $k=2\pi/\lambda$ and the angular frequency $\omega=2\pi\nu$, we may write the relations in the more symmetric form

$$p=\frac{h}{2\pi}k, \qquad E=\frac{h}{2\pi}\omega;$$

or recalling that $\hbar=h/2\pi=1.0546\times10^{-34}$ Js, we have

$$p=\hbar k, \qquad E=\hbar\omega. \tag{18.2}$$

If our assumption as expressed by Eqs. (18.1) or (18.2) is correct, we may expect that whenever the motion of a particle is disturbed in such a way that the matter field associated with the particle cannot propagate freely, interference, diffraction, and scattering of the matter waves should be observed as is the case for elastic and electromagnetic waves. This expectation is indeed what happens.

Before discussing the experiments that reveal interference and diffraction of the matter field, let us estimate the value of the de Broglie wavelength λ associated with a particle in terms of its energy. Electrons accelerated by an electric potential V gain an energy eV; hence their kinetic energy is

$$\frac{p^2}{2m_e}=eV \qquad \text{so that} \qquad p=\sqrt{2m_e eV}.$$

Therefore introducing the values of e, m_e, and k, we obtain the de Broglie wavelength of such accelerated electrons as

$$\lambda = \frac{h}{\sqrt{2m_e eV}} = \frac{1.23 \times 10^{-9}}{\sqrt{V}} \text{ m} \tag{18.3}$$

where V is expressed in volts. Similarly when the kinetic energy of the electron is expressed in electron volts, $E_k = p^2/2m_e$ or $p = \sqrt{2m_e E_k}$; and one may write

$$\lambda = \frac{1.23 \times 10^{-9}}{\sqrt{E_k}} \text{ m},$$

where E_k is expressed in electron volts.

18.3 Scattering of Particles by Crystals

Consider electrons with energy of the order of 10^4 V (this is in the range of the voltage used in TV tubes). Equation (18.3) indicates that the wavelength of the electrons is about 10^{-11} m, comparable to the wavelength of X-rays. Thus if we send a beam of fast electrons through a crystal, we should obtain patterns that result from scattering of the matter field. These patterns corresponding to the interaction of the incoming electrons should be similar to those observed for X-rays of the same wavelength as discussed in Section 17.8.

In 1927 the British scientist G. P. Thomson (1892–1975) began a series of experiments to study the passage of a beam of electrons through a thin film of crystalline material. After the electrons passed through the film, they struck a photographic plate as shown in Fig. 18-1. If the electrons had behaved as particles in the macroscopic sense, a blurred image would have been observed because each electron generally would undergo a different scattering by the atoms in the crystal. However, the result obtained was identical to the Debye-Scherrer patterns (see Ch. 17) for X-ray scattering by a polycrystalline substance as indicated in the photograph of Fig. 18-2. Similarly

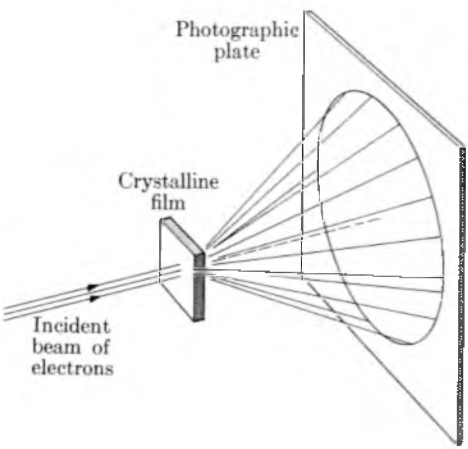


Fig. 18-1. Experimental arrangement for observing electron scattering through crystalline material.

Fig. 18-2. Scattering of electrons by crystal powder. (Courtesy of Dr. Lester Germer)

when an electron beam passes through a single crystal, Laue spot patterns (also observed with X-rays) are produced as seen in the photograph of Fig. 18-3. From the structure of these patterns we can compute the de Broglie wavelength λ when the spacing between the crystal planes is known and if the formulas derived for X-rays are applied. The resulting values of λ can be compared with those obtained from Eq. (18.3). The result is complete agreement within the limits of experimental error.

In the celebrated experiments (made at about the same time as those of Thomson) by the Americans C. Davisson (1881–1958) and L. Germer (1896–1971), a beam of electrons was sent at an angle to the face of a crystal. The scattered electrons were

Fig. 18-3. Scattering of electrons by a single carbon (graphite) crystal. (Courtesy of R. Heidenreich, Bell Telephone Laboratories)

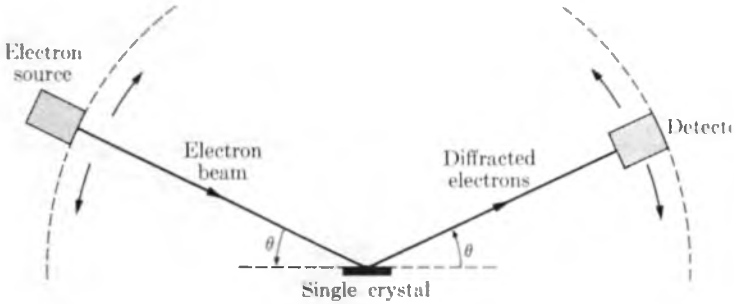


Fig. 18-4. Davisson and Germer arrangement for observing Bragg scattering of electrons.

observed by means of a detector symmetrically located as indicated in Fig. 18-4. This arrangement is similar to the Braggs' for observing X-ray scattering (Section 17.8). Davisson and Germer found that the electron current registered by the detector was a maximum every time the Bragg condition, derived for X-rays, was fulfilled. The Bragg condition is expressed by Eq. (17.31),

$$2d \sin \theta = n\lambda \tag{18.4}$$

where d is the separation of successive atomic layers in the crystal, and λ is given by Eq. (18.3).

The same phenomenon of Bragg scattering has been observed in experiments with protons and neutrons. Neutron scattering is especially useful since it is one of the most powerful means of studying crystal structure. Experimenters use monoenergetic beams of neutrons and analyze their passage through the crystal. The neutrons emerging from a nuclear reactor through a porthole (Fig. 18-5) have a wide spectrum of energy (that is, they vary widely in momentum). In other words, the neutron beam is not monochromatic; rather it contains a spectrum composed of many de Broglie wavelengths. When the neutron beam from the reactor falls on a crystal (LiF, for example), the neutrons observed in the symmetric direction correspond only to the wavelength λ given by Bragg's condition (18.4). Therefore they have a well-defined energy and momentum. The crystal then acts as an *energy filter* or *monochromator*.

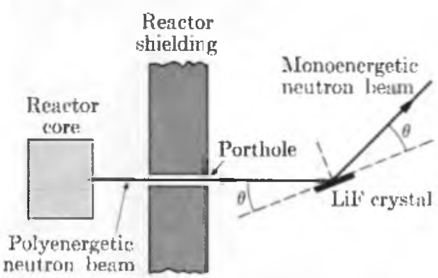


Fig. 18-5. Neutron crystal spectrometer.

The monoenergetic neutron beam is in turn used to study other materials or to analyze nuclear reactions involving neutrons.

Example 18.1. The de Broglie wavelength of thermal neutrons at a temperature of 300 K.

▼ By *thermal neutrons* we mean neutrons that are in thermal equilibrium with matter at a given temperature. Thus the neutrons have an average kinetic energy identical to that of the molecules of an ideal gas at the same temperature. Therefore the average kinetic energy of thermal neutrons is $E_{ave}=(3/2)kT$ where T is the absolute temperature and k is Boltzmann's constant. Given that the temperature is 300 K,

$$E_{ave}=\frac{3}{2}kT=6.21\times10^{-21}\ \mathrm{J}=3.88\times10^{-2}\ \mathrm{eV}.$$

The corresponding momentum is

$$p=\sqrt{2m_n E_{ave}}=4.56\times10^{-24}\ \mathrm{m\ kg\ s^{-1}}.$$

Then using Eq. (18.1), we find that the average de Broglie wavelength of thermal neutrons is

$$\lambda=1.45\times10^{-10}\ \mathrm{m}.$$

Incidentally, noting that the separation of the planes in a NaCl crystal is $d=2.82\times10^{-10}$ m, we see that the first Bragg maximum for neutrons of this wavelength occurs at an angle $\theta=14.9^\circ$. ▲

18.4 Particles and Wave Packets

Using relations (18.1), we may represent the field corresponding to a free particle moving with a well-defined momentum p and energy $E=p^2/2m$ by a harmonic wave of constant amplitude as shown in Fig. 18-6. Symmetry demands that the amplitude of the wave be the same throughout all space since there are no forces acting on the particle to distort the associated matter field more in some regions of space than in others. The *phase velocity* of the field of the free particle is

$$v_p=\lambda\nu=\frac{h}{p}\frac{E}{h}=\frac{E}{p}=\frac{p}{2m}=\tfrac{1}{2}v.$$

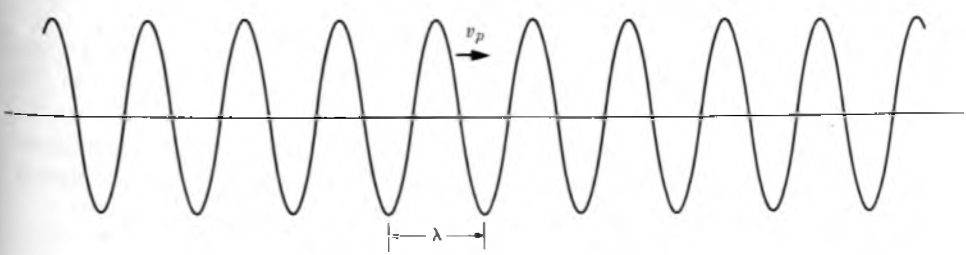


Fig. 18-6. Continuous wave train corresponding to an unlocalized packet.

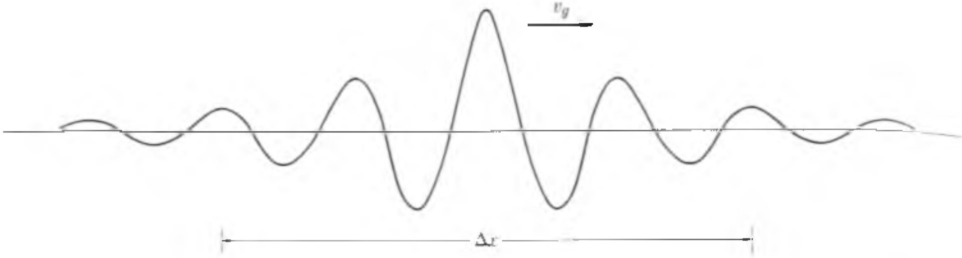


Fig. 18-7. Wave packet corresponding to a particle localized within the distance Δx.

That is, the phase velocity of the matter field is one-half the particle velocity. This equation has no experimental consequence, however, since we cannot measure the phase velocity of a pure harmonic wave directly. We can measure only the *group velocity* of the waves. The fact that the amplitude of the matter field is the same throughout all space suggests that the matter field of a free particle does not give information about the localization in space of a free particle of well-defined momentum. In other words, this matter field is independent of the position of the particle, and an observation of the field by some method would not reveal the position of the particle.

From our physical intuition and our knowledge of fields and waves, we know that a particle localized within a certain region Δx of space should correspond to a matter field whose amplitude or intensity is large in that region and very small outside it. A field may be built up in a certain region and attenuated outside that region through the process of interference by superposing waves of different frequencies and wavelengths. The result is a *wave packet* as shown in Fig. 18-7. The velocity with which the wave packet propagates is the group velocity v_g, defined by Eq. (10.67) as

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

Using relations (18.2) and $E = p^2/2m$, we may rewrite the group velocity of the matter field corresponding to a free particle as

$$v_g = \frac{dE}{dp} = \frac{p}{m} = v.$$

Thus just as our intuition tells us, the group velocity of the matter field (i.e., the velocity of propagation of the packet) is equal to the velocity of the particle. We conclude then that a particle localized in a certain region of space is associated with a field or wave packet whose amplitude is important only in the region occupied by the particle; the velocity of the particle is the group velocity of the field or wave packet.

18.5 Heisenberg's Uncertainty Principle for Position and Momentum

Now we encounter a special situation that cannot be explained in terms of classical mechanics. For a wave packet to be localized in space, it is necessary to superpose several fields of different wavelengths λ (or with different values of the wave number k). If the wave packet extends over a region Δx, the values of the wave numbers of the interfering waves composing the wave packet and having an appreciable amplitude fall within a range Δk such that according to the analysis developed in Example 10.3,

$$\Delta x \Delta k \sim 2\pi.$$

But according to Eqs. (18.1) or (18.2), different wavelengths λ or wave numbers k mean that there are several values of p such that

$$\Delta p = \hbar \Delta k.$$

Therefore when we recall that $h = 2\pi\hbar$, the expression above becomes

$$\Delta x \Delta p \sim h. \tag{18.5}$$

The physical meaning of Eq. (18.5) is this: if a particle is within the region $x - \frac{1}{2}\Delta x$ and $x + \frac{1}{2}\Delta$ (that is, Δx is the *uncertainty* in the position of the particle), its associated field is represented by superposing waves of momenta between $p - \frac{1}{2}\Delta p$ and $p + \frac{1}{2}\Delta p$ where Δp is related to Δx by Eq. (18.5). We say that Δp is the uncertainty in the momentum of the particle. Equation (18.5) implies that the larger Δx, the smaller Δp, and conversely. In other words, information about the localization of a particle in space is obtained at the expense of knowledge about the momentum. The more precise our knowledge of the position of the particle, the more imprecise is our information about its momentum, and conversely. For this reason a particle of well-known momentum ($\Delta p = 0$) is represented by a wave of constant amplitude extending over all space ($\Delta x \sim \infty$) so that our knowledge of the position is nil. We cannot accurately determine both the position and the momentum of a particle simultaneously so that $\Delta x = 0$ and $\Delta p = 0$ at the same time. Such knowledge does not conform with Eq. (18.5).

The result expressed by relation (18.5) is called *Heisenberg's uncertainty principle*, which may be stated thus:

> *It is impossible to know simultaneously and with exactness both the position and the momentum of a particle.*

This principle expresses one of the fundamental facts of nature and to a certain extent may be considered as more fundamental than Eqs. (18.2) even though we have here proceeded in the opposite manner.

The uncertainty principle implies that the path of a particle can never be defined with the absolute precision postulated in classical mechanics. Classical mechanics still holds true for large bodies, such as those of usual concern to the engineer, because the uncertainty implied by Eq. (18.5) is much smaller for a macroscopic body than the experimental errors in the measured values of x and p for the body. However, *for particles of atomic dimensions, the concept of trajectory has no meaning since it cannot be defined precisely*; therefore a picture of the motion different from the picture of classical physics is required. For the same reason, concepts such as velocity, acceleration, and force are of limited use in quantum mechanics. On the other hand, the concept of energy is of primary importance since energy is related more to the "state" of the system than to its "path."

18.6 Illustrations of Heisenberg's Principle

Some simple situations serve to illustrate Heisenberg's principle. Suppose, for example, that we want to determine the x coordinate of a particle by observing whether or not the particle passes through a hole (of width b) in a screen (Fig. 18-8). The precision with which we know the position of the particle is limited by the size of the hole; that is, $\Delta x \sim b$; but the hole disturbs the matter field associated with the particle, and the result is a corresponding change in the motion of the particle as shown by the diffraction pattern produced. The uncertainty in the particle's momentum parallel to the X-axis is determined by the angle θ, corresponding to half of the width of the central maximum of the diffraction pattern since the particle, after

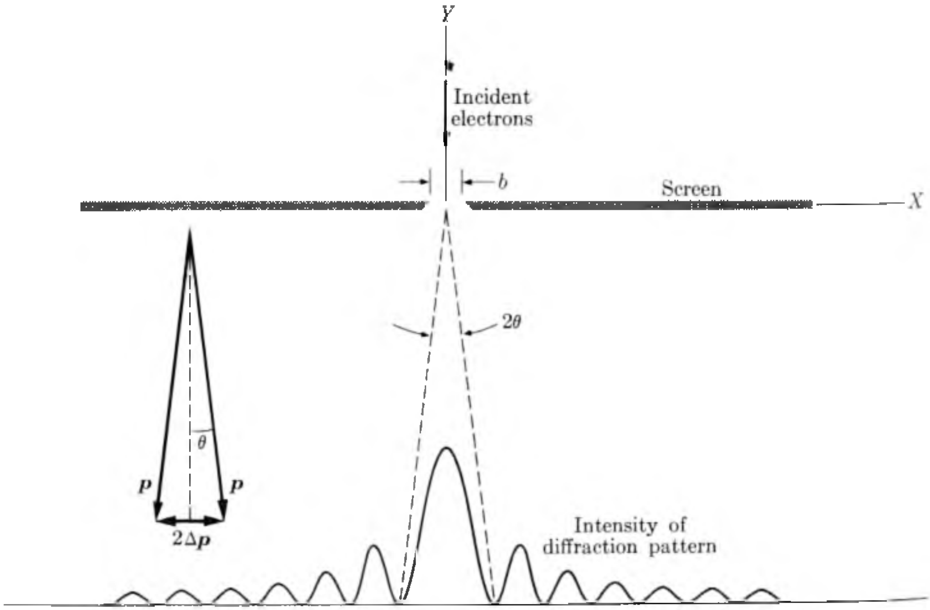


Fig. 18-8. Measurement of position and momentum of a particle passing through a slit.

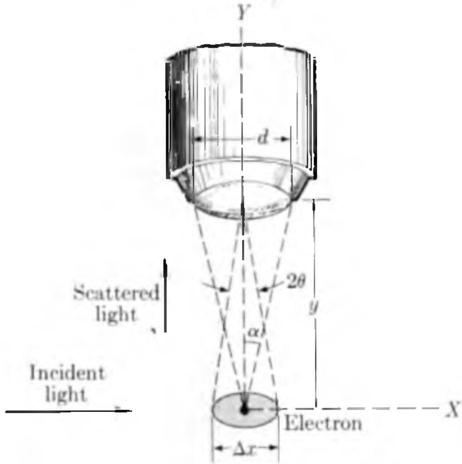


Fig. 18-9. Use of a microscope to measure position and momentum of a particle.

traversing the slit, is most probably moving within the angle 2θ. According to the theory of diffraction produced by a rectangular slit (Section 17.3), the angle θ is given by $\sin\theta = \lambda/b$. Then

$$\Delta p \sim p \sin\theta = \frac{h}{\lambda}\frac{\lambda}{b} = \frac{h}{b}$$

is the uncertainty in the momentum parallel to the X-axis. Therefore $\Delta x \Delta p \sim h$, in agreement with Eq. (18.5). Note that to improve our ability to determine the position of the particle, we must use a very narrow slit; but a very narrow slit produces a very wide central maximum in the diffraction pattern, and therefore a large uncertainty in our knowledge of the X component of the momentum of the particle. Conversely in order to reduce the uncertainty in our knowledge of the X component of the momentum, the central maximum in the diffraction pattern must be very narrow. This condition requires a very wide slit which, in turn, results in a large uncertainty in the x coordinate of the particle.

Another situation illustrating Heisenberg's principle is the case in which we try to determine the position of an electron by means of a microscope (Fig. 18-9). To observe the electron, we must illuminate it with light of some wavelength λ. The light that passes into the microscope has been scattered by the electron under observation. The momentum of the scattered photons is $p_{\text{photons}} = h/\lambda$; and to enter the objective lens, the photons must move within the cone of angle α so that the X component of their momenta has an uncertainty

$$\Delta p \sim p_{\text{photons}} \sin\alpha \sim \frac{h}{\lambda}\frac{d}{2y}$$

since $\sin\alpha \sim d/2y$. This is also the uncertainty in the X component of the electron momentum after the scattering of light since in the scattering process some momentum is exchanged between the electron and the photon. On the other hand, the exact position of the electron is uncertain because of the diffraction of light when it passes through the objective of the microscope. The uncertainty in the position of the electron is thus equal to the diameter of the central disk in the diffraction pattern. This diameter is given by $2y \sin\theta$ with $\sin\theta \sim \lambda/d$.* Hence

$$\Delta x \sim 2y \sin\theta \sim \frac{2y\lambda}{d}.$$

Therefore again $\Delta x \Delta p \sim h$. Note that to improve the accuracy of our knowledge in the position of the electron, we must use a radiation of very small wavelength, but the result is a large disturbance in the momentum. Conversely in order to produce a small disturbance in the momentum, we must use radiation of very long wavelength, which in turn gives rise to a great uncertainty in the position.

These two examples clearly show that the uncertainty principle is a direct consequence of the process of measurement. At the atomic level, measurement inevitably introduces a significant perturbation in the system because of the interaction between the measuring device and the measured quantity.

18.7 The Uncertainty Relation for Time and Energy

In addition to the uncertainty relation $\Delta x \Delta p \sim h$ between a coordinate and the corresponding momentum of a moving particle, there is an uncertainty relation between time and energy. Suppose that we want to measure not only the energy of a particle but also the time at which the particle has such energy. If Δt and ΔE are the uncertainties in the values of these quantities, we will show that the relation

$$\Delta t \Delta E \sim h \tag{18.6}$$

holds. We can understand Eq. (18.6) in the following way. If we want to define the time at which a particle passes through a given point, we must represent the particle by a pulse or wave packet having a very short duration Δt. However, to build such a pulse, it is necessary to superpose fields of many different frequencies with an amplitude appreciable only in a frequency range $\Delta\omega$ centered around the frequency ω related to the energy of the particle by Eq. (18.2), $E = \hbar\omega$. The theory of superposition of waves as developed in Ex. 10.3 requires that

$$\Delta t \Delta\omega \sim 2\pi.$$

Multiplying by $\hbar$ and recalling that $E = \hbar\omega$ and $2\pi\hbar = h$, we obtain Eq. (18.6).

The uncertainty relation (18.6) requires that we revise our concept of stationary states. Consider an electron in an excited stationary state of an atom. After a certain time the electron will suffer a radiative transition into another stationary state of less energy. However, we have no means of predicting with certainty how long the electron

*We have disregarded the factor 1.22 that appears in the theory of diffraction of a plane wave by a circular aperture (Section 17.3).

will remain in the stationary state before making the transition. The most we can discuss is the probability per unit time that the electron will jump into a lower energy state. Therefore the *lifetime* of the state, the average length of time the electron is in a stationary state, is known within an uncertainty Δt. Hence the energy of the stationary state of the electron is not known precisely but has an uncertainty ΔE such that Eq. (18.6) holds. Often ΔE is designated as the *energy width* of the state whose energy is most probably between the limits $E-\frac{1}{2}\Delta E$ and $E+\frac{1}{2}\Delta E$. Thus the shorter the lifetime of an excited state, the larger the uncertainty in the energy of the state. For the ground state, whose lifetime is infinite because a system that is in its ground state cannot suffer a transition to a stationary state of lower energy, we have $\Delta t \sim \infty$. This yields $\Delta E=0$, and the energy of the ground state can be determined accurately.

18.8 Stationary States and The Matter Field

We are now in a position to give a theoretical justification for the idea of stationary states that had been introduced in Section 7.4.

When a particle is in a bound state and confined to move within a limited region of space, such as an electron in an atom or a proton in a nucleus, the associated matter field must also be confined to that region. The situation is similar to that of waves on a string with fixed ends or within a cavity. We know that in such cases only certain wavelengths are possible and the allowed waves are called standing waves. Therefore we may expect that in the case of a bound particle only the states corresponding to the allowed wavelengths of the matter field are possible.

Consider the very simple example of a particle, such as a gas molecule in a box, constrained to move in the region from $x=0$ to $x=a$. The molecule moves freely until it hits the wall, which forces the molecule to bounce back. The situation for a free electron in a piece of metal is similar if we neglect the electron's interactions with the positive ions and if the height of the potential barrier at the metal surface is much larger than the electron's kinetic energy. The electron can move freely through the metal but cannot escape from it.

We may represent each of these physical situations by the rectangular potential energy diagram of Fig. 18-10. This is an oversimplification of the potential energies that actually occur in nature. The simplified potential-energy diagram is called a *potential box*. We have

$$E_p(x)=0 \qquad \text{for} \qquad 0<x<a \tag{18.7}$$

since the particle moves in that region. The potential energy increases sharply to infinity at $x=0$ and $x=a$. This increase means that very strong forces act on the particle at those two points and force the particle to reverse its motion. Then no matter what the value of the energy E, the particle cannot be to the left of $x=0$ or to the right of $x=a$. The situation is formally identical to that corresponding to standing waves on a string with fixed ends. The student may recall that in order to have standing

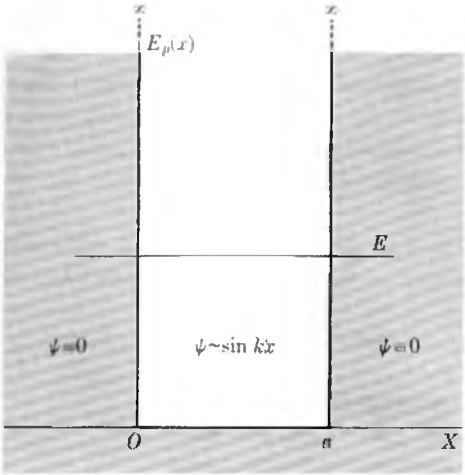


Fig. 18-10. One-dimensional potential box of width a.

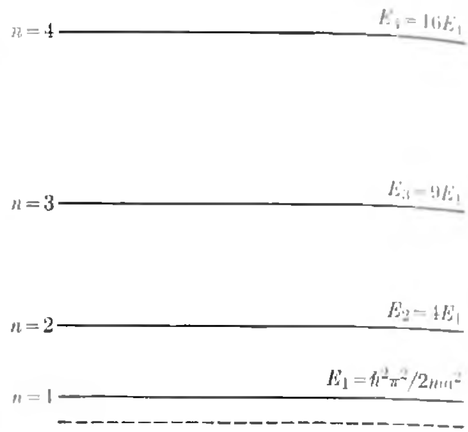


Fig. 18-11. Energy levels for a one-dimensional potential box.

waves on a string with fixed ends a distance a apart, the wavelength λ must have the values $\frac{1}{2}\lambda = a, \frac{1}{2}a, (1/3)a, \ldots, (1/n)a$ or

$$\lambda = \frac{2a}{n}, \qquad n = 1, 2, 3, \ldots. \tag{18.8}$$

We may thus assume that the same relation applies to the wavelength of the matter field of a particle within a potential box of length a. Hence according to Eq. (18.1), the only possible values of the momentum of the particle are

$$p = \frac{h}{\lambda} = \frac{nh}{2a} = \frac{n\pi\hbar}{a}. \tag{18.9}$$

The energy of the particle corresponding to the values given by Eq. (18.9) is

$$E = \frac{p^2}{2m} = \frac{n^2\pi^2\hbar^2}{2ma^2}, \tag{18.10}$$

or if $E_1 = \hbar^2\pi^2/2ma^2$ is the energy for $n = 1$, then

$$E = E_1, 4E_1, 9E_1, \ldots n^2E_1.$$

We conclude then that the particle cannot have any arbitrary energy, but only those values given by Eq. (18.10) and shown in Fig. 18-11; that is, the energy of the particle is quantized.

It is interesting to note that the minimum energy of a particle in a potential box is $E_1 = \hbar^2\pi^2/2ma^2$ and not zero as one would suspect. This minimum energy is related to the uncertainty principle in the following way. The uncertainty in the position of the particle is obviously $\Delta x \sim a$. The particle is moving back and forth with a momentum

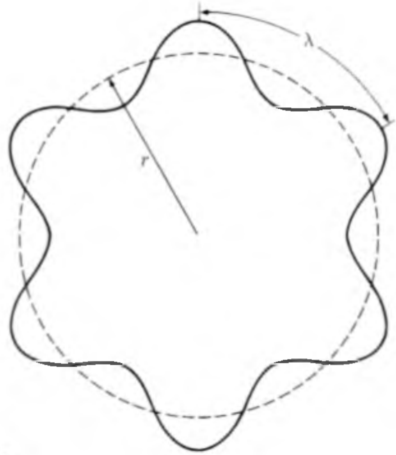


Fig. 18-12. Standing waves on a circle.

p; the uncertainty in the momentum is then $\Delta p \sim 2p$. The uncertainty principle requires that $\Delta x \Delta p \sim h$. Therefore

$$a(2p) \sim h \qquad \text{or} \qquad p \sim \frac{\pi \hbar}{a},$$

giving $E = p^2/2m \sim E_1$. The existence of a *zero-point energy*, as E_1 is sometimes called, is typical of all problems in which a particle is confined to move in a limited region.

As a second example, consider an electron in a hydrogenlike atom. Suppose that the electron describes a circular orbit as shown in Fig. 18-12. The momentum p of the electron is constant for a circular orbit. In order that the orbit correspond to a stationary state, it seems logical that it must be able to sustain standing waves of wavelength $\lambda = h/p$. We can see from Fig. 18-12 that the length of the orbit required must be equal to an integral multiple of λ; that is, $2\pi r = n\lambda = nh/p$, or

$$rp = \frac{nh}{2\pi}. \tag{18.11}$$

Noting that rp is the angular momentum of the electron, we see that the stationary states are those for which the angular momentum is an integral multiple of $\hbar = h/2\pi$. Since $p = mv$, we may also write Eq. (18.11) as

$$L = mvr = n\hbar, \tag{18.12}$$

which expresses the quantization of angular momentum. Equation (18.12) was used in Section 7.4 for obtaining the energy of the stationary states of hydrogen. However as we pointed out when discussing Heisenberg's uncertainty principle, it is impossible to define clearly the orbit of an electron in a hydrogen atom. Instead we talk of the region where it is more likely that the electron will be found. Therefore Eq. (18.12) cannot be rigorously valid. Instead, a more detailed analysis shows that the allowed values of the orbital angular momentum are given by

$$L = \sqrt{l(l+1)}\,\hbar \tag{18.13}$$

where $l = 0, 1, 2, \ldots$ This relation was already introduced in Section 7.5.

18.9 Wave Function and Probability Density

We have argued that we cannot talk about the trajectory of an atomic particle in the sense of classical mechanics. We cannot, for example, ask whether or not the electrons move in elliptical orbits around the nucleus in an atom. This question would be meaningless even if the forces acting on the particles produced such classical orbits. If we cannot talk about the trajectory of an electron or of any other atomic particle, how may we describe its motion?

The information to answer this question is provided by the matter field. To obtain such information, we are guided by our knowledge of waves. We recall that in standing waves the amplitude of the wave is fixed at each point of space. At points at which the amplitude is larger, the wave motion is more intense.

A similar situation occurs in the case of atomic particles. Consider, for example, an electron in an atom. The electron never moves too far away from the nucleus: the electron is essentially confined to a small region of space with dimensions of the order of 10^{-9} m. Thus the associated matter field of the electron may be expressed in terms of standing waves localized in this region with the amplitude varying from point

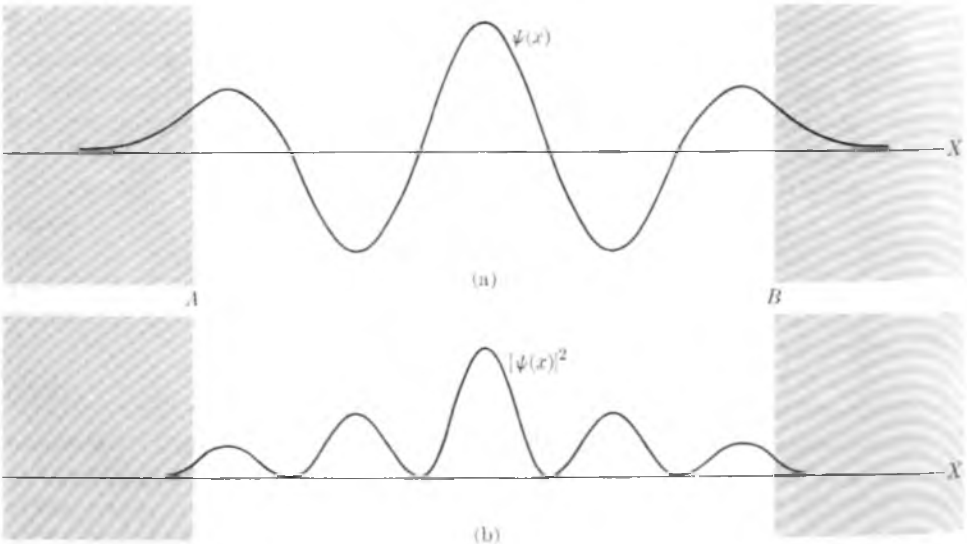


Fig. 18-13. (a) Wave function of a particle moving between A and B. (b) Probability distribution corresponding to the wave function shown in (a).

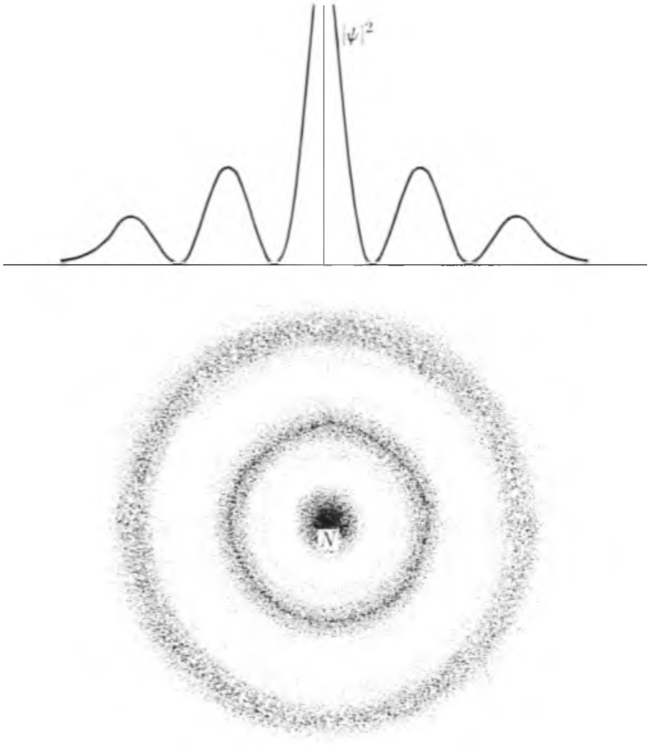


Fig. 18-14. Probability distribution for an electron in an atom.

to point within the region and being practically zero outside this region. Let us designate the amplitude of the matter field by $\psi(x)$. For historical reasons, this amplitude $\psi(x)$ is currently called the *wave function* although the name is misleading. Perhaps it would be better just to call it the *matter-field amplitude.*

We know that the intensity of a wave motion is proportional to the square of the amplitude. Therefore the *intensity of the matter field is given by* $|\psi(x)|^2$. The wave function $\psi(x)$ is sometimes expressed by a complex function; that is, a function containing $i=\sqrt{-1}$. The complex conjugate of a complex function is obtained by replacing each i by $-i$. The complex conjugate of a function ψ is designated by ψ^*. Then $|\psi(x)|^2 \equiv \psi^*(x)\psi(x)$. For a real function $\psi=\psi^*$.

Next consider what physical meaning may be ascribed to the intensity of the matter field. Since the matter field describes the motion of a particle, we may say that *the regions of space in which the particle is more likely to be found are those in which* $|\psi(x)|^2$ *is large.*

For example, the wave function $\psi(x)$ for a particle confined mainly to the region between A and B is shown in Fig. 18-13. Note that $\psi(x)$ decreases very rapidly outside the region AB, while the wave function is oscillating within such a region. The intensity of the matter-field, given by $|\psi(x)|^2$, is indicated in Fig. 18-13b.

To be more quantitative, we say that

> *the probability of finding the particle described by the wave function $\psi(x)$ in the interval dx around the point x is $|\psi(x)|^2\,dx$.*

In other words, the probability per unit length of finding the particle at x is

$$P(x)=|\psi(x)|^2. \tag{18.14}$$

We are assuming for simplicity that the motion is in one direction only. In the general case of motion in space, the wave function (or matter-field amplitude) depends on the three coordinates x, y, z [that is, $\psi(x, y, z)$]. Then $|\psi(x, y, z)|^2 dx\,dy\,dz$ is the probability of finding the particle in the volume $dx\,dy\,dz$ around the point having coordinates x, y, z; or

$$P=|\psi(x, y, z)|^2 \tag{18.15}$$

is the probability per unit volume, or the *probability density*, of finding the particle at x, y, z. For example, suppose that we compute ψ for an electron in an atom, and plot $|\psi|^2$ as in Fig. 18-14, in which N is the nucleus and the degree of darkness is proportional to the value of $|\psi|^2$. Thus the darker zones represent the regions in which the probability of finding the electron is greatest. This statement is the most we can say about the localization of the electron in the atom, and it is impossible to talk about the precise orbit of the electron.

18.10 The Schrödinger Equation

The next step in our investigation of the matter-field amplitude must be to find a rule by which the field amplitude or wave function ψ can be obtained for each dynamical problem. Surely the wave function $\psi(x)$ must depend on the dynamical state of the particle. This dynamical state is determined by the forces acting on the particle and by the particle's total energy. If the forces are conservative, the motion is determined by the potential energy $E_p(x)$ of the particle. Thus we may expect that the wave function $\psi(x)$ must depend in some way on the potential energy and therefore on the total energy,

$$E=\frac{p^2}{2m}+E_p(x), \tag{18.16}$$

of the particle. In fact, the rule for finding $\psi(x)$ is expressed in the form of a differential equation, called the *Schrödinger wave equation*, which was formulated in 1926 by Erwin Schrödinger. This equation (for one-dimensional problems) is

$$-\frac{\hbar^2}{2m}\frac{d^2\psi}{dx^2}+E_p(x)\psi=E\psi \tag{18.17}$$

where m is the mass of the particle. Schrödinger's equation is as fundamental to quantum mechanics as Newton's equation $\boldsymbol{F}=d\boldsymbol{p}/dt$ is to classical mechanics or Maxwell's equations are to electromagnetism. Clearly the solutions ψ of Eq. (18.17) depend on the form of the potential energy $E_p(x)$.

When we solve Schrödinger's equation, we obtain not only the wave function $\psi(x)$ but also the energy of the stationary states of the system.

Schrödinger's equation for the amplitude of the matter field can be easily understood by comparison with the wave equation for the amplitude of standing waves. (Recall Section 16.5.) In effect according to Eq. (16.29), the wave equation for waves propagating in one dimension is

$$\frac{d^2\psi}{dx^2}+k^2\psi=0$$

where $k=2\pi/\lambda$. Recalling that $p=h/\lambda$, we may write $k=2\pi p/h=p/\hbar$, which when substituted in the above equation reduces to

$$\frac{d^2\psi}{dx^2}+\frac{p^2}{\hbar^2}\psi=0. \tag{18.18}$$

For motion in a region in which the potential energy is E_p and the total energy is E, by rearranging Eq. (18.16) we have

$$p^2=2m[E-E_p(x)]$$

so that Eq. (18.18) may be written as

$$\frac{d^2\psi}{dx^2}+\frac{2m}{\hbar^2}[E-E_p(x)]\psi=0, \tag{18.19}$$

which is another way of writing Eq. (18.17).

We shall not make a detailed justification of Schrödinger's equation since it is beyond the scope of this book. However, we may state a practical rule to relate Eq. (18.16) to Eq. (18.17). If in Eq. (18.16),

$$E=\frac{p^2}{2m}+E_p(x),$$

we substitute the term $-i\hbar(d/dx)$ for p, we obtain

$$E=\frac{1}{2m}\left(-i\hbar\frac{d}{dx}\right)^2+E_p(x)=-\frac{\hbar^2}{2m}\frac{d^2}{dx^2}+E_p(x).$$

When we "operate" with this expression on $\psi(x)$, the result is Eq. (18.17).

If the motion is not restricted to one dimension and the particle may move in three-dimensional space, the Schrödinger equation becomes

$$\frac{\partial^2\psi}{\partial x^2}+\frac{\partial^2\psi}{\partial y^2}+\frac{\partial^2\psi}{\partial z^2}+\frac{2m}{\hbar^2}[E-E_p(x, y, z)]\psi=0. \tag{18.20}$$

18.11 The Wave Function of a Free Particle

In the case of a free particle the potential energy is zero [that is, $E_p(x)\equiv 0$], and Schrödinger's equation becomes

$$-\frac{\hbar^2}{2m}\frac{d^2\psi}{dx^2}=E\psi,$$

which may be written in the form

$$\frac{d^2\psi}{dx^2}+\frac{2mE}{\hbar^2}\psi=0. \tag{18.21}$$

However for a free particle. $E=p^2/2m$. Setting $p=\hbar k$, according to Eq. (18.2) where k is the wave number, we have

$$E=\frac{\hbar^2k^2}{2m}.$$

Then Eq. (18.21) becomes

$$\frac{d^2\psi}{dx^2}+k^2\psi=0, \tag{18.22}$$

which is identical to the equation for the amplitude of standing waves with a wavelength $\lambda=2\pi/k=h/p$ as required by Eq. (18.1). This equation is obeyed, for example, by the amplitude of standing waves on a string or in a gas column or electromagnetic waves trapped in a cavity.

Remembering that $i=\sqrt{-1}$ and $i^2=-1$, we see that by direct substitution the differential equation (18.22) admits as solutions the wave functions

$$\psi(x)=e^{ikx} \qquad \text{and} \qquad \psi(x)=e^{-ikx}. \tag{18.23}$$

The wave function $\psi=e^{ikx}$ represents a free particle of momentum $p=\hbar k$ and energy $E=p^2/2m=\hbar^2k^2/2m$ moving in the $+X$-direction, and the wave function $\psi=e^{-ikx}$ represents a free particle of the same momentum and energy but moving in the opposite or $-X$-direction.

Note that either solution in Eq. (18.23) yields

$$|\psi(x)|^2=\psi^*(x)\psi(x)=e^{-ikx}e^{ikx}=1.$$

The fact that $|\psi(x)|^2=1$, or a constant, means that the probability of finding the particle is the same at all points. In other words, $\psi=e^{\pm ikx}$ describes a situation in which we have complete uncertainty about position. This outcome is in agreement with the uncertainty principle because $\psi=e^{\pm ikx}$ describes a particle whose momentum, $p=\hbar k$, we know precisely: that is, $\Delta p=0$, which requires that $\Delta x\to\infty$.

18.12 The Wave Function of a Particle in a Potential Box

Since the particle is free in the region $0<x<a$ (Fig. 18-10), we have that Schrödinger's equation for a particle in a potential box is

$$\frac{d^2\psi}{dx^2}+k^2\psi=0, \qquad k^2=2mE/\hbar^2 \qquad (0<x<a). \tag{18.24}$$

Since the particle is moving back and forth between $x=0$ and $x=a$, the wave function is

$$\psi(x)=Ae^{ikx}-Be^{-ikx}, \tag{18.25}$$

which contains motion in both directions. The boundary conditions require that $\psi(x)=0$ at $x=0$ and $x=a$ since the particle cannot be outside the walls of the box. Then the first boundary condition at $x=0$ gives

$$\psi(x)_{x=0}=A+B=0 \qquad \text{or} \qquad B=-A.$$

Therefore we can rewrite Eq. (18.25) as

$$\psi(x)=A(e^{ikx}-e^{-ikx})=2iA\sin kx=C\sin kx \tag{18.26}$$

where $C=2iA$. The boundary condition at $x=a$ gives

$$\psi(x)_{x=0}=C\sin ka=0. \tag{18.27}$$

Since C cannot be zero (because then there would be no wave function), we conclude that

$$\sin ka=0 \qquad \text{or} \qquad ka=n\pi$$

where n is an integer. Solving for k, we have

$$k=\frac{n\pi}{a} \qquad \text{or} \qquad p=\hbar k=\frac{n\pi\hbar}{a}, \tag{18.28}$$

which gives the possible values of the momentum $p=\hbar k$ of the particle. This result is identical to Eq. (18.8) obtained using a more intuitive method. Substituting the quantized value of k found in Eq. (18.28) into the wave function of Eq. (18.26) yields the general wave function

$$\psi(x)=C\sin\left(\frac{n\pi}{a}x\right); \tag{18.29}$$

and from Eq. (18.24) we may write the quantized energy levels as

$$E_n=\frac{k^2\hbar^2}{2m}=\frac{\hbar^2\pi^2}{2ma^2}n^2. \tag{18.30}$$

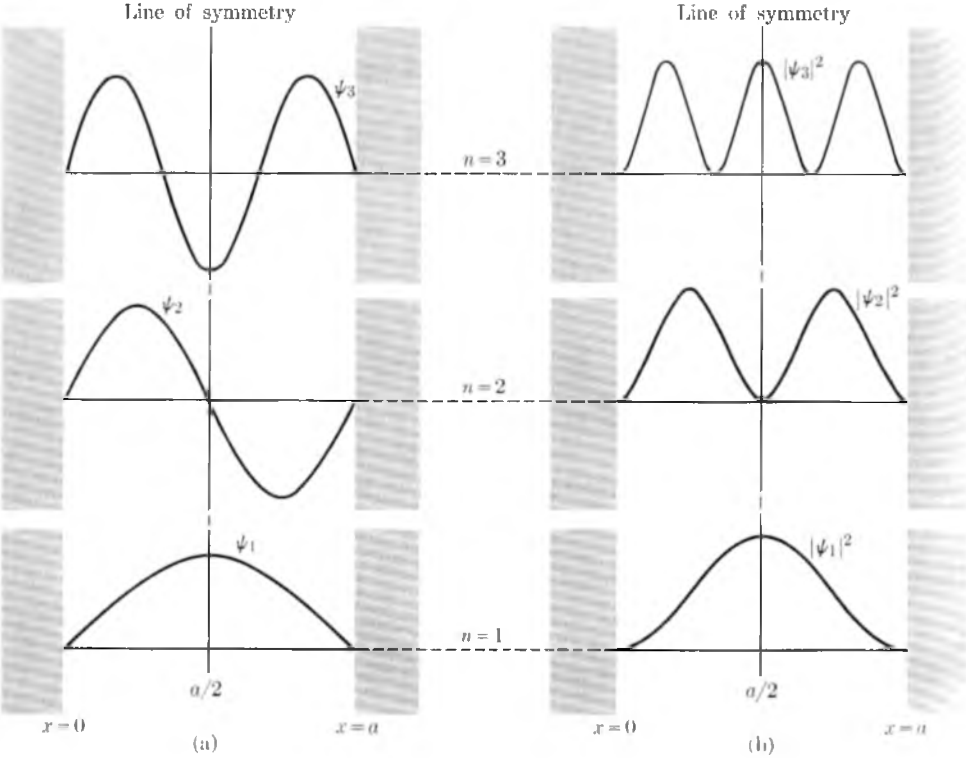


Fig. 18-15. (a) First three wave functions for a particle in a potential box. (b) Corresponding probability densities.

In Fig. 18-15a the wave functions corresponding to $n=1$, 2, and 3 are illustrated; and in Fig. 18-15b the corresponding probability densities for a particle in a potential box are given.

18.13 The Wave Function of the Simple Harmonic Oscillator

An interesting and important physical system is the simple harmonic oscillator, corresponding to the potential energy

$$E_p(x)=\tfrac{1}{2}kx^2.$$

A potential-energy diagram for a simple harmonic oscillator is shown in Fig. 18-16. The Schrödinger equation is

$$-\frac{\hbar^2}{2m}\frac{d^2\psi(x)}{dx^2}+\tfrac{1}{2}kx^2\psi(x)=E\psi(x). \qquad (18.31)$$

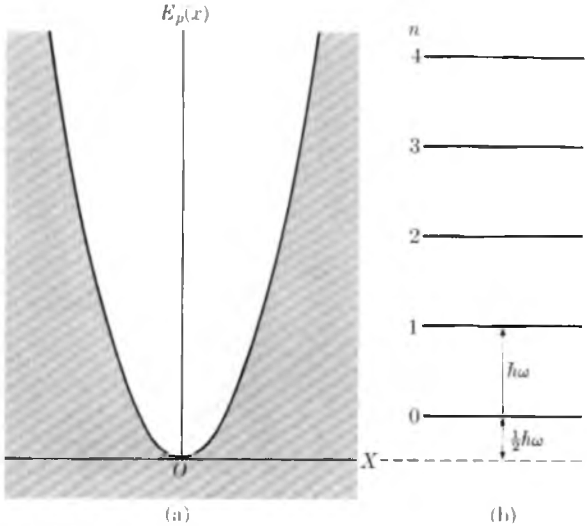


Fig. 18-16. (a) Potential energy of a simple harmonic oscillator. (b) Energy levels.

The solution of this differential equation is too elaborate a mathematical problem to be discussed here. Nevertheless, it may be shown that the possible values of the energy are

$$E_n = (n + \tfrac{1}{2})\hbar\omega \tag{18.32}$$

where $n = 0, 1, 2, 3, \ldots$, integer, and

$$\omega = \sqrt{k/m} \tag{18.33}$$

is the angular frequency of the oscillator.

The zero-point energy of the oscillator is $\frac{1}{2}\hbar\omega$, again a consequence of the uncertainty principle. The energy levels, represented in Fig. 18-16, are equally spaced by an amount $\hbar w$.

Table 18-1. Wave Functions for the Simple Harmonic Oscillator ($a = m\omega/\hbar$)

n	E_n	$\psi_n(x)$
0	$(1/2)\hbar\omega$	$(a/\sqrt{\pi})^{1/2}e^{-a^2x^2/2}$
1	$(3/2)\hbar\omega$	$(a/2\sqrt{\pi})^{1/2}2axe^{-a^2x^2/2}$
2	$(5/2)\hbar\omega$	$(a/8\sqrt{\pi})^{1/2}(4a^2x^2-2)e^{-a^2x^2/2}$
3	$(7/2)\hbar\omega$	$(a/48\sqrt{\pi})^{1/2}(8a^3x^3-12ax)e^{-a^2x^2/2}$

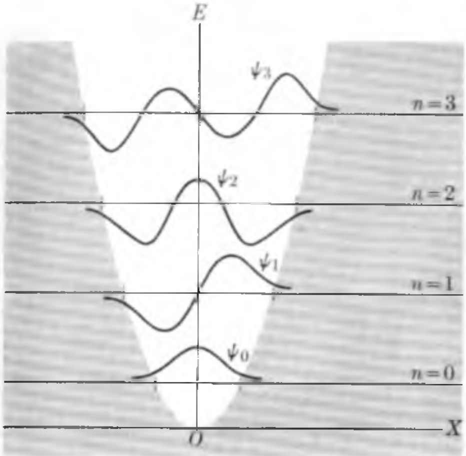


Fig. 18-17. Wave functions corresponding to the first four energy levels of a harmonic oscillator.

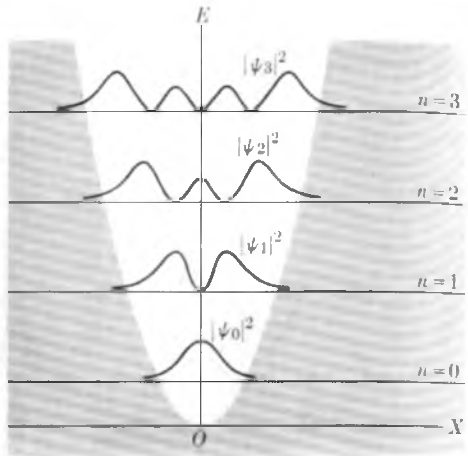


Fig. 18-18. Probability densities corresponding to the first four energy levels of a harmonic oscillator.

We shall not derive the expression for the wave functions. The wave functions corresponding to $n=0$, 1, 2, and 3 have been given in Table 18-1 and represented in Fig. 18-17. The corresponding probability densities are shown in Fig. 18-18.

18.14 The Hydrogen Atom

Although the simplest of all atoms are the one-electron atoms, such as hydrogen (H), single ionized helium (He^+), double ionized lithium (Li^{++}), and so on, their mathematical treatment is too complicated to be given here in full detail. However, there are several features that can be analyzed without great difficulty.

For atoms with one electron the potential energy of the electron relative to the nucleus is $E_p = -Ze^2/(4\pi\epsilon_0 r)$ where Z is the atomic number of the nucleus, and Schrödinger's equation becomes

$$\frac{\partial^2\psi}{\partial x^2}+\frac{\partial^2\psi}{\partial y^2}+\frac{\partial^2\psi}{\partial z^2}+\frac{2m}{\hbar^2}\left(E+\frac{Ze^2}{4\pi\epsilon_0 r}\right)\psi=0. \tag{18.34}$$

The fact that the potential energy depends only on the distance r from the electron to the nucleus suggests that some solutions of Eq. (18.34) are spherically symmetric; that is, ψ depends only on r, or $\psi(r)$. To test this idea, we must first find the form adopted by Eq. (18.34) for this case. First note that since $r^2 = x^2+y^2+z^2$, then $2r(\partial r/\partial x)=2x$ and thus

$$\frac{\partial r}{\partial x}=\frac{x}{r} \tag{18.35}$$

with similar results for $\partial r/\partial y$ and $\partial r/\partial z$. Then

$$\frac{\partial\psi(r)}{\partial x}=\frac{d\psi}{dr}\frac{\partial r}{\partial x}=\frac{x}{r}\frac{d\psi}{dr} \tag{18.36}$$

and

$$\frac{\partial^2\psi(r)}{\partial x^2}=\frac{\partial}{\partial x}\left[\frac{x}{r}\frac{d\psi}{dr}\right]=\frac{1}{r}\frac{d\psi}{dr}-\frac{x}{r^2}\frac{\partial r}{\partial x}\frac{d\psi}{dr}+\frac{x}{r}\frac{d^2\psi}{dr^2}\frac{\partial r}{\partial x}.$$

Substituting Eq. (18.35) into this equation gives

$$\frac{\partial^2\psi(r)}{\partial x^2}=\frac{1}{r}\frac{d\psi}{dr}-\frac{x^2}{r^3}\frac{d\psi}{dr}+\frac{x^2}{r^2}\frac{d^2\psi}{dr^2}. \tag{18.37}$$

Writing the equivalent expressions for $\partial^2\psi/\partial y^2$ and $\partial^2\psi/\partial z^2$ and adding the three together, we finally obtain Schrödinger's equation in the form

$$\frac{d^2\psi}{dr^2}+\frac{2}{r}\frac{d\psi}{dr}+\frac{2m}{\hbar^2}\left(E-\frac{Ze^2}{4\pi\epsilon_0 r}\right)\psi=0. \tag{18.38}$$

Since the probability of finding the electron at a certain distance from the nucleus must decrease as the distance increases, we may assume that the wave function is an expression of the form $f(r)e^{-ar}$ where a is a constant to be determined. Assume for simplicity that $\psi(r)=e^{-ar}$. Substituting this function and its derivatives into Eq. (18.38) will give

$$a^2e^{-ar}-\frac{2a}{r}e^{-ar}+\frac{2m}{\hbar^2}\left(E+\frac{Ze^2}{4\pi\epsilon_0 r}\right)e^{-ar}=0.$$

This equation may be rewritten as

$$\left[a^2+\left(\frac{2m}{k^2}\right)E\right]+\left[-2a+\left(\frac{2mZe^2}{4\pi\epsilon_0\hbar^2}\right)\right]\frac{1}{r}=0.$$

The two terms within the square brackets must independently be zero; therefore we have

$$a^2+\left(\frac{2m}{k^2}\right)E=0 \qquad \text{and} \qquad -2a+\frac{2mZe^2}{4\pi\epsilon_0\hbar^2}=0. \tag{18.39}$$

From the second relation in Eq. (18.39) we find

$$a=\frac{mZe^2}{4\pi\epsilon_0\hbar^2}. \tag{18.40}$$

which when placed in the first relation of Eq. (18.39) gives the energy as

$$E=-\frac{a^2\hbar^2}{2m}=-\frac{mZ^2e^4}{(4\pi\epsilon_0)^2 2\hbar^2}. \tag{18.41}$$

Comparing this value of the energy with Eq. (7.26), we see that the result above corresponds to $n=1$ in Bohr's energy formula. Thus the wave function $\psi=e^{-ar}$ corresponds to the ground state of a hydrogenlike atom.

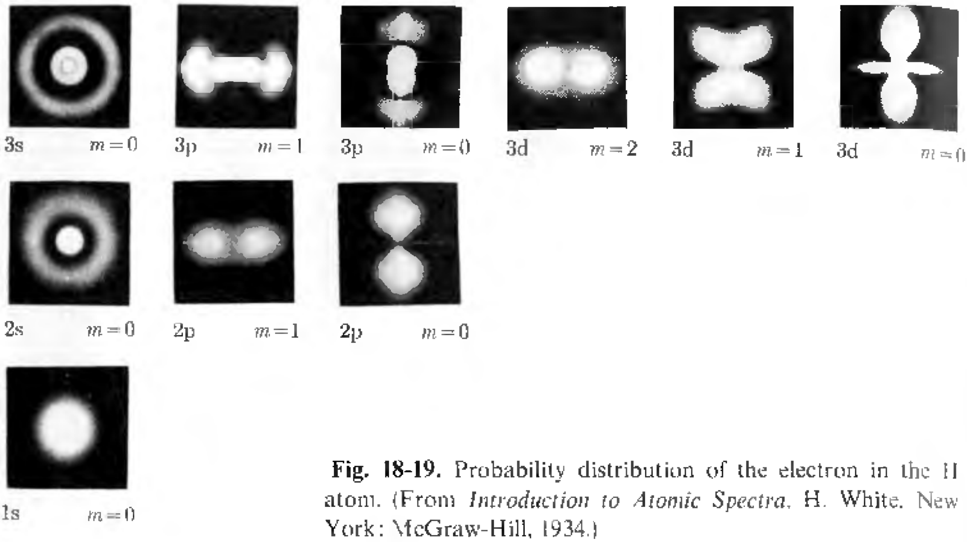


Fig. 18-19. Probability distribution of the electron in the H atom. (From *Introduction to Atomic Spectra*, H. White. New York: McGraw-Hill, 1934.)

By a similar analysis we can easily verify that

$$\psi = e^{-ar/2}(2 - ar) \tag{18.42}$$

is also a solution of Eq. (18.38) corresponding to an energy

$$E = -\frac{mZ^2e^4}{2\hbar^2(4\pi\epsilon_0)^2(2^2)}. \tag{18.43}$$

Again comparing with Eq. (7.26), we see that this energy corresponds to $n=2$ in Bohr's formula. In this way we recognize that the wave function given in Eq. (18.42) corresponds to the first excited state of the atom. Similarly we may find the solutions of Eq. (18.38) for $n=3$, 4, etc.

The question now arises: Are the only solutions of Schrödinger's equation those that are spherically symmetric? The answer is that the spherically symmetric solutions correspond to zero angular momentum ($l=0$) or s-states, but that states with nonzero values of angular momentum are not spherically symmetric. The wave functions for such states depend on the angles that r makes with the coordinate axes. For example, it can be shown that there are solutions of the form $xf(r)$, $yf(r)$, and $zf(r)$ where the probability of finding the electron is concentrated along the X, Y, and Z axes; these states correspond to an angular momentum $l=1$ or p-states. The lowest p-state corresponds to $n=2$ in Bohr's energy formula.

Thus we see that Schrödinger's equation not only accounts for the correct energy values of the electron motion but also provides relevant information about the angular momentum of the electron. In Fig. 18-19 the probability distribution $|\psi|^2$ for different energy levels and different values of the angular momentum is shown.

In concluding this brief introduction to quantum mechanics we may state that the theory of the matter field based on Schrödinger's equation provides an adequate, but not complete algorithm for a discussion of the structure of matter.

Problems

18.1 Calculate the de Broglie wavelength of an electron when its energy is (a) 1 eV, (b) 100 eV, (c) 1 keV, and (d) 1 MeV.

18.2 Calculate which electrons given in Problem 18.1 would be significantly diffracted in a nickel crystal, in which the atomic separation is about 2.15 Å.

18.3 Calculate the energy of electrons that are Bragg-diffracted at an angle of 30° by the crystal in Problem 18.2.

18.4 Monochromatic X-rays ($\lambda = 0.5$ Å) are incident on a sample of KCl powder. A flat photographic plate is placed perpendicular to the incident beam, at a distance of 1.0 m from the powder. Determine the first- and second-order Bragg radii, given that the Bragg plane separation is 3.14 Å.

18.5 A narrow beam of thermal neutrons produced by a nuclear reactor falls on a crystal with lattice spacing of 1.60 Å. Determine the Bragg angle if 2-eV neutrons are strongly diffracted.

18.6 Suppose that a beam of electrons with a de Broglie wavelength of 10^{-5} m passes through a slit 10^{-4} m wide. What angular spread is introduced because of diffraction by the slit?

18.7 A probe must always be smaller (at least by a factor of 10) than the object being studied; otherwise there will be significant perturbation of the position and velocity of the object. Calculate the minimum particle energy if (a) photons, (b) electrons, and (c) neutrons are used to probe a nucleus whose diameter is 10^{-14} m.

18.8 The velocity of a proton in the X-direction is measured to an accuracy of 10^{-7} m s^{-1}. Determine the limit of accuracy with which the proton can be located simultaneously (a) along the X-axis, (b) along the Y-axis.

18.9 Repeat Problem 18.8 for the case in which the particle is an electron.

18.10 The position of an electron is determined with an uncertainty of 0.1 Å. (a) Find the uncertainty in its momentum. (b) If the electron's energy is of the order of 1 keV, estimate the uncertainty in its energy.

18.11 Repeat Problem 18.10 for a proton confined to a nuclear diameter ($\approx 10^{-14}$ m) and having energy on the order of 2 MeV.

18.12 A particle moves rectilinearly under the action of a uniform electric field $\mathscr{E}$ so that its potential energy is $E_p = -\mathscr{E}x$. (a) Write the Schrödinger equation for this motion. (b) Make a sketch of the wave functions for an energy E larger and smaller than zero. (c) Is the energy quantized?

18.13 Find the frequency spread for a 1-nanosecond (10^{-9} s) pulse from a ruby laser ($\lambda = 6.3 \times 10^{-7}$ m).

18.14 What is the effect on the energy levels of a one-dimensional potential box as the size of the box (a) decreases? (b) increases?

18.15 Consider an electron in a one-dimensional potential box of width 2.0 Å. (a) Calculate the zero-point energy. (b) Using the uncertainty principle, discuss the effect of incident radiation designed to locate the electron with a 1% accuracy (that is, $\Delta x = 0.02$ Å).

18.16 (a) Estimate the zero-point energy of an electron confined inside a region of size 10^{-14} m, the order of magnitude of nuclear dimensions. (b) Compare this energy with both the gravitational potential energy and the coulomb potential energy of an electron and a proton separated the same distance. (c) On the basis of this comparison, discuss the possibility that an electron can exist within a nucleus.

18.17 Calculate the zero-point energy of a neutron confined within a nucleus 10^{-15} m in size.

18.18 A particle is represented by the wave function

$$\psi(x)=e^{-(x-x_0)^2/2\alpha}\sin kx,$$

where α is a constant $\gg \lambda^2$. (a) Plot the wave function $\psi(x)$ and the probability distribution $|\psi(x)|^2$ from $x=x-3x_0$ to $x=x+3x_0$. (b) Estimate the uncertainty in the position and in the momentum of the particle.

18.19 Show that the energy levels and wave functions of a particle moving in the XY-plane within a two-dimensional potential box of sides a and b are

$$E=\left(\frac{\hbar^2\pi^2}{2m}\right)\left(\frac{n_1^2}{a^2}+\frac{n_2^2}{b^2}\right),$$

$$\psi=C\sin\left(\frac{n_1\pi x}{a}\right)\sin\left(\frac{n_2\pi y}{b}\right).$$

Discuss the degeneracy of energy levels when $a=b$.

18.20 Find the energy levels and wave functions for a particle moving within a three-dimensional box of sides a, b, and c. Note that this is an extension of the previous problem.

18.21 Calculate the zero-point energy and the spacing for the energy levels (a) in a 1-dimensional harmonic oscillator with an oscillatory frequency of 400 Hz, (b) in a three-dimensional harmonic oscillator with an oscillatory frequency of 400 Hz, (c) in the CO molecule, if the two atoms oscillate with a frequency of 6.43×10^{11} Hz.

18.22 A particle moves due to a potential energy

$$E_p(x)=-E_0e^{-\alpha x^2}.$$

(a) Plot $E_p(x)$. (b) Make a sketch of the wave functions when the total energy is (i) negative and (ii) positive. (c) Do you expect to have quantized energy levels?

CHALLENGING PROBLEMS

18.23 What is the momentum of an electron whose de Broglie wavelength is 10^{-10} m? [AP-B; 1969]

18.24 (a) Describe and interpret an experiment in which electromagnetic radiation exhibits particle-like behavior. (b) Describe and interpret an experiment in which electrons exhibit wavelike behavior. [AP-B; 1973]

18.25 A particle of mass m moves back and forth between two solid walls separated by a distance L. Theory predicts that the only allowed values of the particle's kinetic energy are given by the equation

$$E_n=\frac{n^2h^2}{8mL^2},$$

where h is Planck's constant and $n=1, 2, 3, \ldots$ (a) Determine the longest wavelength of light, $\lambda_{\max}$, that this system can absorb when it is in its ground state, $n=1$. (b) Show that the de Broglie wavelength of the particle for any value of n is given by

$$\lambda_n=\frac{2L}{n}.$$

(c) Describe one other physical system for which the same relation given in part (b) between a wavelength λ_n and a distance L holds [AP-B; 1978]

18.26 A helium atom of mass m moving with speed v zigzags back and forth between two parallel walls of length L separated by distance a as shown in Fig. 18-20. (a) In terms of a, v, and θ, calculate the time interval Δt between two successive collisions with the right-hand wall. (b) In terms of m, v, and θ, calculate the magnitude of the momentum Δp imparted to the right-hand wall each time the atom collides with it. (c) Calculate the average force that the atom exerts on the right-hand wall, and express

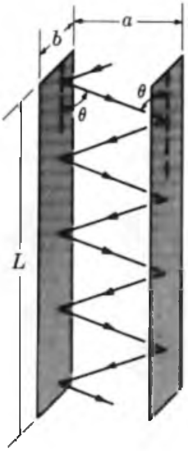


Figure 18-20

the resulting pressure P on the wall in terms of θ, the volume V of the region bounded by the walls, and the kinetic energy E of the atom. (d) Suppose, instead, that a photon of energy E is following the zigzag path. Calculate the magnitude of the momentum Δp it imparts to the right-hand wall in each collision, and express the resulting pressure P in terms of E, θ, and V [AP-B: 1980].

18.27 Show that the ratio of the de Broglie wavelength to the Compton wavelength for the same particle is equal to

$$\sqrt{(c/v)^2 - 1}.$$

18.28 Verify the fact that the group velocity of a wave packet is equal to the particle velocity, even under relativistic conditions. Also show that the phase velocity of the matter field at relativistic speeds is equal to c^2/v.

18.29 If a source moves with a velocity u relative to an observer, the frequency of the radiation measured by the observer suffers a shift $\Delta\nu = \nu u/c$, where u is positive (negative) when the motion is toward (away from) the observer, and where ν would be the frequency if the source were stationary. This is called the *electromagnetic Doppler shift*. Since the molecules in a gas are in random motion, the Doppler shift is different for each molecule. This introduces a line broadening, given by $\delta = 2(\nu/c)\sqrt{2kT \ln 2/m}$, where m is the mass of the molecule and T is the absolute temperature of the gas. Compute the *Doppler broadening* at room temperature (300 K) for (a) the 4.86-eV atomic transition in mercury and for (b) the 1.33-MeV nuclear transition in ^{60}Ni. Discuss in each case the effect on resonance absorption.

18.30 The gamma-ray line emitted by ^{191}Ir has a mean energy of 129 keV, and the measured width of the line at half-maximum intensity is 4.6×10^{-6} eV. Estimate (a) the mean lifetime of the excited state emitting this line, (b) the relative velocity of source and observer that is required to give a first-order Doppler shift equal to the measured line width.

18.31 (a) From the definition of probability density as $P = |\psi|^2$, show that

$$\int_{-\infty}^{\infty} P\,dV = 1$$

where $dV = dx\,dy\,dz$. (*Hint:* Note that the argument here is logical, not mathematical.) (b) From the result of (a), calculate the value of the constant in the simple harmonic oscillator wave function.

18.32 The general expression of the solutions of Schrödinger's equation for the harmonic oscillator is

$$\psi_n = N_n H_n(ax) e^{-a^2x^2/2},$$

where $N_n = \sqrt{a/\pi^{1/2} 2^n n!}$ is the normalization constant, $a = \sqrt{m\omega/\hbar}$, and the functions $H_n(ax)$ are called *Hermite polynomials*, defined by

$$H_n(\xi) = (-1)^n e^{\xi^2} \frac{d^n}{d\xi^n}(e^{-\xi^2}).$$

Write the first four wave functions ($n = 0, 1, 2, 3$) and compare with the expressions given in Table 18-1.

APPENDIX

MATHEMATICAL RELATIONS AND TABLES

This appendix, in which we present certain mathematical formulas that are frequently used in the text, is intended as a quickly available reference for the student. In a few cases we have inserted some mathematical notes in the text proper. Proofs and a discussion of most of the formulas may be found in any standard calculus text; e.g., *Calculus and Analytic Geometry*, fifth edition, by G. B. Thomas (Addison-Wesley, 1979). A short introduction to the basic concepts of the calculus, in a programmed format, may be found in *Quick Calculus: A Short Manual of Self Instruction*, by D. Kelpner and N. Ramsey (John Wiley & Sons, New York, 1963). The student will also have to refer to a number of tables which are in book form. Among these are the *C.R.C. Standard Mathematical Tables* (Chemical Rubber Company, Cleveland, Ohio, 1963), and *Tables of Integrals and Other Mathematical Data*, fourth edition, by H. B. Dwight (Macmillan Company, New York, 1961). We recommend that the student have at his disposal the *Handbook of Chemistry and Physics*, yearly editions of which are issued by the Chemical Rubber Company, Cleveland, Ohio. This handbook also contains a wealth of mathematical, chemical, and physical data.

1. Trigonometric Relations

Referring to Fig. M-1, we can define the following relations:

$$\sin\alpha = y/r, \qquad \cos\alpha = x/r, \qquad \tan\alpha = y/x; \tag{M.1}$$

$$\csc\alpha = r/y, \qquad \sec\alpha = r/x, \qquad \cot\alpha = x/y; \tag{M.2}$$

$$\tan\alpha = \sin\alpha/\cos\alpha; \tag{M.3}$$

$$\sin^2\alpha + \cos^2\alpha = 1, \qquad \sec^2\alpha - 1 = \tan^2\alpha; \tag{M.4}$$

$$\sin(\alpha \pm \beta) = \sin\alpha\cos\beta \pm \cos\alpha\sin\beta; \tag{M.5}$$

$$\cos(\alpha \pm \beta) = \cos\alpha\cos\beta \mp \sin\alpha\sin\beta; \tag{M.6}$$

$$\sin\alpha \pm \sin\beta = 2\sin\tfrac{1}{2}(\alpha \pm \beta)\cos\tfrac{1}{2}(\alpha \mp \beta); \tag{M.7}$$

$$\cos\alpha + \cos\beta = 2\cos\tfrac{1}{2}(\alpha + \beta)\cos\tfrac{1}{2}(\alpha - \beta); \tag{M.8}$$

$$\cos\alpha - \cos\beta = -2\sin\tfrac{1}{2}(\alpha + \beta)\sin\tfrac{1}{2}(\alpha - \beta); \tag{M.9}$$

$$\sin\alpha\sin\beta = \tfrac{1}{2}[\cos(\alpha - \beta) - \cos(\alpha + \beta)]; \tag{M.10}$$

$$\cos\alpha\cos\beta = \tfrac{1}{2}[\cos(\alpha - \beta) + \cos(\alpha + \beta)]; \tag{M.11}$$

$$\sin\alpha\cos\beta = \tfrac{1}{2}[\sin(\alpha - \beta) + \sin(\alpha + \beta)]; \tag{M.12}$$

$$\sin 2\alpha = 2\sin\alpha\cos\alpha, \qquad \cos 2\alpha = \cos^2\alpha - \sin^2\alpha; \tag{M.13}$$

$$\sin^2\tfrac{1}{2}\alpha = \tfrac{1}{2}(1 - \cos\alpha), \qquad \cos^2\tfrac{1}{2}\alpha = \tfrac{1}{2}(1 + \cos\alpha). \tag{M.14}$$

Referring to Fig. M-2, we can formulate, for any arbitrary triangle:

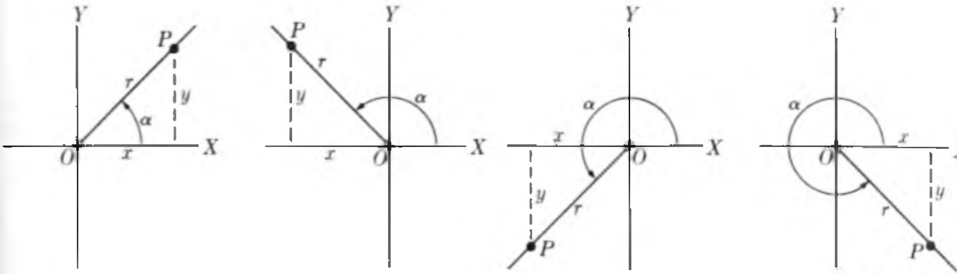


Figure M-1

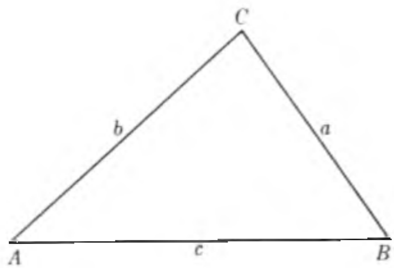


Figure M-2

$$\text{Law of sines: } \frac{a}{\sin A} = \frac{b}{\sin B} = \frac{c}{\sin C}, \tag{M.15}$$

$$\text{Law of cosines: } \quad a^2 = b^2 + c^2 - 2bc \cos A. \tag{M.16}$$

2. Logarithms

(i) Definition of e:

$$e = \lim_{n \to \infty} \left(1 + \frac{1}{n}\right)^n = 2.7182818\ldots \tag{M.17}$$

The exponential functions $y = e^x$ and $y = e^{-x}$ are plotted in Fig. M-3.
(ii) Natural logarithm, base e (see Fig. M-4):

$$y = \ln x \qquad \text{if } x = e^y. \tag{M.18}$$

Common logarithm, base 10:

$$y = \log x \qquad \text{if } x = 10^y. \tag{M.19}$$

The natural and common logarithms are related by

$$\ln x = 2.303 \log x, \qquad \log x = 0.434 \ln x. \tag{M.20}$$

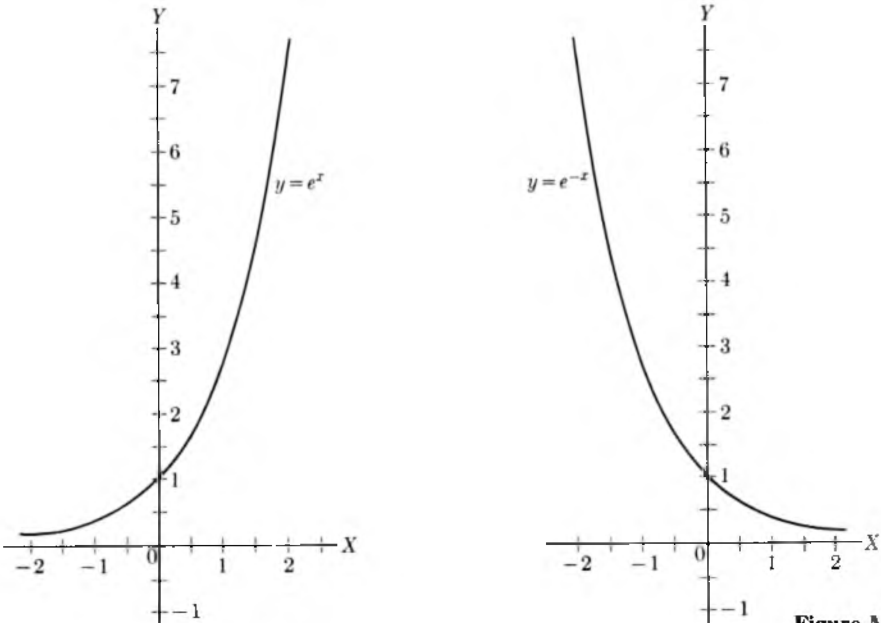


Figure M-3

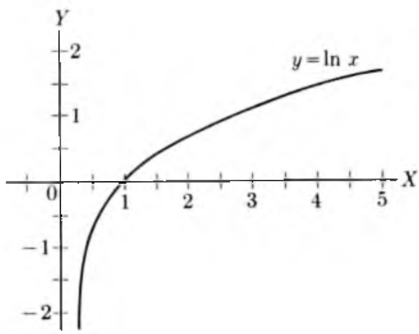


Figure M-4

3. Power Expansions

(i) The binomial expansion:

$$(a+b)^n = a^n + na^{n-1}b + \frac{n(n-1)}{2!}a^{n-2}b^2 + \frac{n(n-1)(n-2)}{3!}a^{n-3}b^3 + \cdots + \frac{n(n-1)(n-2)\cdots(n-p+1)}{p!}a^{n-p}b^p + \cdots \tag{M.21}$$

When n is a positive integer, the expansion has $n+1$ terms. In all other cases, the expansion has an infinite number of terms. The case for which a is 1 and b is a quantity x is used numerous times in the text. Therefore the binomial expansion of $(1+x)^n$ is written

$$(1+x)^n = 1 + nx + \frac{n(n-1)}{2!}x^2 + \frac{n(n-1)(n-2)}{3!}x^3 + \cdots \tag{M.22}$$

(ii) Other useful series expansions:

$$e^x = 1 + x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 + \cdots \tag{M.23}$$

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \cdots \tag{M.24}$$

$$\sin x = x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5 - \cdots \tag{M.25}$$

$$\cos x = 1 - \frac{1}{2!}x^2 + \frac{1}{4!}x^4 - \cdots \tag{M.26}$$

$$\tan x = x + \frac{1}{3}x^3 + \frac{2}{15}x^5 + \cdots \tag{M.27}$$

For $x \ll 1$, the following approximations are satisfactory:

$$(1+x)^n \approx 1 + nx, \tag{M.28}$$

$$e^x \approx 1 + x, \qquad \ln(1+x) \approx x, \tag{M.29}$$

$$\sin x \approx x, \qquad \cos x \approx 1, \qquad \tan x \approx x. \tag{M.30}$$

Note that in Eqs. (M.25), (M.26), (M.27), and (M.30), x must be expressed in radians.

(iii) Taylor series expansion:

$$f(x) = f(x_0) + (x - x_0)\left(\frac{df}{dx}\right)_0 + \frac{1}{2!}(x - x_0)^2\left(\frac{d^2f}{dx^2}\right)_0 + \cdots + \frac{1}{n!}(x - x_0)^n\left(\frac{d^nf}{dx^n}\right)_0 + \cdots \tag{M.31}$$

If $x - x_0 \ll 1$, a useful approximation is

$$f(x) \approx f(x_0) + (x - x_0)\left(\frac{df}{dx}\right)_0. \tag{M.32}$$

4. Complex Numbers

With the definition $i^2 = -1$ or $i = \sqrt{-1}$,

$$e^{i\theta} = \cos\theta + i\sin\theta, \tag{M.33}$$

$$\cos\theta = \tfrac{1}{2}(e^{i\theta} + e^{-i\theta}), \tag{M.34}$$

$$\sin\theta = \frac{1}{2i}(e^{i\theta} - e^{-i\theta}). \tag{M.35}$$

5. Hyperbolic Functions

In order to visualize the following relations, see Fig. M-5.

$$\cosh\theta = \tfrac{1}{2}(e^{\theta} + e^{-\theta}), \tag{M.36}$$

$$\sinh\theta = \tfrac{1}{2}(e^{\theta} - e^{-\theta}), \tag{M.37}$$

$$\cosh^2\theta - \sinh^2\theta = 1, \tag{M.38}$$

$$\sinh\theta = -i\sin(i\theta), \qquad \cosh\theta = \cos(i\theta), \tag{M.39}$$

$$\sin\theta = -i\sinh(i\theta), \qquad \cos\theta = \cosh(i\theta). \tag{M.40}$$

6. Basic Derivatives and Integrals

$f(u)$	df/dx	$\int f(u)\,du$
u^n	$nu^{n-1}\,du/dx$	$u^{n+1}/(n+1) + C\ (n \neq -1)$
u^{-1}	$-(1/u^2)\,du/dx$	$\ln u + C$
$\ln u$	$(1/u)\,du/dx$	$u\ln u - u + C$
e^u	$e^u\,du/dx$	$e^u + C$
$\sin u$	$\cos u\,du/dx$	$-\cos u + C$
$\cos u$	$-\sin u\,du/dx$	$\sin u + C$
$\tan u$	$\sec^2 u\,du/dx$	$-\ln\cos u + C$
$\cot u$	$-\csc^2 u\,du/dx$	$\ln\sin u + C$
$\arcsin u$	$(du/dx)/\sqrt{1-u^2}$	$u\sin^{-1}u + \sqrt{1-u^2} + C$
$\sinh u$	$\cosh u\,du/dx$	$\cosh u + C$
$\cosh u$	$\sinh u\,du/dx$	$\sinh u + C$

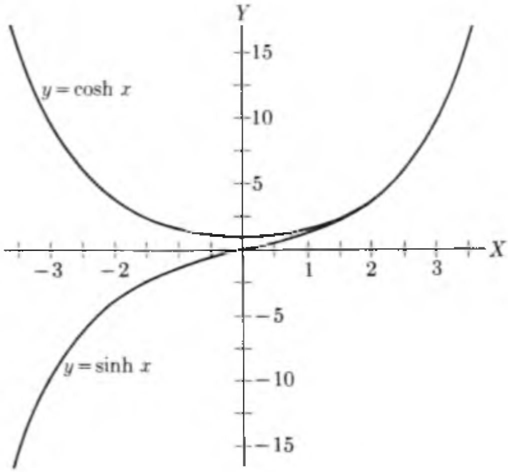


Figure M-5

A useful rule for integration, called *integration by parts*, is

$$\int u\,dv = uv - \int v\,du. \tag{M.41}$$

This method is most frequently used to evaluate the integral on the left by using the integral on the right.

7. Average Value of a Function

The *mean* or *average value* of a function $y = f(x)$ in the interval (a, b) is defined by

$$y_{\text{ave}} = \frac{1}{b-a}\int_a^b y\,dx. \tag{M.42}$$

Similarly, the average value of y^2 is defined by

$$(y^2)_{\text{ave}} = \frac{1}{b-a}\int_a^b y^2\,dx. \tag{M.43}$$

The quantity $\sqrt{(y^2)_{\text{ave}}}$ is called the *root mean square* value of $y = f(x)$ in the interval (a, b), and in general is different from y_{ave}. It is designated y_{rms}.

8. Conic sections

An important family of plane curves are the *conic sections*. A conic section is defined as a curve generated by a point moving in such a way that the ratio of its distance to a point, called the *focus*, and to a line, called the *directrix*, is constant. There are three kinds of conic sections, called ellipse, parabola, and hyperbola, depending on whether this ratio (called the *eccentricity*) is smaller than, equal to, or larger than one. Designating the eccentricity by ϵ, the focus by F, and the directrix by HQD (Fig. M-6), we have

$$\epsilon = PF/PQ.$$

Now $PF = r$, and if we say that $FD = d$, then $PQ = FD - FB = d - r\cos\theta$. Therefore $\epsilon = r/(d - r\cos\theta)$. Or, solving for r,

$$\frac{\epsilon d}{r} = 1 + \epsilon\cos\theta. \qquad \text{(M.44)}$$

This is the form in which the equation of a conic section has been used in the text. The equation for the conic section may also be derived using the angle $\pi - \theta$, and thus the equation has the form

$$\frac{\epsilon d}{r} = 1 - \epsilon\cos\theta.$$

In the case of an ellipse, which is a closed curve, point A corresponds to $\theta = 0$ and point A' to $\theta = \pi$. Thus, according to the polar equation, we have

$$r_1 = \frac{\epsilon d}{1+\epsilon} \quad \text{and} \quad r_2 = \frac{\epsilon d}{1-\epsilon}.$$

Then, since $r_1 + r_2 = 2a$, the semimajor axis is given by

$$a = \tfrac{1}{2}(r_1 + r_2) = \frac{\epsilon d}{1-\epsilon^2}. \qquad \text{(M.45)}$$

The semiminor axis is

$$b = a\sqrt{1-\epsilon^2} \qquad \text{(M.46)}$$

and the area of the ellipse is

$$S = \pi ab = \pi a^2\sqrt{1-\epsilon^2}. \qquad \text{(M.47)}$$

A circle is a special case of an ellipse, when $\epsilon = 0$.

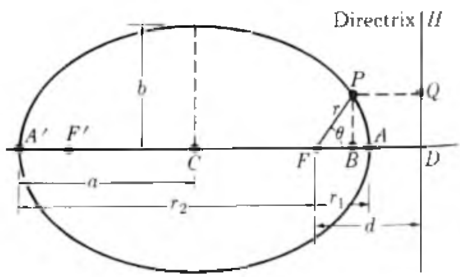


Figure M-6. Geometrical elements of the ellipse.

9. Solid Angles

A *solid angle* is the space included inside a conical (or pyramidal) surface, as in Fig. M-7. Its value, expressed in *steradians* (abbreviated sr), is obtained by drawing, with arbitrary radius R and center at the vertex O, a spherical surface and applying the relation

$$\Omega = \frac{S}{R^2}, \tag{M.47}$$

where S is the area of the spherical cap intercepted by the solid angle. Since the surface area of a sphere is $4\pi R^2$, we conclude that the complete solid angle around a point is 4π steradians. The solid angle formed by the three mutually perpendicular coordinate axes OX, OY, and OZ (Fig. M-8), is $\frac{1}{8}(4\pi)$ or $\pi/2$ steradians.

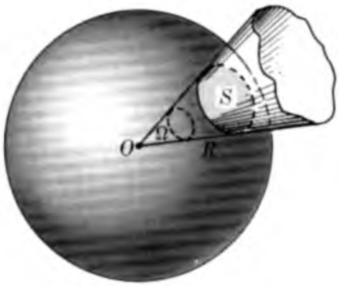


Figure M-7. Solid angle

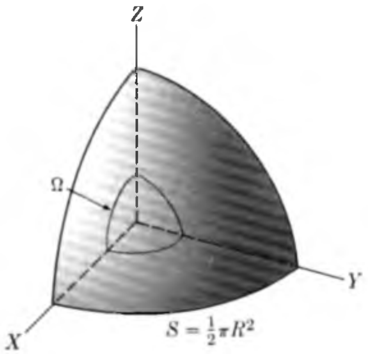


Figure M-8

When the solid angle is small (Fig. M-9), the surface area S becomes dS, and is not necessarily a spherical cap, but may be a small plane surface perpendicular to OP so that

$$d\Omega = \frac{dS}{R^2}. \tag{M.48}$$

In some instances the surface dS is not perpendicular to OP, but its normal N makes an angle θ with OP (Fig. M-10). Then it is necessary to project dS on a plane perpendicular to OP, which gives us the area $dS' = dS \cos\theta$. Thus

$$d\Omega = \frac{dS\cos\theta}{R^2}. \tag{M.49}$$

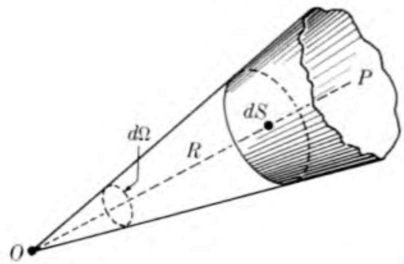


Figure M-9

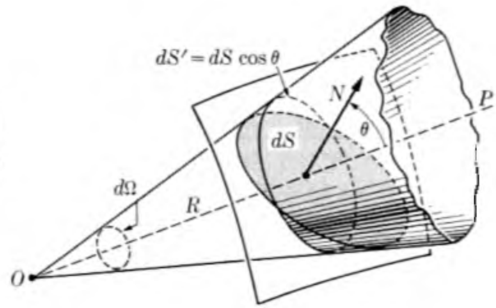


Figure M-10

NATURAL TRIGONOMETRIC FUNCTIONS

Angle					Angle				
Degree	Radian	Sine	Cosine	Tangent	Degree	Radian	Sine	Cosine	Tangent
0°	0.000	0.000	1.000	0.000					
1°	0.017	0.017	1.000	0.017	46°	0.803	0.719	0.695	1.036
2°	0.035	0.035	0.999	0.035	47°	0.820	0.731	0.682	1.072
3°	0.052	0.052	0.999	0.052	48°	0.838	0.743	0.669	1.111
4°	0.070	0.070	0.998	0.070	49°	0.855	0.755	0.656	1.150
5°	0.087	0.087	0.996	0.087	50°	0.873	0.766	0.643	1.192
6°	0.105	0.104	0.994	0.105	51°	0.890	0.777	0.629	1.235
7°	0.122	0.122	0.992	0.123	52°	0.908	0.788	0.616	1.280
8°	0.140	0.139	0.990	0.140	53°	0.925	0.799	0.602	1.327
9°	0.157	0.156	0.988	0.158	54°	0.942	0.809	0.588	1.376
10°	0.174	0.174	0.985	0.176	55°	0.960	0.819	0.574	1.428
11°	0.192	0.191	0.982	0.194	56°	0.977	0.829	0.559	1.483
12°	0.209	0.208	0.978	0.212	57°	0.995	0.839	0.545	1.540
13°	0.227	0.225	0.974	0.231	58°	1.012	0.848	0.530	1.600
14°	0.244	0.242	0.970	0.249	59°	1.030	0.857	0.515	1.664
15°	0.262	0.259	0.966	0.268	60°	1.047	0.866	0.500	1.732
16°	0.279	0.276	0.961	0.287	61°	1.065	0.875	0.485	1.804
17°	0.297	0.292	0.956	0.306	62°	1.082	0.883	0.470	1.881
18°	0.314	0.309	0.951	0.325	63°	1.100	0.891	0.454	1.963
19°	0.332	0.326	0.946	0.344	64°	1.117	0.899	0.438	2.050
20°	0.349	0.342	0.940	0.364	65°	1.134	0.906	0.423	2.144
21°	0.366	0.358	0.934	0.384	66°	1.152	0.914	0.407	2.246
22°	0.384	0.375	0.927	0.404	67°	1.169	0.920	0.391	2.356
23°	0.401	0.391	0.920	0.424	68°	1.187	0.927	0.375	2.475
24°	0.419	0.407	0.914	0.445	69°	1.204	0.934	0.358	2.605
25°	0.436	0.423	0.906	0.466	70°	1.222	0.940	0.342	2.748
26°	0.454	0.438	0.899	0.488	71°	1.239	0.946	0.326	2.904
27°	0.471	0.454	0.891	0.510	72°	1.257	0.951	0.309	3.078
28°	0.489	0.470	0.883	0.532	73°	1.274	0.956	0.292	3.271
29°	0.506	0.485	0.875	0.554	74°	1.292	0.961	0.276	3.487
30°	0.524	0.500	0.866	0.577	75°	1.309	0.966	0.259	3.732
31°	0.541	0.515	0.857	0.601	76°	1.326	0.970	0.242	4.011
32°	0.558	0.530	0.848	0.625	77°	1.344	0.974	0.225	4.332
33°	0.576	0.545	0.839	0.649	78°	1.361	0.978	0.208	4.705
34°	0.593	0.559	0.829	0.674	79°	1.379	0.982	0.191	5.145
35°	0.611	0.574	0.819	0.700	80°	1.396	0.985	0.174	5.671
36°	0.628	0.588	0.809	0.726	81°	1.414	0.988	0.156	6.314
37°	0.646	0.602	0.799	0.754	82°	1.431	0.990	0.139	7.115
38°	0.663	0.616	0.788	0.781	83°	1.449	0.992	0.122	8.144
39°	0.681	0.629	0.777	0.810	84°	1.466	0.994	0.104	9.514
40°	0.698	0.643	0.766	0.839	85°	1.484	0.996	0.087	11.430
41°	0.716	0.656	0.755	0.869	86°	1.501	0.998	0.070	14.301
42°	0.733	0.669	0.743	0.900	87°	1.518	0.999	0.052	19.081
43°	0.750	0.682	0.731	0.933	88°	1.536	0.999	0.035	28.636
44°	0.768	0.695	0.719	0.966	89°	1.553	1.000	0.017	57.290
45°	0.785	0.707	0.707	1.000	90°	1.571	1.000	0.000	∞

COMMON LOGARITHMS

N	0	1	2	3	4	5	6	7	8	9
0	.....	0000	3010	4771	6021	6990	7782	8451	9031	9542
1	0000	0414	0792	1139	1461	1761	2041	2304	2553	2788
2	3010	3222	3424	3617	3802	3979	4150	4314	4472	4624
3	4771	4914	5051	5185	5315	5441	5563	5682	5798	5911
4	6021	6128	6232	6335	6435	6532	6628	6721	6812	6902
5	6990	7076	7160	7243	7324	7404	7482	7559	7634	7709
6	7782	7853	7924	7993	8062	8129	8195	8261	8325	8388
7	8451	8513	8573	8633	8692	8751	8808	8865	8921	8976
8	9031	9085	9138	9191	9243	9294	9345	9395	9445	9494
9	9542	9590	9638	9685	9731	9777	9823	9868	9912	9956
10	0000	0043	0086	0128	0170	0212	0253	0294	0334	0374
11	0414	0453	0492	0531	0569	0607	0645	0682	0719	0755
12	0792	0828	0864	0899	0934	0969	1004	1038	1072	1106
13	1139	1173	1206	1239	1271	1303	1335	1367	1399	1430
14	1461	1492	1523	1553	1584	1614	1644	1673	1703	1732
15	1761	1790	1818	1847	1875	1903	1931	1959	1987	2014
16	2041	2068	2095	2122	2148	2175	2201	2227	2253	2279
17	2304	2330	2355	2380	2405	2430	2455	2480	2504	2529
18	2553	2577	2601	2625	2648	2672	2695	2718	2742	2765
19	2788	2810	2833	2856	2878	2900	2923	2945	2967	2989
20	3010	3032	3054	3075	3096	3118	3139	3160	3181	3201
21	3222	3243	3263	3284	3304	3324	3345	3365	3385	3404
22	3424	3444	3464	3483	3502	3522	3541	3560	3579	3598
23	3617	3636	3655	3674	3692	3711	3729	3747	3766	3784
24	3802	3820	3838	3856	3874	3892	3909	3927	3945	3962
25	3979	3997	4014	4031	4048	4065	4082	4099	4116	4133
26	4150	4166	4183	4200	4216	4232	4249	4265	4281	4298
27	4314	4330	4346	4362	4378	4393	4409	4425	4440	4456
28	4472	4487	4502	4518	4533	4548	4564	4579	4594	4609
29	4624	4639	4654	4669	4683	4698	4713	4728	4742	4757
30	4771	4786	4800	4814	4829	4843	4857	4871	4886	4900
31	4914	4928	4942	4955	4969	4983	4997	5011	5024	5038
32	5051	5065	5079	5092	5105	5119	5132	5145	5159	5172
33	5185	5198	5211	5224	5237	5250	5263	5276	5289	5302
34	5315	5328	5340	5353	5366	5378	5391	5403	5416	5428
35	5441	5453	5465	5478	5490	5502	5514	5527	5539	5551
36	5563	5575	5587	5599	6511	5623	5635	5647	5658	5670
37	5682	5694	5705	5717	5729	5740	5752	5763	5775	5786
38	5798	5809	5821	5832	5843	5855	5866	5877	5888	5899
39	5911	5922	5933	5944	5955	5966	5977	5988	5999	6010
40	6021	6031	6042	6053	6064	6075	6085	6096	6107	6117
41	6128	6138	6149	6160	6170	6180	6191	6201	6212	6222
42	6232	6243	6253	6263	6274	6284	6294	6304	6314	6325
43	6335	6345	6355	6365	6375	6385	6395	6405	6415	6425
44	6435	6444	6454	6464	6474	6484	6493	6503	6513	6522
45	6532	6542	6551	6561	6571	6580	6590	6599	6609	6618
46	6628	6637	6646	6656	6665	6675	6684	6693	6702	6712
47	6721	6730	6739	6749	6758	6767	6776	6785	6794	6803
48	6812	6821	6830	6839	6848	6857	6866	6875	6884	6893
49	6902	6911	6920	6928	6937	6946	6955	6964	6972	6981
50	6990	6998	7007	7016	7024	7033	7042	7050	7059	7067
N	0	1	2	3	4	5	6	7	8	9

COMMON LOGARITHMS (*continued*)

N	0	1	2	3	4	5	6	7	8	9
50	6990	6998	7007	7016	7024	7033	7042	7050	7059	7067
51	7076	7084	7093	7101	7110	7118	7126	7135	7143	7152
52	7160	7168	7177	7185	7193	7202	7210	7218	7226	7235
53	7243	7251	7259	7267	7275	7284	7292	7300	7308	7316
54	7324	7332	7340	7348	7356	7364	7372	7380	7388	7396
55	7404	7412	7419	7427	7435	7443	7451	7459	7466	7474
56	7482	7490	7497	7505	7513	7520	7528	7536	7543	7551
57	7559	7566	7574	7582	7589	7597	7604	7612	7619	7627
58	7634	7642	7649	7657	7664	7672	7679	7686	7694	7701
59	7709	7716	7723	7731	7738	7745	7752	7760	7767	7774
60	7782	7789	7796	7803	7810	7818	7825	7832	7839	7846
61	7853	7860	7868	7875	7882	7889	7896	7903	7910	7917
62	7924	7931	7938	7945	7952	7959	7966	7973	7980	7987
63	7993	8000	8007	8014	8021	8028	8035	8041	8048	8055
64	8062	8069	8075	8082	8089	8096	8102	8109	8116	8122
65	8129	8136	8142	8149	8156	8162	8169	8176	8182	8189
66	8195	8202	8209	8215	8222	8228	8235	8241	8248	8254
67	8261	8267	8274	8280	8287	8293	8299	8306	8312	8319
68	8325	8331	8338	8344	8351	8357	8363	8370	8376	8382
69	8388	8395	8401	8407	8414	8420	8426	8432	8439	8445
70	8451	8457	8463	8470	8476	8482	8488	8494	8500	8506
71	8513	8519	8525	8531	8537	8543	8549	8555	8561	8567
72	8573	8579	8585	8591	8597	8603	8609	8615	8621	8627
73	8633	8639	8645	8651	8657	8663	8669	8675	8681	8686
74	8692	8698	8704	8710	8716	8722	8727	8733	8739	8745
75	8751	8756	8762	8768	8774	8779	8785	8791	8797	8802
76	8808	8814	8820	8825	8831	8837	8842	8848	8854	8859
77	8865	8871	8876	8882	8887	8893	8899	8904	8910	8915
78	8921	8927	8932	8938	8943	8949	8954	8960	8965	8971
79	8976	8982	8987	8993	8998	9004	9009	9015	9020	9025
80	9031	9036	9042	9047	9053	9058	9063	9069	9074	9079
81	9085	9090	9096	9101	9106	9112	9117	9122	9128	9133
82	9138	9143	9149	9154	9159	9165	9170	9175	9180	9186
83	9191	9196	9201	9206	9212	9217	9222	9227	9232	9238
84	9243	9248	9253	9258	9263	9269	9274	9279	9284	9289
85	9294	9299	9304	9309	9315	9320	9325	9330	9335	9340
86	9345	9350	9355	9360	9365	9370	9375	9380	9385	9390
87	9395	9400	9405	9410	9415	9420	9425	9430	9435	9440
88	9445	9450	9455	9460	9465	9469	9474	9479	9484	9489
89	9494	9499	9504	9509	9513	9518	9523	9528	9533	9538
90	9542	9547	9552	9557	9562	9566	9571	9576	9581	9586
91	9590	9595	9600	9605	9609	9614	9619	9624	9628	9633
92	9638	9643	9647	9652	9657	9661	9666	9671	9675	9680
93	9685	8689	9694	9699	9703	9708	9713	9717	9722	9727
94	9731	9736	9741	9745	9750	9754	9759	9763	9768	9773
95	9777	9782	9786	9791	9795	9800	9805	9809	9814	9818
96	9823	9827	9832	9836	9841	9845	9850	9854	9859	9863
97	9868	9872	9877	9881	9886	9890	9894	9899	9903	9908
98	9912	9917	9921	9926	9930	9934	9939	9943	9948	9952
99	9956	9961	9965	9969	9974	9978	9983	9987	9991	9996
100	0000	0004	0009	0013	0017	0022	0026	0030	0035	0039
N	0	1	2	3	4	5	6	7	8	9

EXPONENTIAL FUNCTIONS

x	e^x	e^{-x}	x	e^x	e^{-x}
0.00	1.0000	1.0000	2.5	12.182	0.0821
0.05	1.0513	0.9512	2.6	13.464	0.0743
0.10	1.1052	0.9048	2.7	14.880	0.0672
0.15	1.1618	0.8607	2.8	16.445	0.0608
0.20	1.2214	0.8187	2.9	18.174	0.0550
0.25	1.2840	0.7788	3.0	20.086	0.0498
0.30	1.3498	0.7408	3.1	22.198	0.0450
0.35	1.4191	0.7047	3.2	24.532	0.0408
0.40	1.4918	0.6703	3.3	27.113	0.0369
0.45	1.5683	0.6376	3.4	29.964	0.0334
0.50	1.6487	0.6065	3.5	33.115	0.0302
0.55	1.7332	0.5769	3.6	36.598	0.0273
0.60	1.8221	0.5488	3.7	40.447	0.0247
0.65	1.9155	0.5220	3.8	44.701	0.0224
0.70	2.0138	0.4966	3.9	49.402	0.0202
0.75	2.1170	0.4724	4.0	54.598	0.0183
0.80	2.2255	0.4493	4.1	60.340	0.0166
0.85	2.3396	0.4274	4.2	66.686	0.0150
0.90	2.4596	0.4066	4.3	73.700	0.0136
0.95	2.5857	0.3867	4.4	81.451	0.0123
1.0	2.7183	0.3679	4.5	90.017	0.0111
1.1	3.0042	0.3329	4.6	99.484	0.0100
1.2	3.3201	0.3012	4.7	109.947	0.0091
1.3	3.6693	0.2725	4.8	121.510	0.0082
1.4	4.0552	0.2466	4.9	134.289	0.0074
1.5	4.4817	0.2231	5	148.413	0.0067
1.6	4.9530	0.2019	6	403.428	0.0025
1.7	5.4739	0.1827	7	1096.633	0.0009
1.8	6.0496	0.1653	8	2980.957	0.0003
1.9	6.6859	0.1496	9	8103.084	0.0001
2.0	7.3891	0.1353	10	22026.466	0.00004
2.1	8.1662	0.1224			
2.2	9.0250	0.1108			
2.3	9.9742	0.1002			
2.4	11.023	0.0907			

ANSWERS TO PROBLEMS

INDEX

ANSWERS TO PROBLEMS

Chapter One

1.1 4.22×10^{-8} N; 3.41×10^{-44} N; electrical repulsion is 1.2×10^{36} times stronger than gravitational attraction.
1.2 8.22×10^{-8} N; 3.62×10^{-48} N; electrical attraction is 2.3×10^{39} times stronger than gravitational attraction.
1.3 Electrical repulsion is 4.17×10^{42} stronger for electrons and 1.24×10^{36} for protons.
1.4 $\sin^2 \theta \tan \theta = (K_e q^2)/(4mgl^2)$.
1.5 -39.2 μC.
1.6 (a) 4.81×10^{-15} N; (b) 5.28×10^{15} m s^{-2}; about 5.4×10^{14} times stronger than g.
1.7 (a) 0; (b) -10^{-3} J; (c) 2.3×10^{-3} J.
1.8 (a) 1.01×10^{3} N C^{-1}; (b) 2.67×10^{6} m s^{-1}.
1.9 (a) 7.04×10^{-3} m; (b) 19.4° (0.34 r); (c) 4.22×10^{-2} m.
1.10 (a) 5.69×10^{-9} s; (b) 1.42×10^{-2} m; (c) 9.87×10^{-2} m; (d) path is a parabola.
1.11 (a) 4.33×10^{-4} m s^{-1}; (b) 2.74×10^{-4} m s^{-1} upward.
1.12 2.
1.13 (a) 1.61×10^{-6} m, 1.40×10^{-14} kg; (b) weight is 4.3 times larger.
1.14 (a) 1.88×10^{5} N C^{-1} at 4.76° above the parallel defined by the line joining the charges; (b) on the line joining the charges, 2.414 m from the 5 μC charge.
1.15 951 V.
1.16 (a) 3.15×10^{-2} N, 9.0×10^{-3} N, 4.05×10^{-2} N; (b) 4.50×10^{-3} J, 4.50×10^{-3} J, 5.40×10^{-3} J; (c) 7.2×10^{-3} J; (d) the sum of terms in answer (b) contains each interaction twice.
1.17 (a) 4.50×10^{-3} N, 9.0×10^{-3} N, 1.35×10^{-2} N; (b) 9.00×10^{-4} J, -4.50×10^{-3} J, 0 J; (c) -1.80×10^{-3} J.
1.18 (a) 3.0 m; (b) 0.2 μC.
1.19 (a) 3.60×10^{5} N C^{-1}; (b) 6.33×10^{16} m s^{-2} radially inward; (c) $0.97c$.
1.20 2880 V.
1.21 (a) 3.6×10^{5} N C^{-1} toward 0.2 μC, 9×10^{4} V; (b) 3.75×10^{5} N C^{-1} toward 0.3 μC, 9×10^{4} V; (c) 1.26×10^{6} N C^{-1} away from 0.2 μC, 6.43×10^{4} V; (d) 3.92×10^{5} N C^{-1} at 6.6° from vertical, 4.5×10^{4} V.
1.22 (a) 1.8×10^{6} N C^{-1} toward 0.3 μC, -1.8×10^{4} V; (b) 1.87×10^{6} N C^{-1} toward 0.3 μC, 0 V; (c) 9.9×10^{5} N C^{-1} away from 0.2 μC, 2.57×10^{4} V; (d) 2.38×10^{5} N C^{-1} at 19.1° below horizontal, -9×10^{3} V.
1.23 -7.72×10^{-3} J; no.

1.24 (a) NA; (b) $2K_eq/a$; (c) NA; (d) NA; (e) $\sqrt{3}\,a$; (f) $2K_eqx/(x^2 + a^2)^{3/2}$.
1.25 (a) $2\sqrt{K_eqq'/am}$; (b) NA; (c) $\sqrt{15}\,a$; (d) same as in (a) where q' and m are the negative charge and its mass.
1.26 (a) NA; (b) there are infinities at $y = \pm a$; (c) yes.
1.27 (a) $\boldsymbol{u}_r 50$ N C^{-1}; (b) $\boldsymbol{u}_r 10.8$ N C^{-1}.
1.28 (a) 1.56 N directly away from CM ; 0.18 J; (b) 0 N C^{-1}, 4.67×10^5 V; (c) 0.27 J.
1.29 (a) From the CM, outside the triangle, there are 6 radial lines; (b) at a great distance, the equipotentials are circles centered at the CM.
1.30 6.455×10^7 m s^{-1} (6.57×10^7 m s^{-1} if calculated classically).
1.31 (a) 2.18×10^{-5} m; (b) electron's velocity is 42.9 times faster; (c) same energy, equal to 1600 eV or 2.56×10^{-16} J.
1.32 (a) 6.24×10^{15} protons/second; (b) 800 W; (c) 1.24×10^7 m s^{-1}; (d) 153 cal s^{-1}.
1.33 (a) 3.08×10^7 m s^{-1}, 1.22×10^7 m s^{-1}; (b) 3.32×10^7 m s^{-1} at 21.6° with direction of field; (c) $x \times 0.769$ m, $y = 0.610$ m; (d) 5.02×10^{-16} J or 3.13 keV.
1.34 (a) 1.30×10^7 m s^{-1}; (b) 1.63×10^7 m s^{-1}; (c) 9.83×10^6 m s^{-1}; (d) 3.47×10^{-9} s; (e) will not reach the cathode.
1.35 (a) In the general case $\boldsymbol{F} = (-p^2/4\pi\epsilon_0 r^4)[\boldsymbol{u}_r(6 \cos\theta \cos\theta_2 - 3 \sin\theta \sin\theta_2) + \boldsymbol{u}_\theta(2 \sin\theta \cos\theta_2 + \cos\theta \sin\theta_2)]$ where θ_2 is the angle of the second dipole relative to the orientation of the first. (b) Gravitational attraction is 6×10^{-43} N; dipole-dipole force has a maximum value of 4×10^{-11} N, about 10^{32} stronger.
1.36 (a) 0, $3qa^2$; (b) $Q/(4\pi\epsilon_0 z^3)$, $-3Q/4\pi\epsilon_0 z^4$; (c) $-Q/4\pi\epsilon_0 y^3$, $3Q/4\pi\epsilon_0 y^4$.
1.37 (a) There is no electric dipole; qa^2; (b) $(4q/4\pi\epsilon_0 z) + (Q/4\pi\epsilon_0 z^3)$, $(-4q/4\pi\epsilon_0 z^2) + (-3Q/4\pi\epsilon_0 z^4)$; (c) same as in (b), with y replacing z.

Chapter Two

2.1 (a) 0, 0; (b) Ca^3, $\epsilon_0 Ca^3$.
2.2 (a) ca^3, $\epsilon_0 ca^3$, $2\epsilon_0 cx$; (b) 0, 0, 0.
2.3 NA.
2.4 (a) Inner: $-q$, outer: $+q$; (b) electric field: outside: K_eq/r^2, in sphere: 0, in cavity: K_eq/r^2; electric potential: outside: K_eq/r, in sphere: K_eq/R_1, in cavity: K_eq/r; (c) NA.
2.5 (a) NA; (b) NA.
2.6 1.2×10^{-7} C on the smaller and 1.8×10^{-7} C on the larger.
2.7 NA.
2.8 (a) 16.5; (b) 15.5.
2.9 (a) 1000 V; (b) 2000 V.
2.10 (a) 0.485 m^2; (b) 300 μC.
2.11 1.70 m^2.
2.12 (a) NA; (b) NA.
2.13 9.58×10^{-11} F.

2.14 (a) Series: $\frac{2}{3}$ μF, parallel: 6.5 μF; (b) series: 13.3 μC on each, 9 V, 6.67 V, 4.44 V; parallel: 30 μC, 40 μC, 60 μC, 20 V on each; (c) series: 1.33×10^{-4} J, parallel: 1.30×10^{-3} J.
2.15 (a) 10 μF; (b) 12 μF: 480 μC, 18 μF: 720 μC, 4 μF: 320 μC, 5 μF: 400 μC, 1 μF: 80 μC, 2 μF: 160 μC, 3 μF: 240 μC; (c) 12 μF and 18 μF: 40 V; all others: 80 V.
2.16 C_1: 600 μC, 200 V; C_2: 200 μC, 100 V; C_3: 400 μC, 100 V.
2.17 NA.
2.18 NA.
2.19 (a) 380 μC; (b) 7600 V; (c) 1.44×10^{-3} J; (d) 1.36×10^{-3} J.
2.20 (a) 2×10^{-2} C; (b) 800 V; (c) 8 J; (d) 2 J.
2.21 (a) 7.7×10^{-10} C and 2.3×10^{-10} C; (b) 4.5×10^{-9} J; (c) 3.46×10^{-9} J; (d) through radiation and Joule heating.

Chapter Three

3.1 2.06×10^{-14} s.
3.2 (a) Resistance doubles; (b) resistance halves; (c) resistance quarters.
3.3 (a) 5 Ω; (b) 3 Ω: 60 A, 12 Ω: 10 A, 6 Ω: 20 A, 4 Ω: 30 A; (c) 3 Ω: 180 V, all others: 120 V.
3.4 (a) 6 Ω; (b) 12 Ω: 12 A, 10 Ω: 12 A, 6 Ω: 4 A, 3 Ω: 8 A; (c) 12 Ω: 144 V, 10 Ω: 120 V, 6 Ω and 3 Ω: 24 V.
3.5 (a) 10 Ω; (b) 4 Ω: 12 A, 9 Ω: 8 A, 16 Ω: 4 A, 6 Ω: 1.33 A, 3 Ω: 2.67 A; (c) 4 Ω: 48 V, 9 Ω: 72 V, 16 Ω: 64 V, 6 Ω: 8 V, 3 Ω: 8 V.
3.6 (a) 9 Ω; (b) 5 Ω: 60 A, 20 Ω: 12 A, 3 Ω: 48 A, 12 Ω: 8 A, 6 Ω: 16 A, 4 Ω: 24 A; (c) 5 Ω: 300 V, 20 Ω: 240 V, 3 Ω: 144 V, 12 Ω: 96 V, 6 Ω: 96 V, 4 Ω: 96 V.
3.7 (a) 10 Ω; (b) 7 Ω: 24 A, 12 Ω: 6 A, 10 Ω: 6 A, 3 Ω: 4 A, 6 Ω: 2 A, 18 Ω: 4 A, 9 Ω: 8 A; (c) 7 Ω: 168 V, 12 Ω: 72 V, 10 Ω: 60 V, 3 Ω: 12 V, 6 Ω: 12 V, 18 Ω: 72 V, 9 Ω: 72 V.
3.8 (a) 10 Ω; (b) 10 A in all resistors except the center one, where the current is zero; (c) 100 V, 0 V.
3.9 (a) 8 Ω; (b) 12 V.
3.10 (a) 32 Ω; (b) 20 V.
3.11 27 W.
3.12 90 W.
3.13 NA.
3.14 Place a 4 Ω resistor in parallel with the instrument.
3.15 Place a 24 Ω resistor in series with the instrument.
3.16 (a) NA; no; (b) 1.10 Ω; (c) 1.24 Ω.
3.17 (a) 350 K: 1050 Ω; 370 K: 39 Ω; 470 K: 11.1 Ω; 570 K: 3.82 Ω; (b) NA; (c) 0.1.
3.18 NA.

3.19 NA.
3.20 (a) 2.829 V; (b) 1.33 V.
3.21 (a) 0.5 Ω; (b) 10 V.
3.22 (a) $\frac{1}{3}$ A; (b) 2.67 V; (c) 22 Ω: $\frac{1}{9}$ A, 12 Ω: $\frac{1}{54}$ A, 6 Ω: $\frac{1}{27}$ A, 4 Ω: $\frac{1}{18}$ A, 8 Ω: $\frac{2}{9}$ A, 5 Ω: $\frac{8}{45}$ A, 20 Ω: $\frac{2}{45}$ A.
3.23 2 Ω: 3 A; 9 Ω: 2 A; 4 Ω: 5 A.
3.24 4 Ω: $\frac{10}{38}$ A; 12 Ω: $\frac{35}{38}$ A; 10 Ω: $\frac{45}{38}$ A.
3.25 4 Ω: 1.137 A; $\frac{1}{2}$ Ω: 0.58 A; 3 Ω: 0.558 A; 2 Ω: 0.308 A; 1 Ω: 0.887 A; 6 Ω: 0.251 A.
3.26 (a) 0.22 V; (b) $\frac{13}{28}$ A.
3.27 (a) −12 V; (b) ∞; (c) −12 V; (d) $\frac{12}{7}$ A; (e) 4.5 Ω; (f) 4.2 Ω.

Chapter Four

4.1 (a) 5.68×10^{-5} T; (b) 10^7 rad s^{-1}.
4.2 (a) 7.22×10^{-2} m; (b) 9.58×10^{13} rad s^{-1}.
4.3 (a) 3.48×10^{-2} m; (b) 3.79×10^{-1} m; (c) 2.29×10^7 Hz.
4.4 (a) 2.89×10^7 m s^{-1}; (b) 4.34×10^{-8} s; (c) 8.66×10^6 V.
4.5 (a) 0.528 m; (b) 4.37×10^{-8} s.
4.6 (a) 1 T; (b) 3×10^6 rad s^{-1}.
4.7 2.13×10^{-2} m.
4.8 (a) 5.68×10^{-4} T into the page; (b) 3.14×10^{-8} s.
4.9 (a) 6.38×10^{-3} m; (b) 1.28×10^{-2} m.
4.10 1: $\boldsymbol{u}_z(qv\mathscr{B})$; 2: 0; 3: $\boldsymbol{u}_z(qv\mathscr{B}/\sqrt{2})$; 4: $-\boldsymbol{u}_x(qv\mathscr{B}/\sqrt{2})$; 5: $-\boldsymbol{u}_x(qv\mathscr{B})$; 6: $-\boldsymbol{u}_x(qv\mathscr{B}/\sqrt{3}) - \boldsymbol{u}_z(qv\mathscr{B}/\sqrt{3})$.
4.11 Since the charge will move in a plane (perpendicular to $\mathscr{B}$) in a circular path of radius R. $\boldsymbol{v} = \boldsymbol{u}_x(v_0 \cos \omega t) + \boldsymbol{u}_y(v_0 \sin \omega t)$ where $\omega = (q/m)$ and $\boldsymbol{r} = \boldsymbol{u}_x(R \sin \omega t) - \boldsymbol{u}_y(R \cos \omega t)$ [at $t = 0$, $x = 0$, $y = -R$].
4.12 $\boldsymbol{v} = \boldsymbol{u}_x(v_0 \cos \alpha \cos \omega t) + \boldsymbol{u}_y(v_0 \cos \alpha \sin \omega t) + \boldsymbol{u}_z(v_0 \sin \alpha)$; $\boldsymbol{r} = \boldsymbol{u}_x(R \sin \omega t) - \boldsymbol{u}_y(R \cos \omega t) + \boldsymbol{u}_z(v_0[\sin \alpha]t)$ where $R = (m/q)(v_0 \cos \alpha/\mathscr{B})$.
4.13 $\mathscr{B} = -\boldsymbol{u}_z(0.5)$ T.
4.14 1.5×10^5 m s^{-1}.
4.15 (a) 1.75×10^7 m s^{-1}; (b) NA.
4.16 3.27 T horizontal and normal to the velocity.
4.17 (a) $1.02R$; (b) 960 V.
4.18 (a) classical velocities, around 10^5 m s^{-1}; (b) 0.245 m.
4.19 (a) 0.656 T; (b) 2.51×10^7 m s^{-1}; (c) 5.28×10^{-13} J, 3.30 MeV; (d) about 82 full turns.
4.20 For deuteron: (a) 1.31 T; (b) 2.51×10^7 m s^{-1}; (c) 1.06×10^{-12} J, 6.59 MeV; (d) about 165 turns. For alpha: (a) 1.30 T; (b) 2.51×10^7 m s^{-1}; (c) 2.10×10^{-12} J, 13.1 MeV; (d) 164 turns.
4.21 (a) 4.57×10^7 per second (i.e., a frequency of 2.29×10^7 Hz); (b) 5.03×10^7 m s^{-1}; (c) 1.32×10^7 V.
4.22 (a) 1.31 T; (b) 4.22 MeV; (c) 2.01×10^7 m s^{-1}.

4.23 (a) NA; (b) NA; (c) NA.
4.24 (a) Because of the "focusing" of off-axis electrons; (b) they are mirror images of each other.
4.25 1.57 m.
4.26 (a) For O': with q_1 fixed at O', $\boldsymbol{F}_1' = -\boldsymbol{F}_2' = -\boldsymbol{u}_x(q_1q_2/4\pi\epsilon_0 r')$; for O: $\boldsymbol{F}_1 = -\boldsymbol{F}_2 = \boldsymbol{F}_1'$; (b) For O': $\boldsymbol{F}_1' = -\boldsymbol{F}_2' = -\boldsymbol{u}_y(q_1q_2/4\pi\epsilon_0 r')$; for O: $\boldsymbol{F}_1 = -\boldsymbol{F}_2 = \sqrt{1-(v^2/c^2)}\,\boldsymbol{F}_1'$.
4.27 (a) 1; (b) 1.015; (c) 1.54; (d) 12.08.
4.28 (a) 1; (b) 1.005; (c) 1.155; (d) 2.294.
4.29 (a) 1; (b) 1.005; (c) 1.155; (d) 2.294.
4.30 (a) $2.30997433 \times 10^{-20}$ N repulsion; (b) electric force only, equal to $2.31000000 \times 10^{-20}$ N repulsion; (c) 1.386×10^{-20} N repulsion measured by outside observer, versus 2.31×10^{-20} N measured by internal observer.

Chapter Five

5.1 3.78×10^8 Am^{-2}.
5.2 1: $-\boldsymbol{u}_y(0.3)$ N; 2: 0 N; 3: $\boldsymbol{u}_x(0.3)$ N; 4: $(\boldsymbol{u}_x - \boldsymbol{u}_y)$ 0.3 N; 5: $(\boldsymbol{u}_x + \boldsymbol{u}_y)$ 0.3 N.
5.3 (a) 6×10^{-3} N m; (b) 0.04 Am^2; (c) 8.07×10^{-3} N m.
5.4 (a) On vertical sides: 0.16 N in Z-direction; on horizontal sides: 0.06 N in Y-direction; (b) $\boldsymbol{u}_y$ (8.31×10^{-3}) N m; (c) on vertical sides: 0.16 N in X-direction; on horizontal sides: 0.104 N in Y-direction; (d) $-\boldsymbol{u}_y$ (4.8×10^{-3}) N m; (e) same for both cases.
5.5 (a) $\boldsymbol{u}_y$ (1.32×10^{-2}) T; (b) a magnetic field of $-\boldsymbol{u}_x$ (2.29×10^{-2}) T will hold the loop at any angle or even allow it to rotate uniformly about the Z-axis.
5.6 3.6×10^{-6} N m.
5.7 0.30°.
5.8 (a) $\boldsymbol{u}_x$ 10^{-2} N; (b) $-\boldsymbol{u}_z$ (5×10^{-4}) N m.
5.9 (a) $-\boldsymbol{u}_z$ 10^{-2} N; (b) $[\boldsymbol{u}_x(-5 \times 10^{-4}) + \boldsymbol{u}_y(10^{-3})]$ N m.
5.10 (a) $\mathcal{B}I(2R)$ where R is the radius of the circle; (b) NA.
5.11 (a) Yes; (b) no; (c) $\tau = [3.6 \times 10^{-6} \sin\theta]$ N m; (d) NA; (e) the coil will continue to rotate until restoring torque changes direction of rotation.
5.12 (a) The coil will continue to rotate in one direction; (b) use of "split ring" and "brushes"; (c) could use the device as a motor.
5.13 (a) 4.0×10^{-6} T, 2×10^{-7} T; (b) 0 N C^{-1}.
5.14 2.40×10^{-20} N, resulting in an acceleration normal to the motion equal to 2.64×10^{10} m s^{-2}.
5.15 (a) $\boldsymbol{u}_x$ (2×10^{-6}) T; (b) $(\boldsymbol{u}_x - \boldsymbol{u}_z)$ 10^{-6} T; (c) $-\boldsymbol{u}_x$ (3×10^{-6}) T.
5.16 (a) $\mu_0 I/\pi a$; (b) $\mu_0 I/3\pi a$; (c) 0, $2\mu_0 I/3\pi a$.
5.17 (a) 2 A out of the page; (b) 2.13×10^{-6} T; (c) 1.64×10^{-6} T in toward the wires at an angle of 13° below the horizontal.
5.18 (a) NA; (b) $\mu_0 Ia/\pi(a^2 + x^2)$; (c) NA; (d) $x = 0$.

5.19 (a) NA; (b) $\mu_0 Ia/\pi(a^2 - y^2)$; (c) NA; (d) $y = \pm a$.
5.20 (a) NA; (b) $\mu_0 Ia/\pi(a^2 + x^2)$; (c) NA; (d) $x = 0$.
5.21 (a) NA; (b) $\mu_0 Iy/\pi(a^2 - y^2)$; (c) NA; (d) $y = \pm a$.
5.22 (a) 1.92×10^{-4} N m^{-1} in the $+Y$-direction; (b) 1.92×10^{-4} N m^{-1} in the $-Y$-direction.
5.23 30.
5.24 6.91×10^{-3}.
5.25 1.02×10^{-2} m.
5.26 (a) 3.2×10^{-5} N repulsion; (b) 0, 0.
5.27 46.4 A.

Chapter Six

6.1 (a) 0, 2.5×10^{-5} T, 5×10^{-5} T; (b) 5×10^{-5} T, 2.5×10^{-5} T, 1.33×10^{-6} T; (c) 4 m and 4×10^{-4} m.
6.2 1.7×10^{-3} T.
6.3 4.40×10^{-2} T.
6.4 NA.
6.5 (a) $r > \Delta R_3$: 0; $R_2 < \Delta r < R_3$: $(\mu_0 I/2\pi r)[(R_3^2 - r^2)/(R_3^2 - R_2^2)]$; $R_1 < \Delta r < R_2$: $(\mu_0 I/2\pi r)$; $0 < r < \Delta R_1$: $(\mu_0/2\pi)[Ir/R_1^2]$; (b) NA.
6.6 (a) 2.84×10^3 turns/meter; (b) 1.79×10^4 m.
6.7 NA.
6.8 NA.
6.9 (a) 0.24 Wb; (b) 0; (c) 0.24 Wb.
6.10 $[\mu_0 Ib/2\pi] \ln (1 + a/r)$.
6.11 0.
6.12 NA.
6.13 NA.
6.14 NA.

Chapter Seven

7.1 (a) 2.36×10^{-3} kg; (b) 2.24×10^{22}.
7.2 (a) 1.12 A; (b) 9.65×10^4 C.
7.3 (a) 1.139×10^{-13} m; (b) 1.139×10^{-13} m; (c) 1.1478×10^{-13} m; (d) 178°, 160°.
7.4 (a) 2.88×10^{-14} m; (b) about 3 nuclear radii away.
7.5 About 14.
7.6 (a) -27.2 eV (-4.36×10^{-18} J); (b) 13.6 eV (2.18×10^{-18} J); (c) -13.6 eV; (d) 6.57×10^{14} Hz.
7.7 (a) $E_{av} = \frac{1}{2}E_{p_{av}} = -\frac{1}{2}K_e Ze^2/r$; (b) -13.6 eV (same value).
7.8 (a) 8.22×10^6 orbits; (b) 1.95×10^4 orbits.
7.9 NA.
7.10 (a) H: -13.6 eV, -3.40 eV, -1.51 eV, -0.85 eV; He: -54.4 eV.

-13.6 eV, -6.04 eV, -3.40 eV; (b) H: 10.2 eV, He: 40.8 eV; (c) NA; (d) the mth even level of He coincides with the mth energy level of H.
7.11 (a) 2.19×10^6 m s^{-1}; (b) NA.
7.12 (a) 8.8×10^{10} rad s^{-1}; (b) 1.34×10^8 rad s^{-1}.
7.13 (a) 9.284×10^{-24} A m^2; (b) 8.79×10^5 rad s^{-1}.
7.14 1.
7.15 (a) $2.45\,\hbar$; (b) $m_l = 0$: 90°, $m_l = \pm 1$: 67.8°, $m_l = \pm 2$: 50.8°.
7.16 (a) 17, if spin is not considered; (b) 2.26×10^{-4} eV.
7.17 $2.17 \times 10^4 : 1$.

Chapter Eight

8.1 (a) 4π; (b) 4π; (c) 8π; (d) 4π; (e) 8π.
8.2 $(\mu_0/2\pi) b I_0 \omega \ln[1 + a/r] \cos \omega t$.
8.3 NA.
8.4 (a) Circles (pointing clockwise); (b) 5×10^{-3} N C^{-1} clockwise, 3.14×10^{-3} V; (c) 1.57 m A; (d) 0; (e) in a time-dependent field, potential difference is not a useful concept; (f) 3.14×10^{-3} V.
8.5 NA.
8.6 (a) 0.65 V; (b) $0.2t$ V; (c) 0; (d) 0.
8.7 (a) $V = 6\pi \sin(10\omega t) - 0.4\pi \sin(10\omega t)\cos(10\omega t)$, $I = V/R$; (b) 11.6 V
8.8 (a) end a; (b) 45 kV.
8.9 A: 0 V; C: 7.1×10^{-2} V; D: 0.1 V.
8.10 $(\mu_0 I b a v/2\pi)[(r_0 + vt)(a + r_0 + vt)]^{-1}$, where $r_0 = r$ at $t = 0$.
8.11 NA.

Chapter Nine

9.1 (a) 0.4 H; (b) 4 V; (c) 4 V; (d) 8 V.
9.2 $\mu_0 n A$, where n is the number of turns per unit length and A is the cross sectional area of the toroid.
9.3 (a) NA; (b) NA; (c) NA; (d) NA.
9.4 (a) NA; (b) $q = (q_0/C_2 - C_1)[C_2 e^{(C_2 - C_1)t/(C_1C_2R)}] - C_1$, $I = (q_0/C_1R)(e^{(C_2 - C_1)t/(C_1C_2R)})$
9.5 (a) NA; (b) NA; (c) NA.
9.6 (a) NA; (b) NA; (c) NA; (d) NA; (e) NA.
9.7 (a) NA; (b) NA; (c) R.
9.8 (a) $I = -VC\omega \cos \omega t$; (b) NA; (c) NA.
9.9 (a) 5 mA; (b) 50 mA; (c) 0.5 A; (d) NA.
9.10 (a) 50 mA; (b) 5 mA; (c) 0.5 mA; (d) NA.
9.11 (a) 0.5 mA; (b) 50 mA; (c) 5 A; (d) NA.
9.12 (a) 5 A; (b) 50 mA; (c) 0.5 mA; (d) NA.
9.13 (a) $I = (V_0/\omega L) \cos \omega t$; (b) NA; (c) NA.
9.14 NA.

9.15 (a) $[R^2 + (1/\omega C)^2]^{1/2}$, current leads by $\tan^{-1}(1/\omega RC)$; (b) $[\omega L - (1/\omega C)]$, angle is $\pm(\pi/2)$.
9.16 (a) 4.52×10^{-4} H; (b) 9.04×10^{-4} V.
9.17 (a) 7.20×10^{-2} H; (b) 1.44×10^{-3} Wb in each turn; (c) 0.48 V.
9.18 (a) $L_1N_2I_0/5N_1t$; (b) $(L_1N_2I_0\omega/5N_1)\cos\omega t$.
9.19 NA.
9.20 NA.

Chapter Ten

10.1 3.33 m.
10.2 (a) 0.10 m; (b) 0.5 m; (c) 100 Hz; (d) 50 m s^{-1}; (e) NA.
10.3 (a) 10 m; (b) 5 Hz; (c) 0.2 s; (d) 50 m s^{-1}; (e) 0.02 m; (f) $+X$-direction: (g) $\xi = 0.02\sin[2\pi(0.1x + 5t)]$ m.
10.4 (a) NA; (b) NA; (c) NA; (d) NA; (e) two waves are the same, but move in opposite directions.
10.5 NA.
10.6 NA.
10.7 (a) NA; (b) NA.
10.8 (a) 0, 8.87×10^{-3} m, 1.69×10^{-2} m, 2.35×10^{-2} m; (b) 8.87×10^{-3} m, 3.00×10^{-3} m, -3.00×10^{-3} m; (c) $v = -0.06\cos(3x - 2t)$ m s^{-1}; (d) 6×10^{-2} m s^{-1}; (e) $\frac{2}{3}$ m s^{-1}.
10.9 (a) NA; (b) $\xi(x,t) = (8A/\pi^2)[\sin\theta - (\frac{1}{9}\sin 3\theta + (\frac{1}{25}\sin 5\theta - \ldots]$ where $\theta = kx - \omega t$.
10.10 (a) NA; (b) $\xi(x,t) = (4A/\pi)[\sin\theta + \frac{1}{3}\sin 3\theta + \frac{1}{5}\sin 5\theta + \ldots]$ where $\theta = kx - \omega t$.
10.11 (a) $F = F_0\cos(2\pi[x/\lambda - t/P])$ where $F_0 = (2\pi\lambda A\rho\xi_0/P^2)$; (b) NA; (c) NA.
10.12 35.4 m s^{-1}.
10.13 39.6 m s^{-1}.
10.14 (a) 1255.5 m s^{-1} (99%); (b) 336.8 m s^{-1} (99.3%); (c) 351.1 m s^{-1} (99%).
10.15 0.58 m s^{-1} K^{-1}.
10.16 28.9 amu compared to 28.8 amu for 80% N_2 & 20% O_2.
10.17 (a) Increases by $\sqrt{2}$; (b) decreases by $\sqrt{2}$; (c) increases 4×; (d) decreases to $\frac{1}{4}$.
10.18 903.4 m s^{-1}.
10.19 99 m s^{-1}.
10.20 (a) 38.3 m s^{-1}; (b) 0.16 m.
10.21 (a) 12.8 m s^{-1}; (b) 0.27 m s^{-1}; (c) $\xi = 10^{-3}\sin(20.9x - 268t)$ m.
10.22 (a) 20 m s^{-1}; (b) 0.95 Hz; (c) 21 m; (d) $\xi = 0.1\sin(6t - 0.3)$ m, $\xi = 0.1\sin(6t - 0.9)$ m; (e) NA; (f) 0.1 m; (g) NA.
10.23 (a) 1.25×10^{-2} m; (b) 6.25×10^{-3} m, 9.8×10^2 N; (c) 5.06×10^3 m s^{-1}, 3.19×10^3 m s^{-1}.
10.24 (a) 3.5×10^{-3}; (b) 0.35; (c) 3.5×10^3; (d) 1.68×10^{-2} m.
10.25 (a) 0.24 m s^{-1}; (b) 1.25 m s^{-1}; (c) 3.95 m s^{-1}; (d) 6.20 m s^{-1}.

10.26 NA.
10.27 (a) 4.49×10^{-13} W m^{-2}, -3.48 db; 0.881 W m^{-2}, 119 db; (b) 1.43×10^{-11} m, 2.00×10^{-5} m.
10.28 (a) i: 10, ii: 100; (b) i: 3.16, ii: 10.
10.29 (a) Goes up by 6 db; (b) 3.16.
10.30 (a) 3 db; (b) 6 db.
10.31 (a) 57.9; (b) 2.98×10^{-4}.
10.32 (a) 532.1 Hz; (b) 471.6 Hz.
10.33 (a) 918.9 Hz, 0.37 m; (b) 1096.8 Hz, 0.31 m.
10.34 (a) 911.8 Hz, 0.373 m; 1088.2 Hz, 0.312 m; (b) yes, the difference is measurable.

Chapter Eleven

11.1 (a) $\mathscr{E} = (\boldsymbol{u}_y + \boldsymbol{u}_z)(\mathscr{E}_0/\sqrt{2}) \sin\theta$, $\mathscr{B} = (\boldsymbol{u}_y - \boldsymbol{u}_z)(\mathscr{E}_0/\sqrt{2}c) \sin\theta$; (b) $\mathscr{E} = (-\boldsymbol{u}_y + \boldsymbol{u}_z\sqrt{3})(1/2)\,\mathscr{E}_0 \sin\theta$, $\mathscr{B} = (-\boldsymbol{u}_y\sqrt{3} - \boldsymbol{u}_z)(1/2c)\,\mathscr{E}_0 \sin\theta$; (c) $\mathscr{E} = \boldsymbol{u}_y\mathscr{E}_0 \cos\theta + \boldsymbol{u}_z\mathscr{E}_0 \sin\theta$, $\mathscr{B} = -\boldsymbol{u}_y\mathscr{E}_0 \sin\theta + \boldsymbol{u}_z(\mathscr{E}_0/c)\cos\theta$; (d) $\mathscr{E} = \boldsymbol{u}_y(2\mathscr{E}_0/\sqrt{5})\cos\theta + \boldsymbol{u}_z(\mathscr{E}_0/\sqrt{5})\sin\theta$, $\mathscr{B} = -\boldsymbol{u}_y(\mathscr{E}_0/\sqrt{5}c)\sin\theta + \boldsymbol{u}_z(2\mathscr{E}_0/c\sqrt{5})\cos\theta$, where $\theta = kx - \omega t$.
11.2 (a) I: right-hand circular, II: linearly polarized at $-45°$ with XY-plane, III: right-hand elliptically polarized at $-45°$, IV: left-hand elliptically polarized at $+45°$; (b) in each case $\mathscr{B}_y = -\mathscr{E}_z/c$ and $\mathscr{B}_z = \mathscr{E}_y/c$.
11.3 (a) 3 m, linearly polarized, wave travels out $+X$-axis; (b) $\mathscr{B} = \boldsymbol{u}_z(-0.5/c)\cos[2\pi \times 10^8(t - x/c)]$ T; (c) 3.3×10^{-4} W m^{-2}.
11.4 (a) 15 m, right-hand circularly polarized, wave travels out $+X$-axis; (b) $\mathscr{B} = \boldsymbol{u}_y(-0.5/c)\sin[4\pi \times 10^7(t - x/c)] + \boldsymbol{u}_z(0.5/c)\cos[4\pi \times 10^7(t - x/c)]$; (c) 6.6×10^{-4} W m^{-2}.
11.5 (a) i: $0.5\sqrt{6}\mathscr{E}_0$, 35.3°; $0.5\sqrt{6}\,\mathscr{E}_0$, 90°; $0.5\sqrt{6}\mathscr{E}_0$, 215.3°; $0.5\sqrt{2}\mathscr{E}_0$, 270°; $0.5\sqrt{6}\mathscr{E}_0$, 35.3°; ii: $0.5\sqrt{2}\mathscr{E}_0$, 270°; $0.5\sqrt{6}\mathscr{E}_0$, 35.3°; $0.5\sqrt{2}\mathscr{E}_0$, 90°; $0.5\sqrt{6}\mathscr{E}_0$, 215.3°; $0.5\sqrt{2}\mathscr{E}_0$, 270°; (b) in each case $\mathscr{B} = \mathscr{E}/c$ and the angle is 180° greater than given.
11.6 $\mathscr{E} = \boldsymbol{u}_z(\sqrt{24\pi})\cos\theta$ N C^{-1}, $\mathscr{B} = (\boldsymbol{u}_x - \boldsymbol{u}_y)(\sqrt{24\pi}/c)\cos\theta$ T where $\theta = \omega t - k_x x - k_y y = 2\pi \times 10^6(6 \times 10^8 t - \sqrt{2}x - \sqrt{2}y)$.
11.7 $\mathscr{E} = \boldsymbol{u}_y(\sqrt{24\pi})\cos\theta$ N C^{-1}, $\mathscr{B} = \boldsymbol{u}_z(\sqrt{24\pi}/c)\cos\theta$ T where $\theta = 4\pi \times 10^6(3 \times 10^8 t - x)$.
11.8 (a) 3.33×10^{-11} T; (b) 8.85×10^{-16} J m^{-3}; (c) 8.85×10^{-16} N m^{-2}; (d) 1.77×10^{-15} N m^{-2}.
11.9 1050 N C^{-1}, 3.52×10^{-6} T.
11.10 69.28 N C^{-1}, 2.31×10^{-7} T.
11.11 (a) 3.33×10^{-10} T; (b) 1.33×10^{-5} W m^{-2}; (c) 4.4×10^{-14} J m^{-3}; (d) 167 W.
11.12 (a) $\mathscr{E}_z = \mathscr{E}_0[\cos(\omega t - kx) + \cos(\omega t - ky)]$; (b) $\mathscr{B} = -\boldsymbol{u}_y(\mathscr{E}_0/c)\cos(\omega t - kx) + \boldsymbol{u}_x(\mathscr{E}_0/c)\cos(\omega t - ky)$; (c) $\epsilon_0\mathscr{E}_z^2$; (d) X-component $= c^2\epsilon_0\mathscr{E}\mathscr{B}_y$, Y-component $= c^2\epsilon_0\mathscr{E}\mathscr{B}_x$.
11.13 NA.

11.14 NA.
11.15 (a) 795.8 W m^{-2}; (b) 1.55×10^3 N C^{-1}, 5.16×10^{-6} T; (c) 2.65×10^{-6} J m^{-3}, 8.83×10^{-15} kg m^{-2} s^{-1}.
11.16 (a) 5.76×10^{-12} W; (b) 1.74×10^{11} molecules (2.8×10^{-13} moles).
11.17 280 W.
11.18 $(dE/dt) = [q^4 v^2 \mathcal{B}^2/6\pi\epsilon_0 c^3 m^2(1 - v^2 - c^2)^2]$.
11.19 (a) 4.6×10^{-8} W (2.9×10^{11} eV s^{-1}), 4.4×10^{-5} eV rev^{-1}; (b) 1.1×10^{-2} eV s^{-1}, 5.2×10^{-10} eV rev^{-1}; (c) 1.2×10^{-7} eV s^{-1}, 2.4×10^{-13} eV rev^{-1}.

Chapter Twelve

12.1 (a) 1.0243×10^{-10} m; (b) 4.72×10^{-17} J at 44.3° from the incident direction.
12.2 (a) 1.012×10^{-10} m, 58.8°; (b) 2.297×10^{-17} J or 143.4 eV.
12.3 (a) NA; (b) NA.
12.4 (a) −96.9 eV, 2.34×10^{16} Hz, 1.214×10^{-12} m; (b) 96.9 eV, 5.32×10^{-24} kg m s^{-1}, 59.5°.
12.5 (a) 2.418×10^{17} Hz, 1.24×10^{-9} m; (b) 2.418×10^{14} Hz, 1.24×10^{-6} m; (c) 2.418×10^{12} Hz, 1.24×10^{-4} m.
12.6 3.37×10^{-19} J or 2.1 eV.
12.7 (a) 2.66×10^{15} Hz; (b) 1.13×10^{-7} or 1127.9 A.
12.8 (a) Electron: 5.40×10^{-23} kg m s^{-1}, proton: -4.86×10^{-23} kg m s^{-1}; (b) electron: 1.60×10^{-19} J = 9.98 keV, proton: 7.07×10^{-19} J = 4.41 eV.
12.9 (a) 273.6 MeV; (b) 40 MeV.
12.10 (a) 4.01×10^{-22} kg m s^{-1} (0.75 MeV/c), 0.908 MeV; (b) 0.998 MeV; (c) yes.
12.11 (a) NA; (b) 4.6×10^{14} Hz; (c) 1.89 eV; (d) 6.59×10^{-34} J s.
12.12 (a) 1.45 V; (b) 1.45 eV (= 2.32×10^{-19} J), 7.13×10^5 m s^{-1}.
12.13 (a) 6.04×10^9 electrons s^{-1} m^{-2}; (b) 3×10^{-9} J s^{-1} m^{-2} or 1.21×10^{10} eV s^{-1} m^{-2}.
12.14 (a) 2.89×10^{15} Hz; (b) $n^2 = 1 + 4.85 \times 10^{25}(\Sigma f_i/[(2.89 \times 10^{15})^2 - \nu^2])$; (c) 1.000003, 0.99992.
12.15 NA.
12.16 NA.
12.17 NA.
12.18 NA.

Chapter Thirteen

13.1 (a) NA; (b) NA; (c) NA.
13.2 (a) 2.87×10^{-2} m; (b) NA.
13.3 (a) 26.2°, 22.5°; (b) 2.20×10^{-2} m.
13.4 16.6°.
13.5 1.11, 0.11.
13.6 (a) NA; (b) there is no loss of energy.

13.7 (a) 8.52×10^{-7} m; (b) 2.70×10^{-9} m; (c) 3.81×10^{-11} m.
13.8 (a) $\Delta\theta = [2.4 \times 10^{-4}\ \rho_0(pT_0/p_0T) \tan\theta]$ radians where ρ_0, p_0 and T_0 are atmospheric density, pressure and temperature at STP; (b) 58.7″ of arc.

Chapter Fourteen

14.1 (a) 9.20×10^{-2}, 0.728; (b) −0.3033, 0.697.
14.2 (a) NA; (b) NA.
14.3 (a) −0.2, 0.8; (b) 0.2, 1.2; (c) in the first case there is a phase change of π for the reflected wave.
14.3 NA.
14.5 56.3°, 33.7°.
14.6 35.3°.
14.7 (a) 36.9°; (b) normal to the plane of incidence.
14.8 (a) $\frac{1}{2}c\epsilon_0\mathscr{E}_0^2$; (b) $\mathscr{E}_0/n$; (c) $\frac{1}{2}v\epsilon\mathscr{E}_0^2$; (d) answers (a) and (b) are identical, as they should be since no energy is reflected.
14.9 (a) NA; (b) NA.
14.10 NA.
14.11 (a) 0.848, −0.128; (b) 0.845, −0.155; (c) 0.502, −0.333; (d) 0.441, −0.559; only parallel components, no parallel components in (a) and (c) and the opposite case for (b) and (d).
14.12 (a) 1.177, −0.115; (b) 1.168, 0.168; (c) 1.399, 0.052; (d) 1.325, 0.325.
14.13 Reflected beam is elliptically polarized with opposite-handedness from the incident beam; the normal component = 0.311, parallel component = 0.097; refracted beam is elliptically polarized in the same sense as the incident beam; normal component = 0.689, parallel component = 0.722.
14.14 (a) 0.75; (b) 0.50; (c) 0.25; (d) 0; (e) 0.25; (f) 0.50; (g) 0.75; (h) 1; (i) NA.
14.15 (a) 3.969×10^{-7} m, 3.558×10^{-7} m; (b) 5.085×10^{14} Hz.
14.16 (a) 1.620; (b) 2.35.
14.17 (a) Odd integer × 8.726×10^{-7} m; (b) integer × 1.745×10^{-6} m; (c) integer × 3.49×10^{-6} m.
14.18 (2) linearly polarized at the 45° angle; (3) elliptically polarized, axis orientation dependent on thickness of plate; (4) linearly polarized at the 60° angle.
14.19 (a) 45° with respect to the Y-axis, counterclockwise; (b) perpendicular to the answer in (a); (c) 5.83:1.
14.20 (a) 6.05×10^{-7} m, 6.30×10^{-7} m, 6.56×10^{-7} m, 6.84×10^{-7} m; (b) 6.17×10^{-7} m, 6.42×10^{-7} m, 6.70×10^{-7} m, 7.00×10^{-7} m; (c) 6.05×10^{-7} m, 6.56×10^{-7} m.
14.21 299°.
14.22 1.19×10^{-3} kg.

Chapter Fifteen

15.1 (a) 0.78 m, 0.56; (b) 1.0 m, 1.0; (c) 1.33 m, 1.67; (d) ∞, ∞−; (e) −0.75 m, 2.5 (upright); (f) 0.27, 0.45 (upright).

15.2 (a) −0.27 m; (b) 0.45 (upright); (c) 0.75 m, 0.25 (upright); (d) −1.33 m, 1.66 (inverted).
15.3 (a) 0.48 m; (b) 1.92 m; (c) −1.20 m; (d) 0.80 m; (e) 2.40 m; (f) 0.30 m; (g) −0.60 m [the minus sign indicates a convex mirror].
15.4 (a) 1.07 m; (b) 3.2 m; (c) 0.53 m.
15.5 (a) −1.07 m; (b) −3.2 m; (c) −0.53 m.
15.6 0.08 m, 2.1.
15.7 0.07 m.
15.8 0.375 m or 0.40 m.
15.9 (a) NA; (b) NA.
15.10 (a) $f_o = -1.2$ m, $f_i = -1.8$ m; (b) i: −1.2 m, 0.33 (upright); ii: −1.03 m, 0.43 (upright); iii: −0.6 m, 0.67 (upright).
15.11 (a) $f_o = 1.2$ m, $f_i = 1.8$ m; (b) i: 3.6 m, 1.0 (inverted); ii: 7.2 m, 3.0 (inverted); iii: −1.8 m, 2.0 (upright).
15.12 (a) 1.5 m beyond the first surface, real; (b) 4.0 mm; (c) 0.25 m outside the rod, real; (d) 1.5 mm.
15.13 (a) $f_o = 0.05$ m (in the figure 5 cm to the left of the 10 cm surface), $f_i = -0.2$ m [i.e., at the center of curvature of the 20 cm surface]. Note that for light passing right to left through this system, the focal points are at the same positions; however, f_i is now at f_o and vice versa; (b) NA.
15.14 (a) −0.438 m (i.e., about 3.8 cm outside the rod on the hemisphere side); (b) 1.71 (upright).
15.15 (a) 0.357 m; (b) real for the second surface; (c) upright; (d) 1.71×.
15.16 0.367 m.
15.17 (a) NA (there are four); (b) NA (2 are converging); (c) ±0.133 m, ±0.40 m.
15.18 (a) 0.24 m; i: (b) 0.343 m, (c) 0.43; ii: (b) 0.48 m, (c) 1; iii: (b) 0.60 m, (c) 1.5; iv: (b) ∞, (c) NA; v: (b) −1.2 m, (c) 6; (d) 0.10 m, 0.5.
15.19 (a) −0.24 m; (b) −0.11 m; (c) 0.545; (d) −0.60 m, 1.5; (e) 1.20 m, 6.
15.20 0.0545 m.
15.21 (a) 0.12 m and 0.06 m from the screen; (b) 2 and 0.5.
15.22 (a) 1.07 m; (b) 0.67 m; (c) 0.80 m; (d) 0.27 m; (e) −0.10 m. All are real except (d).
15.23 (a) 0.80 m; (b) 2.4 m; (c) 0.40 m.
15.24 (a) 0.48 m; (b) 1.92 m; (c) −1.2 m; (d) 0.80 m; (e) 2.4 m; (f) 0.3 m; (g) −0.60 m.
15.25 NA.
15.26 4.5×10^{-3} m.
15.27 (a) $F_o = 0.17$ m, $F_i = 0.09$ m; (b) $F_o = 0.08$ m, $F_i = -0.30$ m; (c) $F_o = F_i = \infty$; (d) $F_o = 0.60$ m, $F_i = 1.8$ m. All distances are from the 0.6 m lens.
15.28 (a) $F_o = -1.2$ m, $F_i = -3.0$ m; (b) $F_o = 0.15$ m, $F_i = 2.4$ m; (c) $F_o = -0.15$ m, $F_i = 1.2$ m; (d) $F_o = -2.0$ m, $F_i = 1.0$ m. All the distances are from the 0.6 m lens.

15.29 $F_o = 1.67 \times 10^{-2}$ m, $F_i = 4.17 \times 10^{-2}$ m from the first lens.
15.30 (a) 360; (b) 2.8×10^{-7} m.
15.31 0.37 rad or 21°.
15.32 (a) 38.5°; (b) 37.2°; (c) 48.6°.
15.33 (a) 1.52; (b) 40°.
15.34 by a least-squares fit: $A_0 = 1.51322$, $B_0 = 2.155 \times 10^{-15}$.
15.35 0.304° (0.739° if the n values in Table 15-3 are used).
15.36 (a) $F_D = 0.266$ m; (b) 3.33×10^{-3} m.

Chapter Sixteen

16.1 (a) 6.5×10^{-4} m; (b) 1.6×10^{-3} m, 3.3×10^{-3} m.
16.2 4.2×10^{-3} m.
16.3 (a) i: no phase change, ii: phase change of π radians; (b) reflection from a more dense medium "always" does this.
16.4 4.5×10^{-5} m.
16.5 Rings of bright and dark will appear.
16.6 (a) On the line *between* the sources at the midpoint and 0.25 m from either and *everywhere* outside; (b) none; (c) for a coordinate system with origin midway between the two sources and the X-axis containing the sources, points of minimum intensity will be hyperbolas given by $r_1 - r_2 = (2n + 1)/4$, where $r_1^2 = (x + \frac{3}{8})^2 + y^2$ and $r_2^2 = (x - \frac{3}{8})^2 + y^2$; (d) the intensity is never zero because the two sources are always at different distances from the point in question.
16.7 (a) NA; (b) NA.
16.8 4.07×10^{-5} rad.
16.9 Main maxima at angles given by $\sin^{-1}(n\lambda/a)$ where n is any integer, including zero, with intensities of $9I_0$; zeros at angles given by $\sin^{-1}(n'\lambda/3a)$, where n' is any integer, excluding 0, 3, 6, 9 ...; a single lower maximum of intensity I_0 at angles given by $\sin^{-1}([2n'' + 1]\lambda/2a)$ where n'' is any integer.
16.10 $1.18\ I_0$.
16.11 (a) λ; (b) maxima at 0° ($16I_0$), at 21.5° ($1.18I_0$), at 39.4° ($1.18I_0$), at 90° ($16I_0$) and similarly in each quadrant; (c) NA.
16.12 (a) 0.11° or 1.9×10^{-3} rad; (b) 1.72°.
16.13 2.95×10^{-4} rad.
16.14 1.6×10^{-6} m.
16.15 (a) 5.6 wavelengths; (b) almost 35 radians; (c) 6 wavelengths or about 38 radians.
16.16 1.18×10^{-5} m.
16.17 (a) $\rho_r = 2.19 \times 10^{-3}\sqrt{N}$ m, where N is any positive integer; (b) 83 rings.
16.18 1.4×10^{-3} m.
16.19 (a) 298.7 Hz, 597 Hz, 896 Hz; (b) 2 m, 1 m, 0.67 m; (c) NA; (d) $\xi = \xi_0 \sin(k_n x)\cos(\omega_n t)$, where $k_n = 2\pi/\lambda_n$ and $\omega_n = 2\pi\nu_n$.
16.20 (a) i: increases by $\sqrt{2}$, ii: decreases by $1/\sqrt{2}$, iii: decreases by $\frac{1}{2}$,

iv: decreases by $\frac{1}{2}$; (b) i: decreases by $1/\sqrt{2}$, ii: increases by $\sqrt{2}$, iii: increases by 2, iv: increases by 2.

16.21 (a) 289.5 Hz, 578.9 Hz; (b) 144.7 Hz, 434.2 Hz; (c) NA; (d) NA.

16.22 0.17%.

16.23 24.4 Hz.

16.24 (a) 0.339 m, 1.02 m, 1.70 m; (b) NA.

16.25 (a) NA; (b) NA; (c) NA.

16.26 (a) NA; (b) ν_0: doubly degenerate on 1, 0 & 0; ν_1: single 1 & 1; ν_2: double on 0 & 2; ν_3: double on 1 & 2; ν_4: single on 2 & 2; ν_5: double on 0 & 3; ν_6: double on 1 & 3; ν_7: double on 2 & 3.

16.27 (a) $\nu = (v/2a)\sqrt{n_1^2 + n_2^2 + n_3^2}$; ν_0: triply degenerate on 1, 0, & 0; $\nu_1 = \sqrt{2}\nu_0$: triple on 1, 1, & 0; $\nu_2 = \sqrt{3}\nu_0$: single on 1, 1, 1; $\nu_3 = 2\nu_0$: triple on 2, 0, 0; $\nu_4 = \sqrt{5}\nu_0$: triple on 2, 1, 0; etc.

16.28 2.25×10^{20} modes/volume.

Chapter Seventeen

17.1 5.6×10^{-4} m.

17.2 6.75×10^{-5} m.

17.3 600 m.

17.4 5×10^{-7} m, 4×10^{-7} m.

17.5 (a) A diffraction pattern of rectangles similar to Fig. 17-11. (b) 2.4×10^{-3} m $\times$ 1.2×10^{-3} m.

17.6 (a) 2.05×10^{-6} m; (b) 6.83×10^{-6} m.

17.7 9.7 km (197 km).

17.8 4.8×10^{-5} m.

17.9 (a) 3; (b) NA; (c) NA.

17.10 8.9 m.

17.11 $a = 5 \times 10^{-5}$ m, $b = 1.25 \times 10^{-5}$ m.

17.12 (a) 17.5°; (b) 36.9°; (c) 64.2°.

17.13 12.5°.

17.14 (a) 5.70×10^{-7} m; (b) 43.2°.

17.15 5.00×10^{-7} m.

17.16 First order: 13.9° to 24.8°; second order: 28.7° to 57.1°.

17.17 (a) 16,000; (b) yes, they are 0.84′ apart (see Example 17.6); (c) 13.7°; (d) 4.1×10^5 m^{-1}.

17.18 NA.

17.19 (a) 7.75×10^{-4} m; (b) 1.55×10^{-3} m.

17.20 7.07×10^{-4} m.

17.21 (a) 6.67×10^{-5} m; (b) 5.0×10^{-5} m; (c) bright for 3 zones, dark for 4 zones.

17.22 (a) Bright; (b) 0.25 m *toward* the aperture.

17.23 (a) Minimum intensity: $8/n$ m, where $n = 2, 4, 6, \ldots$; maximum intensity: $8/n$ m, where $n = 1, 3, 5, \ldots$; (b) 4 m.

17.24 (a) λ_1 only at $20/n$ m, where $n = 1, 2, 3, \ldots$; (b) λ_2 only at $40/3n'$ m, where $n' = 1, 2, 3, \ldots$ *except* whenever $n' = 2n/3$ (e.g.: $n = 6$, $n' = 4$, or 3.33 m).
17.25 (a) 9.79×10^{-11} m; (b) 20.3°.
17.26 (a) 6.98×10^{-10} m; (b) 12.4°, 25.4°.
17.27 1.25°.
17.28 7.96×10^{-3} m and 1.59×10^{-2} m.

Chapter Eighteen

18.1 (a) 1.23×10^{-9} m; (b) 1.23×10^{-10} m; (c) 3.89×10^{-11} m; (d) if the *total energy* of the electron is 1 MeV, the problem is done relativistically and the solution is 8.7×10^{-13} m.
18.2 100 eV at 16.6° and 1 keV at 5.2°.
18.3 32.7 eV.
18.4 7.99×10^{-2} m and 0.161 m.
18.5 3.6°.
18.6 11.5°.
18.7 (a) 1.24 GeV; (b) 1.24 GeV; (c) 616 MeV.
18.8 (a) 3.9 m; (b) complete accuracy.
18.9 (a) 7.3×10^3 m; (b) complete accuracy.
18.10 (a) 6.6×10^{-23} kg m s^{-1}; (b) 7.7 keV.
18.11 (a) 6.6×10^{-20} kg m s^{-1}; (b) 8.1 MeV.
18.12 (a) $[-\hbar^2/2m]d^2\psi/dx^2 - (\mathscr{E}x)\psi = E\psi$; (b) NA; (c) no.
18.13 10^9 Hz.
18.14 (a) Levels get further apart; (b) levels become closer together.
18.15 (a) 9.36 eV; (b) the electron's energy would be raised enough to necessitate a relativistic treatment, giving 0.4 MeV.
18.16 (a) 5.36×10^{-10} J = 3760 MeV; (b) gravitational energy = 10^{-34} eV while coulomb energy is 0.14 MeV; (c) no electron can exist within a nucleus.
18.17 204 MeV.
18.18 (a) NA; (b) NA.
18.19 NA.
18.20 $E = [\hbar^2\pi^2/2m][n_1^2/a^2 + n_2^2/b^2 + n_3^2/c^2]$,
$\psi = C \sin(n_1\pi x/a) \sin(n_2\pi y/b) \sin(n_3\pi z/c)$.
18.21 (a) 1.325×10^{-31} J, 2.65×10^{-31} J; (b) 3.98×10^{-31} J, 2.65×10^{-31} J; (c) 2.01×10^{-22} J, 4.02×10^{-22} J.
18.22 (a) NA; (b) NA; (c) yes.