1958

Physics [complete work]

Henry Semat
City College of New York

Robert Katz
University of Nebraska-Lincoln, rkatz2@unl.edu

Physics

Books by Henry Semat

Fundamentals of Physics THIRD EDITION

Introduction to Atomic and Nuclear Physics THIRD EDITION

Physics in the Modern World

With Robert Katz

Physics

Physics

HENRY SEMAT, Ph.D.

Professor of Physics
The City College of New York

ROBERT KATZ, Ph.D.

Professor of Physics
Kansas State College of
Agriculture and Applied Science

RINEHART & COMPANY, INC., NEW YORK

Printed in the United States of America
Library of Congress Catalog Card Number: 58-5761

Preface

This book is intended for students of science and engineering; it aims to develop both an understanding of the important concepts of physics and some analytical skill in the solutions of problems. The mathematical level of the book is such that it may be used by students who are taking a course in calculus concurrently.

The notations and methods of the calculus are introduced early in the text, beginning with the concept of a derivative in the discussion of motion, and are then extended to more complex problems as the student progresses both in physics and in mathematics. Vector algebra is, of course, also used. The vector notation is introduced at the beginning of the text in treating displacements; it is then extended to include the use of the dot product and cross product of two vectors, and the resolution of a vector into components with the aid of unit vectors. These vector methods are used extensively in the sections on Mechanics and Electricity.

The method of exposition and the division of the subject matter into six parts—Mechanics, Heat, Wave Motion and Sound, Electricity and Magnetism, Optics, and Atomics and Nucleonics—follow closely those of the senior author's *Fundamentals of Physics,* now in its third edition. However, the treatment of much of the material is entirely new, as are over two hundred of the figures. The problems at the end of each chapter are graded in difficulty, and many illustrative examples are provided in the text both to clarify concepts and to guide the student in the analytical approach to the solutions of problems. Included among these problems are some involving selected derivations and some requiring the use of calculus. Answers to odd-numbered problems are given in the Appendix, and a booklet containing all the answers is available to the instructor and may be distributed to his students if he so desires.

Three systems of units are developed and used in the sections on Mechanics, Heat, and Sound; these are the British engineering system, and the cgs and mks metric systems. In the section on Electricity and Magnetism, the rationalized mks system is the primary one, but the unrationalized Gaussian system is also developed because of its wide use in all of physics and much of engineering. The Gaussian system is introduced in the early chapters of Electricity; in subsequent

chapters it is used only when new concepts are introduced or when important equations are developed. The forms taken by each important equation in both the mks and the Gaussian systems are presented in a table at the end of the appropriate chapter, while the necessary conversion factors for the units and constants of each system appear in another table. Problems at the end of each chapter are given in both sets of units.

In developing the subject of Electricity and Magnetism we were guided by the recommendations of the Coulomb Law Committee of the American Association of Physics Teachers and we followed these recommendations fairly closely. We take this opportunity of expressing our indebtedness and appreciation to this Committee.

We have used modern concepts and examples throughout the book, and in addition have devoted an entire separate section to Atomics and Nucleonics. The reason for this section is that the traditional two- or three-semester course in physics is actually a terminal course for a great many students; this course will be their only opportunity to learn about atomic and nuclear physics from a physicist's point of view. Wherever a separate course in atomic and nuclear physics is part of the curriculum, this section may be omitted from the general course.

We wish to thank Professor Karl D. Larsen, Head of the Department of Physics at Lafayette College, Professor I. Wallerstein of Purdue University, and Professor Robert A. Becker of the University of Illinois, each of whom read the manuscript and made many valuable criticisms and helpful suggestions.

We are also indebted to several colleagues at our respective institutions who read different parts of the manuscript and made valuable criticisms and helpful suggestions. At Kansas State College Professor R. M. Kerchner, Head of the Electrical Engineering Department, read the section on Electricity and Magnetism; Professor L. D. Ellsworth of the Physics Department read the sections on Electricity and Magnetism and Optics, and Professor Basil Curnutte, Jr., of the Physics Department read the section on Optics. At the City College of New York, Professor M. W. Zemansky, Chairman of the Physics Department, read the section on Heat, Professor Fred C. Rose read the section on Electricity and Magnetism, and Professor Robert I. Wolff read the section on Optics.

HENRY SEMAT
City College of New York

ROBERT KATZ
Kansas State College

January, 1958

Contents

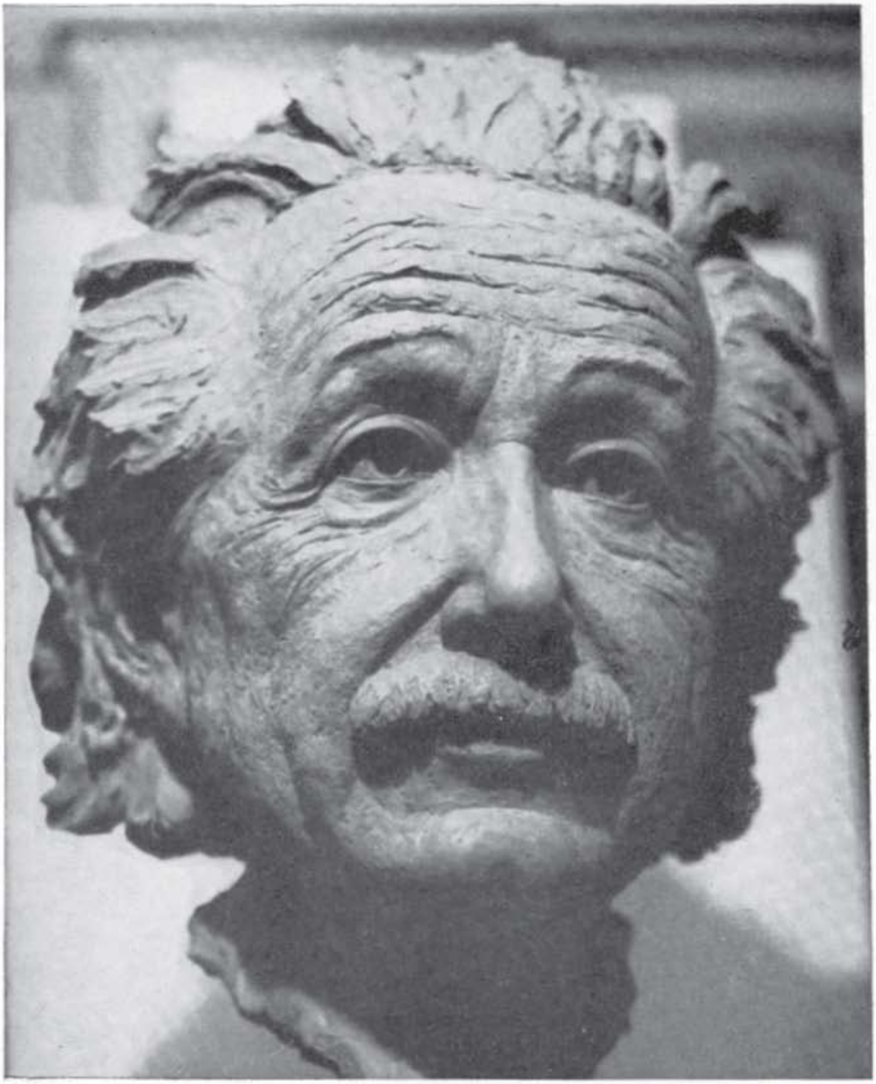
Part Four ELECTRICITY AND MAGNETISM

Part Five LIGHT

Part Six ATOMICS AND NUCLEONICS

APPENDICES

Physics



Sculpture of Albert Einstein by Emil Seletz, M.D., Los Angeles.

Part One

MECHANICS

1

Fundamental Quantities

1-1 The Scope of Physics

Physics is a fundamental science dealing with matter and energy. By convention, the subject matter of physics has been divided into such topics as mechanics, heat, sound, light, and electricity. In addition to these general classifications, present-day physics includes atomic physics, nuclear physics, solid-state physics, chemical physics, biophysics, and many other subdivisions. It is impossible to include all aspects of physics in a single definition or paragraph, and to distinguish physics clearly from its nearest neighbors, the other physical sciences—astronomy, chemistry, and geology.

Like other scientists, the physicist studies nature, and, although the scientist is himself part of the world, he attempts to describe nature as it exists without his interference. Inspired by the conviction that nature is orderly and "rational," the physicist has sought to find that order and to express it as elegantly and as concisely as possible. Physics has attempted to study the least complex aspects of nature, or, if you will, the most fundamental, and has found that quantitative descriptions have been most fruitful. Physics has thus become a quantitative science involving precise measurements of certain quantities and the formulation of mathematical relationships among them.

Like other sciences, physics is an ahistorical subject. This does not mean that there is no history of physics but rather that, at a particular moment in time, the history of the development of physical ideas is not pertinent to the question of the accuracy of the description of nature. Because we can devote but little space to the history of physical ideas, the student will find that the time spent in consulting an encyclopedia on some of the topics studied in this text, and in reading biographies of physicists, will enrich his knowledge of the subject and may even stimulate his own inventiveness.

1-2 The Relationship of Physics to Engineering

Engineering has often been called applied physics, as though the physicist discovered all fundamental truth and the engineer merely applied it; as though the entire structure of engineering rested on the shoulders of physics. This is true only in the very limited sense that fundamental knowledge is called physics, and that the application of that knowledge is called engineering. In actual fact, there is so great an interplay between physics and engineering that activity in one field is stimulated by the needs and the successes of the other. The entire radio industry owes its birth to the researches of physicists such as Maxwell, Hertz, and J. J. Thomson during the latter half of the last century. Much of today's experimental physics would not be possible without the developments which have taken place in the technology of radio through the efforts of such men as Edison, DeForest, Fleming, Marconi, and Armstrong.

1-3 The Fundamental Concepts of Mechanics

Mechanics deals with the behavior of various types of bodies such as *particles*, *rigid bodies*, *liquids*, and *gases*, when subjected to the action of *forces*. Under special conditions the bodies may be in *equilibrium* under the action of these forces; under other conditions the motions of these bodies will be changed or *accelerated*. The quantitative concepts used in mechanics can be classified into two groups, one known as the *fundamental concepts* consisting of three quantities, *length*, *time*, and *mass*, which form the bases of mechanics, and a second group known as the *derived concepts* consisting of the other quantities used in mechanics.

The difference between these two sets of quantities lies in the fact that each one of the derived concepts can be defined accurately in terms of the fundamental concepts. But the fundamental concepts of length, time, and mass cannot be defined in any such manner; they require special consideration. The method of treating each of these concepts is to set forth a series of rules for its measurement or to outline a series of experiments for determining its magnitude. Two steps are usually involved in each case: one, that of setting up a *standard* and, two, that of setting up rules for producing multiples or subdivisions of the standard. These steps will be discussed for each of the fundamental concepts.

The mechanics of particles and rigid bodies is subdivided under the more formal titles of *statics*, *kinematics*, and *dynamics*. *Statics is the study of bodies at rest and the conditions under which they remain at rest. Kinematics is the study of the motion of bodies without regard to the cause of that motion.* It is the language in whose terms motion is described. *Dynamics is the study of motion in relation to the forces which cause that motion.*

1-4 Length. Standards and Units

Most of us are familiar, in a general way, with the operations involved in measuring the length of an object or of measuring the distance between two points. Such a measurement involves the use of some measuring rod or tape whose length is known in terms of a standard of length. Only in comparatively modern times have standardized measuring rods been readily available. Such units as the length of a man's foot, the distance between his outstretched hands (the fathom), and the distance between his footprints (the pace) have always been available and are still in common use, but these are inadequate for recording the measurements of the physicist or the engineer. Today, *the legal standard of length in the United States is the meter*, defined as the distance between two lines marked on a special platinum-iridium bar known as the *standard meter* when that bar is at 0°C. This standard meter is deposited at the International Bureau of Weights and Measures at Sèvres, France. Accurate copies of this standard meter have been distributed to various national physical laboratories such as the National Bureau of Standards in Washington, D.C. As a result of subsequent developments in physics, it has been possible to measure the meter in terms of the wavelength of a particular kind of light. If this wavelength measurement, or a similar one, is adopted as a standard by international agreement, we would then have an indestructible standard of length.

In English-speaking countries other units of length find conventional usage. In the United States the yard is defined legally as 3,600/3,937 of the standard meter.

TABLE 1-1 PREFIXES USED IN THE METRIC SYSTEM

Latin Base			Greek Base		
deci	1/10	10^{-1}	deka	10	10^{1}
centi (c)	1/100	10^{-2}	hecto	100	10^{2}
milli (m)	1/1,000	10^{-3}	kilo (k)	1,000	10^{3}
micro (μ)	1/1,000,000	10^{-6}	mega (M)	1,000,000	10^{6}

Symbols for prefixes in common use are shown in parentheses.

While multiples and submultiples of the yard are defined in a rather complex way in the English system of units, in the metric system multiples and submultiples of the meter are based upon the decimal system. Latin prefixes are attached to the name of the standard to indicate smaller units, while Greek prefixes indicate larger units. The prefixes in common use are indicated in Table 1-1.

Some of the more common multiples and submultiples of the yard and meter are given in the appropriate columns of Table 1-2.

TABLE 1-2 UNITS OF LENGTH

English	Metric
3 feet = 1 yard	100 centimeters = 1 meter
36 inches = 1 yard	10 millimeters = 1 centimeter
5,280 feet = 1 mile	1,000 microns = 1 millimeter
1,000 mils = 1 inch	1,000 meters = 1 kilometer

It is often necessary to convert the value of a quantity expressed in one set of units to its value expressed in a different set of units. For this purpose the appropriate conversion factor must be used. To obtain this, it is necessary to refer to the original definitions of the standards of the two systems and then compute the conversion factor. Once done, these factors may be listed in a table for convenience and may be used with confidence. Table 1-3 lists some conversion factors for units of length. Others may be calculated when necessary.

TABLE 1-3 CONVERSION FACTORS

1 kilometer	= 0.6214 mile
1 meter	= 39.37 inches
1 foot	= 30.48 centimeters
1 inch	= 2.540 centimeters

Illustrative Example. Convert miles per hour to feet per second. The conversion from one set of units to another is executed by a series of multiplications by unity. Since multiplication by the pure number, 1, cannot affect the truth of an equation, the resulting calculation, correctly performed, must also be a true equality. Note that the equation 5,280 ft = 1 mi can be written as

$$\frac{5{,}280 \text{ ft}}{1 \text{ mi}} = 1;$$

similarly,

$$\frac{1 \text{ hr}}{3{,}600 \text{ sec}} = 1.$$

We can thus write

$$1\,\frac{\text{mi}}{\text{hr}} = 1\,\frac{\text{mi}}{\text{hr}} \times \frac{5{,}280 \text{ ft}}{1 \text{ mi}} \times \frac{1 \text{ hr}}{3{,}600 \text{ sec}}$$

$$= \frac{5{,}280}{3{,}600}\,\frac{\text{ft}}{\text{sec}},$$

or

$$1\,\frac{\text{mi}}{\text{hr}} = 1.467\,\frac{\text{ft}}{\text{sec}}.$$

All conversions from one set of units to another may be accomplished in this way.

1-5 Angular Measure

While many systems of linear measure are in general use, only two systems of angular measurement are ordinarily encountered. In the degree-minute-second system, a circle is first divided into 360 equal parts called degrees. Each degree is divided into 60 equal parts called minutes, and each minute is divided into 60 equal parts called seconds. This method of subdivision comes from the Babylonians, who used a number system based on 60; the terminology derives from the Latin for "first small part" and "second small part" (small = minute). The other system of angular measurement is based on the radian as the unit of angle, where a radian is defined as the angle subtended at the center of a circle by an arc whose length is equal to the radius. To determine an angle in radian measure, we draw a circle of arbitrary size centered at the intersection of the sides of the angle, as shown in Figure 1-1. If r is the radius of the circle and s is the length of the inscribed arc, then the angle ϕ (Greek letter phi) is determined by dividing the length of the arc s by the length of the radius r; thus

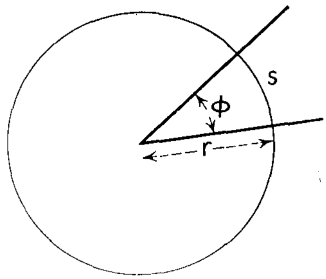


Fig. 1-1

$$\phi = \frac{s}{r}\cdot \tag{1-1}$$

Note that neither the degree nor the radian has any dimensions. The radian is a pure number. It is simply a length divided by length. The degree, minute, or second is likewise a pure number, simply indicating the fraction of a full circle represented. The designation "radian" or "degree" is merely a description of the manner in which the measurement was made.

To convert from degrees to radians, we observe that a complete circle may be represented as 360° or as 2π radians, so that

$$1 \text{ radian} = \frac{360°}{2\pi} = 57.3°.$$

Some interesting examples of the measurement of distances through the measurement of angles are the determination of the radius of the earth, the diameter of the earth's orbit, and the distance of some of the nearer stars. For example, to measure the radius of the earth, simultaneous sightings of a distant star may be made from two points A and B on the earth's surface a known distance s apart (see Figure 1-2). Because of the great

distance of the star from the earth, the lines of sight of the star from the two locations may be assumed to be parallel. We assume the plumb bob to point to the center of the earth (although this is not strictly true). The

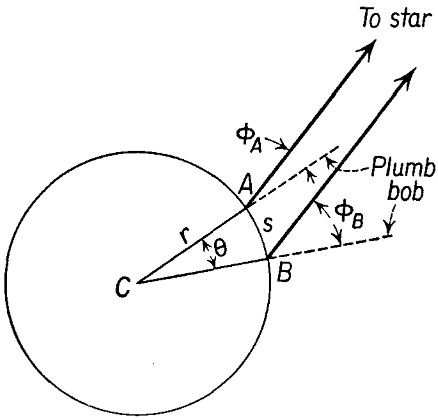


Fig. 1-2 Method of determining the radius of the earth.

angle between the line of sight and the plumb bob is measured at each of these two points; let us call these angles ϕ_A and ϕ_B, respectively. From geometry, the angle θ (theta) subtended at the center of the earth is given by the difference between the angles ϕ_B and ϕ_A. Thus

$$\theta = \phi_B - \phi_A.$$

Now

$$\theta = \frac{s}{r},$$

or

$$r = \frac{s}{\theta},$$

from which

$$r = \frac{s}{\phi_B - \phi_A}.$$

Since the quantities on the right-hand side of the equation are measurable, the radius of the earth can be determined. As the result of a great many such measurements, the earth is known to be an oblate spheroid, a slightly squashed-down sphere, of a radius of approximately 4,000 miles.

1-6 The Concept of Time

Just as man has a built-in concept of length, so he has a built-in concept of time. The natural cycles of the day, the lunar month, and the year have always provided man with measures of long time intervals, while the pulse has provided something of a measure of smaller time intervals. It is said that a musician's sense of tempo is based upon his pulse beat, and that

variations in the tempo of performance of a symphony orchestra have been observed to accompany changes in the conductor's pulse.

The measure of time is still based upon the period of the earth's rotation on its axis and on the period of its revolution about the sun. For this reason the fundamental measurement of time has been the domain of the astronomer. Because of the very practical importance of accurate knowledge of time in ship navigation in the determination of longitude, the accurate measurement of time in the United States is in the hands of the U.S. Naval Observatory. *The standard time interval used in scientific work is the mean solar day.* This standard makes use of the apparent motion of the sun with respect to an observer on the earth; this apparent motion is due to the rotation of the earth on its axis. The instant at which the sun passes the meridian (directly overhead) at the observer's place is called *local noon.* The solar day is the time between two successive noons. Careful measurements have shown that solar days differ slightly at different times of the year, owing to the earth's position in its orbit and the speed and direction of its motion. The mean solar day is an average value taken over the year.

The *unit* of time universally used is the *second;* this is $1/(24 \times 60 \times 60) = 1/86{,}400$ of the mean solar day. The instrument for measuring time is the clock. The astronomical clock makes use of the periodic motion of a pendulum or of the periodic motion of a balance wheel connected to a fine spring. The clock is used to subdivide the day into submultiples.

Basic principles of physics such as the principles of the conservation of energy and of angular momentum, which will be studied in detail later in this text, indicate that the period of rotation of a rigid earth should be constant. The earth loses about a second a year owing to mechanical energy losses in the motion of the tides, and so on. The shape of the earth is continually undergoing slight changes as a result of volcanic action, earthquakes, and the action of the tides. Consequently, although the average second is well defined, two successive seconds of time may not correspond to the same change in the earth's rotational position. Physicists have sought to establish a means of measuring time which does not depend on the motion of the earth and have only recently succeeded in building a clock which bases its measurement of time on the internal vibrations of the ammonia molecule. This clock is said to be accurate to 1 part in 24 billion, or to about 1 sec in 300 years and can, perhaps, be used to determine the irregularities in the motion of the earth.

1-7 The Concept of Mass

Most of us have some idea of the meaning of the term *mass of a body,* but unfortunately this is generally confused with the term *weight.* Our muscles provide us with an intuitive conception of weight, and it is often difficult

for us to make the required distinction between weight and mass. The concept of mass is related to the "amount" of matter present in a particular object, while weight refers to the force of the earth's attraction on that object. While the weight may vary from place to place on the earth, and certainly will vary if the object is removed to the moon, the mass remains constant, except at speeds comparable to the speed of light, and is independent of position in the universe.

A real understanding of the concept of mass can come only after a study of dynamics; for the present we shall confine our discussion to a statement of the standard of mass, the method of measuring mass, and the units commonly used.

The standard of mass is a *kilogram* (abbreviated kg). This is the mass of a certain piece of platinum kept at the International Bureau of Weights and Measures at Sèvres, France. The instrument used in comparing masses is the equal-arm balance. The two bodies whose masses are to be compared are placed in the pans of this instrument and, if a balance is obtained, are said to have equal masses. Two accurate copies of the standard kilogram are deposited at the National Bureau of Standards.

The kilogram is not only the standard of mass, but it is widely used as the *unit* of mass in science and engineering. The system of units based upon the *meter* as a unit of length, the *kilogram* as the unit of mass, and the *second* as the unit of time is called the *mks system.* A second metric system of units is based upon the *centimeter* (cm) as the unit of length, the *gram* (gm) as the unit of mass, and the *second* as the unit of time. This system is known as the *cgs system.*

Other units of mass are commonly used. Thus the *pound mass* (lb mass) is legally defined as $\frac{1}{2.20462}$ kg. The *British gravitational* system of units utilizes a unit of mass called the *slug,* which cannot be properly defined until we study dynamics. In this system of units the unit of length is the *foot* (ft), and the unit of time is the *second.*

1-8 Force

The concept of force is best derived formally through the definitions of length, mass, and time, and the application of Newton's laws of motion. For most purposes the common spring balance is a convenient device for the measurement of force. Although a precise definition of force again must await the study of dynamics, the units of force in the several systems may be named now. The unit of force in the mks system is called the *newton* (nt). The unit of force in the cgs system is called the *dyne.* The unit of force in the British gravitational system is the weight of a pound

mass at a place where the acceleration of gravity has the "standard" value of 32.17398 ft/sec^2; we shall call it a *pound force* (lb f).

1-9 Position

In many physical problems it is necessary to describe the position of an object, and this is most conveniently done by coordinate systems such as those used in analytic geometry. A coordinate system has an origin and a net of fixed reference lines. The location of a point on a two-dimensional plane surface is often specified in terms of a rectangular (Cartesian) coordinate system by giving the x and y coordinates of the point. Where three dimensions are involved, a rectangular coordinate system utilizing three mutually perpendicular axes is normally used; the position of a point is specified by its x, y, and z coordinates.

Equations relating the x and y coordinates are used to describe a plane figure. For example, $y = mx + b$ is the equation of a straight line in the xy plane, where m is the slope of the line, and b is the y intercept. Equations relating x, y, and z coordinates are used to describe a surface in space. Thus $x^2 + y^2 + z^2 = r^2$ is the surface of a sphere of radius r whose center is at the origin.

The position of a point in a plane may also be specified through the use of the polar coordinates r and θ. The length of the line from the origin of the coordinates O to the point P is the coordinate r, while the coordinate θ is the angle between the x axis and the line OP, measured in the counterclockwise direction.

1-10 Displacement. Scalar and Vector Quantities

If an object initially at a position P_1 with coordinates x_1, y_1 is displaced to a point P_2 with coordinates x_2, y_2, its *displacement* may be described in terms of the distance it has been moved parallel to the x axis and the distance it has moved parallel to the y axis. We say that the x displacement is the final x coordinate minus the initial x coordinate. Thus if in Figure 1-3 the point P_1 is at $(x, y) = (10, -5)$ and P_2 is at $(5, 10)$, the x displacement is $5 - 10 = -5$; that is, the object has been displaced 5 units in the direction of the negative x axis. Similarly, the y displacement is given as $10 - (-5) = +15$; that is, the object has been displaced 15 units in the positive y direction. The actual displacement from P_1 to P_2 is specified by the arrow in Figure 1-3 and designated by the boldface letter $\mathbf{r}$; the x and y components of the displacements are shown in dotted lines. The magnitude of the displacement is the length of the arrow, and its direction is from P_1 to P_2. To specify completely the displacement of an object, both

the magnitude of the displacement and its direction must be given. One way of specifying a displacement is to give its x and y components. A second way to describe the displacement from P_1, the initial point, to P_2, the final point, is to imagine a new coordinate system with axes parallel to x and y whose origin is at P_1. In terms of the new coordinate system, the displacement is described by the polar coordinates of the point P_2 which give both the magnitude r and the direction θ along which the displacement is to be made.

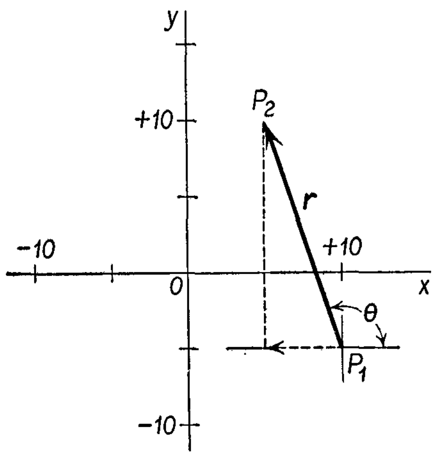


Fig. 1-3 Displacement.

We shall be dealing with two types of quantities in physics: *scalar* quantities, of which length, mass, and time are examples; and *vector* quantities, of which displacement and force are examples.

A scalar quantity has magnitude only and is specified by a number together with the appropriate unit, for example, a length of 2 m or 78.74 in. When describing a vector quantity, we must specify its *direction* as well as its magnitude. The method for doing this is suggested by our study of displacements; a vector quantity is represented by a *vector* which consists of an arrow drawn in the direction of the vector quantity with the head of the arrow showing the *sense* of the vector, and the length of the arrow showing the magnitude of the vector drawn to some arbitrary scale. For example, a displacement of 6 mi 30° north of east can be represented by an arrow 1 in. long, as shown in Figure 1-4, with its line at an angle of 30° to the conventional east direction. The scale of this drawing is 1 in. = 6 mi. Other vector quantities can be represented in a similar manner with a suitable choice of scales.

1-11 Addition of Vectors

The mathematical operations of addition, subtraction, multiplication, and division that are familiar to all of us have been developed for scalar quantities. We must now proceed to develop a set of mathematical operations suitable for vector quantities. In developing these operations we shall be guided by the results of physical processes. For example, the process of addition of vectors can be based upon the process involved in the addition

of displacements. We can then extend this to the addition of other vector quantities, checking the results to see whether the method is applicable.

Suppose that we consider the displacement of a particle from point P_0 to point P_1, as shown in Figure 1-5; this to be followed by the displacement from P_1 to P_2. The same result could have been accomplished by the single displacement from P_0 to P_2. Thus the displacement represented by the vector $\mathbf{P_0P_2}$ is equal to the sum of the displacements $\mathbf{P_0P_1}$ and $\mathbf{P_1P_2}$. We

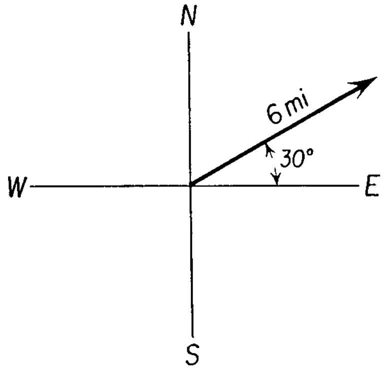


Fig. 1-4

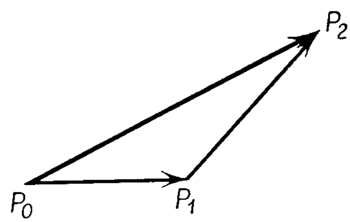


Fig. 1-5 Addition of displacements.

thus have a rule for the addition of two vectors **A** and **B**. This rule may be stated as follows: Starting at any arbitrary point, and using any convenient scale, draw vector **A** equal and parallel to itself and pointing in its direction.

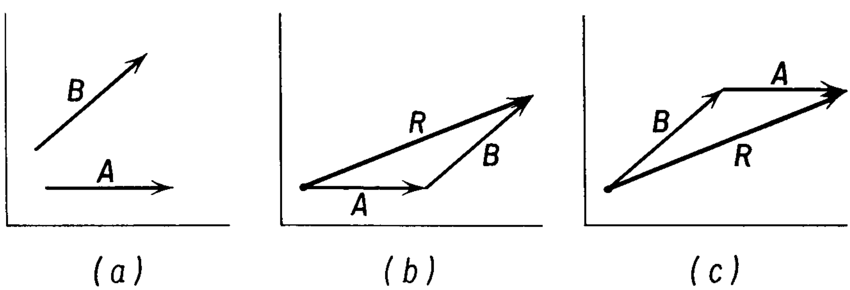


Fig. 1-6 Vector addition.

At the head of vector **A**, start the tail of vector **B** and draw **B** equal and parallel to itself and in its direction. To find the sum of vector **A** and vector **B**, draw a vector from the origin, or tail, of **A** to the end, or head, of **B**. This vector **R** is the sum of **A** and **B**, as shown in Figure 1-6. It is also called the *resultant* of **A** and **B**. Thus

$$\mathbf{R} = \mathbf{A} + \mathbf{B}.$$

The order in which this addition is performed is immaterial, as shown in Figure 1-6(c), in which **B** is drawn first and **A** is added to it. Thus

$$\mathbf{R} = \mathbf{B} + \mathbf{A} = \mathbf{A} + \mathbf{B}. \tag{1-2}$$

It may be remarked here that identical results could have been obtained by a geometrical method by starting with **A** and **B** at a common

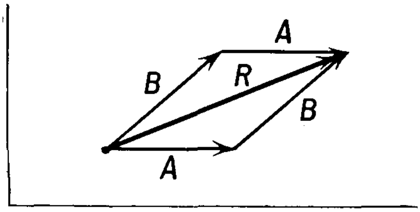


Fig. 1-7 Parallelogram method for the addition of two vectors.

point, as shown in Figure 1-7, and drawing a parallelogram with **A** and **B** as sides; the diagonal of the parallelogram is the resultant **R**. It will be noted that each of the earlier figures is one half of the parallelogram with **R** common to each triangle.

The above method of adding two vectors is called the *parallelogram method.* It can be extended to any number of vectors by first determining the resultant of two of the vectors, then adding a third vector to this resultant, and so forth.

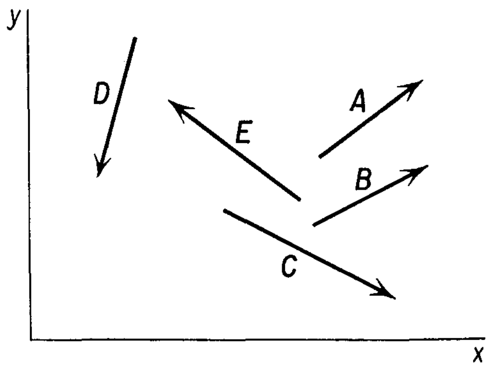


Fig. 1-8

In the preceding discussion we have used **boldface type** to represent *vector* quantities. This procedure will be followed throughout the text.

Suppose that the arrows of Figure 1-8 represent a set of instructions for the displacement of an object from the point P_0. These displacements can be applied to the object in any sequence with the same result. In Figure 1-9 the vector **A** has been transferred to the point P_0. The head

of the arrow at the point P_1 represents the first new position of the object. To the point P_1 as a new origin, we apply the vector **B**, and then successively add vectors **C**, **D**, and **E** in head-to-tail fashion to find the final position of the object at the point P_5. A single displacement given by the vector

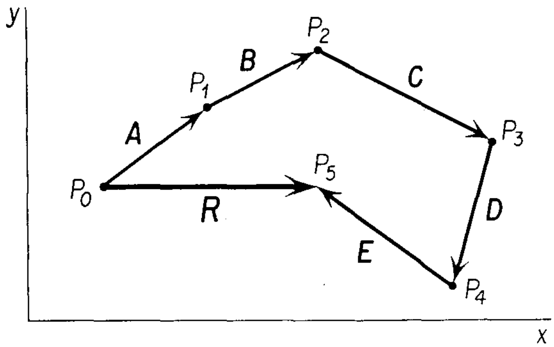


Fig. 1-9 Polygon method for the addition of vectors.

R, Figure 1-9, will also locate the object at the point P_5. The vector **R** is therefore completely equivalent to the set of operations **A**, **B**, **C**, **D**, and **E**. The resultant **R** is said to be the sum of the vectors. This result may be expressed in equation form as

$$\mathbf{R} = \mathbf{A} + \mathbf{B} + \mathbf{C} + \mathbf{D} + \mathbf{E}. \tag{1-3}$$

The procedure just described is called the *polygon method for the addition of vectors*, for the resulting figure is a closed polygon. *To add a number of vectors by the polygon method, the vectors are redrawn in sequence from some arbitrary origin, placing the vectors head to tail. The sum of the several vectors is the single vector drawn from the tail of the first vector to the head of the last vector*. Just as in the case of two vectors, the order in which vectors are added makes no difference in the final result.

When summing two vectors by the triangle method, the resulting figure is a triangle, made up of the two vectors and their resultant. It is often convenient to apply the law of sines or the law of cosines to find both the magnitude and the direction of the resultant.

Illustrative Example. Find the sum of vectors **A** and **B** in Figure 1-10. Vectors **A** and **B** are redrawn, connected head to tail, as shown in the figure. Vector **R** represents their sum. Since A, B, and θ are known, the magnitude of **R** may be determined, using the law of cosines. Thus

$$R^2 = A^2 + B^2 - 2AB\cos\theta.$$

Now that the magnitude of R has been found, the angle ϕ which describes the orientation of **R** with respect to the x axis may be found from a second application

of the law of cosines, or from the law of sines:

$$\frac{\sin\phi}{B} = \frac{\sin\theta}{R},$$

$$\sin\phi = \frac{B\sin\theta}{R},$$

from which
$$\phi = \arcsin\frac{B\sin\theta}{R}.$$

In this example, as elsewhere in the text, the italic letters A, B, and R have been used to represent the magnitude of the vectors **A**, **B**, and **R**.

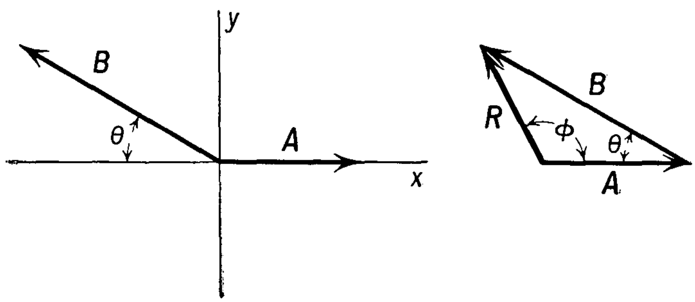


Fig. 1-10

1-12 Subtraction of Vectors

The subtraction of vectors may be accomplished by the rules developed for the addition of vectors once the meaning of a negative vector has been established, for we shall adhere to the notation developed in algebra that

$$\mathbf{A} - \mathbf{B} = \mathbf{A} + (-\mathbf{B}). \tag{1-4}$$

The negative of a vector is a vector of the same length but pointing in the opposite direction. The sequence of operations performed in subtracting a vector **B** from a vector **A** is described in Figure 1-11(a). First the vector **B** is reversed in direction to become the vector $-\mathbf{B}$. Then the vector $-\mathbf{B}$ is added to the vector **A** by the triangle method. Finally the resultant of the operation $\mathbf{A} - \mathbf{B}$ is displayed. In Figure 1-11(b) the sum of the same two vectors is displayed. Just as in algebra, the difference of two quantities differs from their sum.

Because the addition of two vectors requires consideration of direction as well as magnitude, the algebra of vectors is distinctly different from the algebra of scalars, leading to some superficially strange results. For example, a vector of length 5 added to a vector of length 3 does not necessarily

yield a vector of length 8. The addition may lead to a vector of any length between 2 and 8, depending on the orientations of the two vectors. Similarly, a vector of length 3 subtracted from a vector of length 5 does not

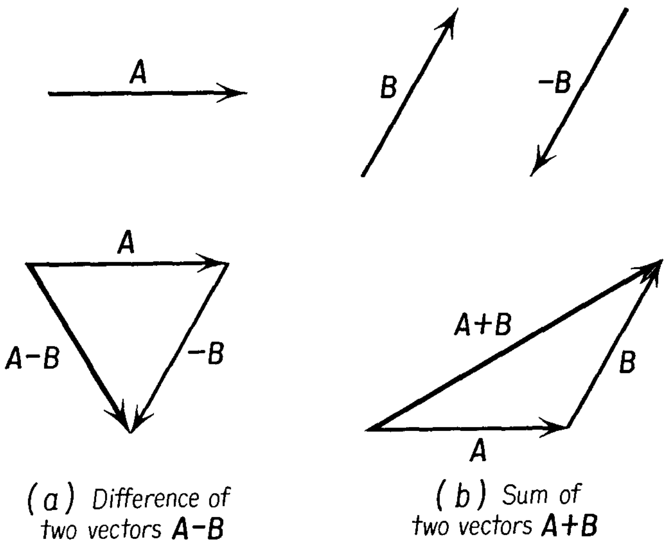


Fig. 1-11 (a) Difference of two vectors, **A** − **B**. (b) Sum of the same two vectors, **A** + **B**.

necessarily yield a vector of length 2. Again, the subtraction may yield a vector of any length between 2 and 8, depending upon the orientations of the two vectors.

1-13 Resolution of Vectors; Components

Any number of vectors which, added together, form a resultant **R** may be considered as *components* of **R**. Thus a vector may be resolved into any desired number of components.

Most frequently, the components which are sought are those which are mutually independent, with one component in some specified direction. Any two vectors which are perpendicular to each other are mutually independent. For example, if we refer the displacement to an x-y Cartesian coordinate system, a displacement of a body in the y direction produces no change in its x coordinate; similarly, a displacement in the x direction produces no change in the y coordinate of the body. Thus, if we are interested in determining two components of a vector **A** one of which is parallel to the x axis, the other component must be parallel to the y axis. Referring

to Figure 1-12, if the vector **A** makes an angle θ with the x axis, its x component is of magnitude

$$A_x = A \cos \theta, \tag{1-5a}$$

and its y component is of magnitude

$$A_y = A \sin \theta. \tag{1-5b}$$

The magnitude of A can be found in terms of its components in a very simple way, since

$$A = (A_x^2 + A_y^2)^{1/2}. \tag{1-6}$$

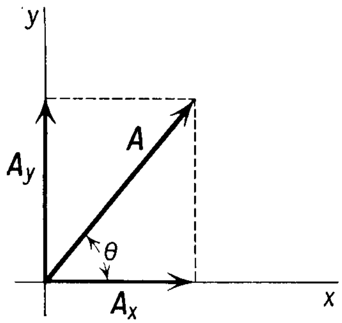


Fig. 1-12 Resolution of a vector into two rectangular components.

The concept of rectangular components of a vector can be utilized for an analytical method of adding vectors. Suppose that we wish to obtain the resultant of a number of coplanar vectors, that is, vectors all in one plane, the x-y plane; these may be written as

$$\mathbf{A} + \mathbf{B} + \mathbf{C} + \cdots = \mathbf{R}. \tag{1-7}$$

Let us call the x components of these vectors $\mathbf{A}_x$, $\mathbf{B}_x$, $\mathbf{C}_x$, . . . respectively, and similarly for the y components $\mathbf{A}_y$, $\mathbf{B}_y$, $\mathbf{C}_y$ . . . . Let θ_1 be the angle between **A** and the x axis, θ_2 the angle between **B** and the x axis, and so forth.

Then $$A_x = A \cos \theta_1, \tag{1-8a}$$

and $$A_y = A \sin \theta_1. \tag{1-8b}$$

In a similar manner

$$B_x = B \cos \theta_2,$$

$$B_y = B \sin \theta_2, \quad \text{and so on.}$$

The x component of the resultant R_x is

$$\mathbf{R}_x = \mathbf{A}_x + \mathbf{B}_x + \mathbf{C}_x + \cdots, \tag{1-9}$$

and similarly for the y component of the resultant.

From a knowledge of the components of the resultant **R**, the magni-

tude of the resultant may be obtained from the Pythagorean theorem.

$$R = (R_x^2 + R_y^2)^{1/2}. \tag{1-10}$$

The angle θ that **R** makes with the x axis is given by

$$\cos\theta = \frac{R_x}{R}, \tag{1-11a}$$

or

$$\sin\theta = \frac{R_y}{R}, \tag{1-11b}$$

or

$$\tan\theta = \frac{R_y}{R_x}. \tag{1-11c}$$

These results may be easily extended to three dimensions.

Illustrative Example. Find the resultant of the vectors **A**, **B**, **C**, in the x-y plane, as shown in Figure 1-13. For purposes of illustration, we shall consider

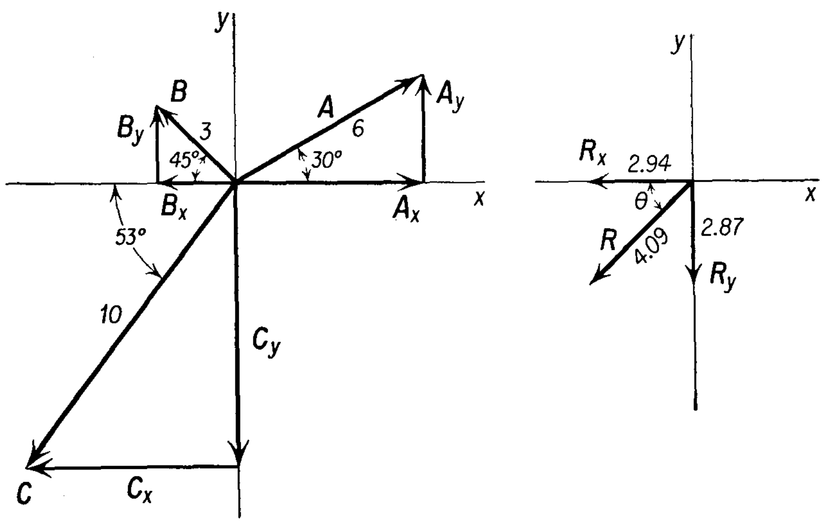


Fig. 1-13

that these vectors represent displacements and that all lengths are given in feet. The x component of the resultant is

$$\mathbf{R}_x = \mathbf{A}_x + \mathbf{B}_x + \mathbf{C}_x, \tag{1-9}$$

so that

$$R_x = A\cos 30° - B\cos 45° - C\cos 53°$$
$$= 6 \times 0.866 - 3 \times 0.707 - 10 \times 0.602,$$

which yields

$$R_x = -2.94 \text{ ft}.$$

The y component of the resultant is

$$\mathbf{R}_y = \mathbf{A}_y + \mathbf{B}_y + \mathbf{C}_y,$$

so that

$$R_y = 6 \sin 30° + 3 \sin 45° - 10 \sin 53°$$
$$= 6 \times 0.500 + 3 \times 0.707 - 10 \times 0.799,$$

which yields

$$R_y = -2.87 \text{ ft.}$$

The magnitude of the resultant is given by

$$R = (R_x^2 + R_y^2)^{1/2} \tag{1-10}$$
$$= [(-2.94)^2 + (-2.87)^2]^{1/2},$$

so that

$$R = 4.09 \text{ ft.}$$

The angle that $\mathbf{R}$ makes with the x axis, denoted by (R, x), is given by

$$\cos (R, x) = \cos \theta = \frac{2.94}{4.09} = 0.719,$$

$$\theta = 46° = 0.80 \text{ radian.}$$

Since $\mathbf{R}$ lies in the third quadrant, we state the result as: The angle made by the resultant with the positive x axis is 226°, or 3.94 radians.

1-14 Other Vector Operations

Methods for obtaining the sum and difference of vectors have already been displayed. One additional operation should be discussed here. A vector may be multiplied by a scalar $+a$, and the result is still a vector. The new vector points in the same direction as the old but has a times the magnitude of the old vector. The result of multiplication of a vector by -1 is a vector of the original length but pointing in the opposite direction. Combining these two definitions, it is now possible to multiply a vector by a negative scalar. The definition further permits the division of a vector by a scalar, for division of a vector by a scalar b is equivalent to multiplication by the scalar $1/b$. The result of this operation is always a vector quantity. Thus velocity is a vector quantity, for it is the result of dividing displacement, a vector, by time, a scalar.

Problems

1-1. Check the conversion factors given in Table 1-3, starting with the legal definition of the yard.

1-2. A ship at sea receives radio signals from two radio transmitters A and B 125 mi apart, one due south of the other. The direction finder shows that transmitter A is 30° south of east, while transmitter B is due east. How far is the ship from each transmitter?

1-3. What are the polar coordinates of a point having the following rectangular coordinates?

(a) $x = 3, y = 4$ (b) $(x, y) = (5, -5)$ (c) $(x, y) = (-12, 6)$

1-4. What are the rectangular coordinates of a point having the following plane polar coordinates?

(a) $(r, \theta) = (6, 30°)$ (b) $(r, \theta) = (10, 53°)$
(c) $(r, \theta) = (12, 0.5 \text{ radians})$ (d) $(r, \theta) = (-9, 3.4 \text{ radians})$

1-5. A displacement vector lying in the x-y plane has magnitude 5 m and makes an angle of 60° with the x axis. Find its x and y components.

1-6. A displacement vector has an x component of +3 ft and a y component of +4 ft. Find the magnitude and direction of the vector.

1-7. A displacement vector has an x component of +3 m, a y component of +5 m, and a z component of +9 m. Find the magnitude of the resultant, and the angles made with the x, y, and z axes.

1-8. Find the resultant of two vectors, the first of length 7 ft making an angle of 25° with the positive x axis, the second of length 15 ft making an angle of 75° with the same axis, all angles being measured counterclockwise: (a) using the method of x and y components; (b) using the law of sines, the law of cosines, or both.

1-9. The coordinates (x, y) of the head and tail of two vectors are:

Vector	Head Coordinates	Tail Coordinates
A	(0, 3)	(3, 7)
B	(5, 2)	(9, −3)

Find the sum **A** + **B** and the difference **A** − **B** of these vectors (a) using the method of components and (b) using the law of sines, the law of cosines, or both.

1-10. Find the components of the resultant of the following displacements: 2 ft north, 6 ft southwest, 5 ft south, 7 ft northwest.

1-11. Three coplanar vectors **A**, **B**, **C** are drawn radially outward from the origin of coordinates; the polar coordinates (r, θ) of their heads are **A**: (5, 30°), **B**: (7, 150°), and **C**: (4, 240°). (a) Find their resultant. (b) Find the magnitude and direction of **R** where

$$\mathbf{R} = 3\mathbf{A} - 2\mathbf{B} + \frac{\mathbf{C}}{2}\cdot$$

1-12. Derive the law of cosines by use of the component method for the addition of vectors.

1-13. Derive the law of sines by making use of vector concepts.

1-14. A vector of length A is in the positive x direction, and a vector of length B is in the positive y direction. Their sum is a vector of length 5 which makes an angle of 37° with the positive x axis. Find the magnitudes of **A** and **B**.

1-15. A vector A of length 4 in the positive x direction is added to a vector B to yield a resultant of length 7. The resultant is in the first quadrant and makes an angle of 45° with the x axis. Find the magnitude and direction of the vector **B**.

2

Motion of a Particle (Kinematics)

2-1 Motion Is Relative

Normally, when we say an object is at rest, we mean that it is at rest with respect to the surface of the earth; when we say a car is moving at a speed of 40 mi/hr, we imply that the motion is taking place at this speed relative to the road. A boat sailing on the river moves with respect to the river's banks, but it also moves with respect to the flowing water in the river. The lift on the wings of an airplane is generated by the motion of the airplane through the air, but it is quite important to know the plane's motion with respect to the ground. When we speak of the motion of a car or a train, we normally mean the motion with respect to the ground, but when we speak of the rated speed of an airplane, we refer to its motion with respect to the air. To avoid confusion in the discussion of motion, it is important to refer the motion to a *frame of reference,* usually thought of as fixed on the earth or fixed relative to the stars, in which the motion is measured. For many problems it is convenient to use moving frames of reference; it is then necessary to specify the nature of the motion of the frame. The frame of reference generally takes the form of a set of coordinate axes in which the motion is pictured.

2-2 Uniform Motion in a Straight Line

The simplest type of motion is that in which a body traverses equal distances along a straight line in equal time intervals; this type of motion is called *uniform motion in a straight line.* The *speed* of the body is defined as the *distance traversed divided by the time elapsed;* that is,

$$\text{Speed} = \frac{\text{distance}}{\text{time}},$$

or, in symbols,

$$v = \frac{s}{t}, \tag{2-1}$$

where v is the speed of the body, and s is the distance traversed in time t.

Whenever a number is used to specify the speed of a body, it must always be accompanied by the appropriate units such as feet per second, usually written as ft/sec, or miles per hour (mi/hr), or meters per second (m/sec), or any other appropriate units of distance and time.

One other aspect of motion is the direction in which it takes place. When we wish to specify that a body has moved from a point A to a point B, we can use a vector directed from A to B; this vector is the *displacement* of the body. If s is the distance from A to B, the displacement is $\mathbf{s}$ (printed in boldface type), a vector drawn *from A to B.*

To specify both the speed of a body and its direction of motion, we use the term *velocity.* The *velocity* $\mathbf{v}$ *of a body in uniform motion in a straight line is defined as the displacement divided by the time during which the displacement occurred,* or, in symbols,

$$\boxed{\mathbf{v} = \frac{\mathbf{s}}{t}\cdot} \tag{2-2}$$

The direction of the velocity is the same as that of the displacement. Velocity is thus a vector quantity. For example, if a train is moving due west with a uniform speed of 60 mi/hr, its velocity is 60 mi/hr west. No statement about the velocity of a body is complete without specifying *both* magnitude and direction.

2-3 Relative Velocities

It is frequently important to be able to determine the velocity of a body with respect to one frame of reference when its velocity has been determined with respect to a second frame of reference which is in motion with respect to the first one. For example, the velocity of a ship relative to the water can easily be measured, but what is usually desired is its velocity relative to the shore, for the water is generally in motion.

To understand how these velocities are related to each other, let us consider the case of a boat in the river in which the water is moving downstream with a velocity $\mathbf{w}$ relative to its banks. Let us assume that the boat, if left free, would float downstream with the current; that is, its velocity relative to the water would be zero, but its velocity relative to the banks would be the same as that of the water. Suppose now that the engines of the boat are started and that the boat moves with a velocity $\mathbf{u}$ relative to the water. Its velocity $\mathbf{v}$ relative to the banks will therefore be the resultant of the two velocities—the velocity $\mathbf{w}$ which it acquires because it is moving with the water, and the velocity $\mathbf{u}$ which it acquires relative to the water.

In the form of an equation,

$$\boxed{\mathbf{v} = \mathbf{w} + \mathbf{u}.} \tag{2-3}$$

The three quantities involved are vector quantities, and the addition must be performed by vector methods. As a simple illustration, suppose that the velocity of the current in a river is 3 mi/hr south and that a boat heads toward the west with a velocity of 4 mi/hr with respect to the water,

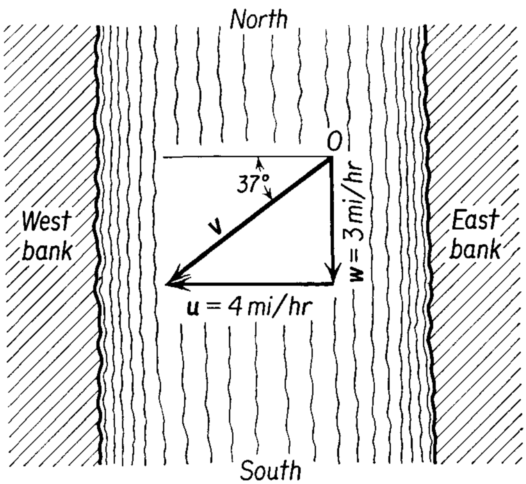


Fig. 2-1 A boat headed west in a river which flows south actually travels south of west.

as shown in Figure 2-1. It is desired to find the velocity of the boat relative to the shore. The velocity **u** of the boat relative to the water is added vectorially to the velocity **w** of the water relative to the shore to get the resultant velocity **v** of the boat relative to the shore. Its value is 5 mi/hr directed at an angle of 37° south of west.

Illustrative Example. Consider the case of an airplane which has to fly from New York to Montreal, due north. Suppose that its normal flying speed in still air is 200 mi/hr. During the flight there is a steady northwest wind of 40 mi/hr. In what direction should the airplane be headed in order to go due north? What will be its speed relative to the ground?

To solve the problem we draw a vector **w** to an appropriate scale, representing the magnitude and direction of the wind, as shown in Figure 2-2. To find the heading of the airplane which will result in a displacement due north, we first draw a straight line due north from O, the tail of vector **w**. From the head of vector **w**, we swing an arc of radius u to intersect this line. The vector **v** from O due north to the arc represents the resultant velocity of the airplane with respect to the ground; its magnitude is the ground speed. The desired heading of the airplane is given by the vector **u**, as shown on the figure. From the figure, if $u = 200$

mi/hr, and w = 40 mi/hr, then, to the same scale, v = 170 mi/hr. The airplane must be headed about 8° west of north.

2-4 Instantaneous Speed and Velocity

We have thus far confined our discussion to the simplest type of motion, that with constant velocity. Of very great interest is the motion of a body in which its velocity changes. Since velocity is a vector quantity, a change in velocity will occur whenever (a) the speed of the body changes while its direction remains the same, (b) the direction of motion changes while its speed remains the same, or (c) its speed and the direction of its motion change simultaneously. Whenever the *velocity* of a body *changes* in any manner whatever, the motion of the body is said to be *accelerated.*

In order to be able to discuss accelerated motion, it is important to know how to specify the speed and the velocity of the body at any instant or at any point in its path. Suppose that the motion takes place along the x axis. The *average speed* of the motion is defined as the *distance traversed divided by the elapsed time.* If the object is at x_1 at a time t_1, and then is at x_2 at a subsequent time t_2, the average speed $\bar{v}$ (read v bar) may be defined in the form of an equation as follows:

Fig. 2-2 Relative velocities of an airplane, **v** with respect to the earth and **u** with respect to the air, when there is a wind of velocity **w** from the northwest.

$$\bar{v} = \frac{x_2 - x_1}{t_2 - t_1}. \tag{2-4}$$

Illustrative Example. Find the average speed with which a car travels down a straight highway 100 mi long if its speed during the first 50 mi is 25 mi/hr and during the second 50 mi, is 75 mi/hr.

Our first reaction may be to say that the average speed is 50 mi/hr, but this is incorrect. To find the average speed, the distance must be divided by the

elapsed time. The distance traversed is 100 mi. The elapsed time is 2 hr for the first 50 mi, and $\frac{50}{75}$ hr for the second 50 mi. The total elapsed time is $2\frac{2}{3}$ hr.

$$\bar{v} = \frac{100 \text{ mi}}{2.67 \text{ hr}} = 37.5 \frac{\text{mi}}{\text{hr}} \cdot$$

The instantaneous speed at any point P in the path of a body moving in a straight line, say the x direction, can be found by taking the average speed during a short time interval $\Delta t = t_2 - t_1$ during which the particle

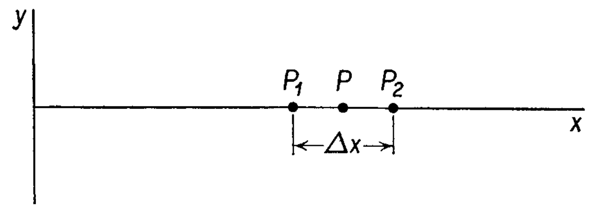


Fig. 2-3

has moved a distance Δx from P_1 to P_2, points on either side of P (see Figure 2-3). The average speed of the body is $\Delta x/\Delta t$. As the time interval Δt is made shorter, the points P_1 and P_2 close in on point P, and, in the language of the calculus, we study the value of the average speed over a sequence of *nested* intervals. The sequence of values of the average speed approaches a constant value called a *limit;* this limit is the instantaneous speed v_x of the body at point P; thus

$$v_x = \lim_{\Delta t \to 0} \frac{\Delta x}{\Delta t} = \frac{dx}{dt} \cdot \tag{2-5}$$

As in the calculus, the symbol Δ (delta) has been used to indicate small increments. Thus if the position may be described as a function of time,

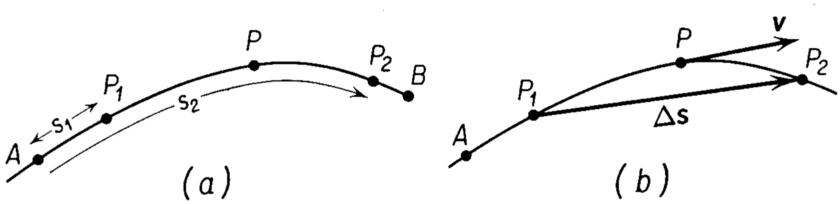


Fig. 2-4 Displacement in curvilinear motion.

the instantaneous speed is the derivative of the position with respect to time. In the case of *uniform* motion in a straight line, the average speed is the same as the instantaneous speed.

Consider the motion along a curved path, sometimes called *curvilinear* motion, such as a motion along APB of Figure 2-4. The average speed of

the body is again the distance traversed divided by the elapsed time. The distance traveled along the curved path is denoted by s, measured from some arbitrary reference point A on the path. If s_1 is the distance of P_1 from A and s_2 the distance of P_2 from A, the average *speed* from P_1 to P_2 [Figure 2-4(a)] will be given by

$$\bar{v} = \frac{s_2 - s_1}{t_2 - t_1}, \tag{2-6}$$

where the body passes point P_1 at time t_1 and point P_2 at time t_2.

If we let

$$\Delta s = s_2 - s_1$$

and

$$\Delta t = t_2 - t_1,$$

then

$$\bar{v} = \frac{\Delta s}{\Delta t}.$$

The instantaneous speed v may then be determined by a limiting process as

$$v = \lim_{\Delta t \to 0} \frac{\Delta s}{\Delta t} = \frac{ds}{dt}. \tag{2-7}$$

The average *velocity* over the interval P_1P_2 is the vector displacement divided by the time, rather than the scalar distance divided by the time. Calling the displacement of the particle from P_1 to P_2 by $\Delta\mathbf{s}$, and the time interval for performing this displacement Δt, the average velocity in the neighborhood of P becomes

$$\bar{\mathbf{v}} = \frac{\Delta \mathbf{s}}{\Delta t}. \tag{2-8}$$

The displacement $\Delta\mathbf{s}$ will not, in general, coincide with the actual path from P_1 to P_2, but as the two points are taken closer and closer to P, the displacement practically coincides with the actual path along the curve. The direction of the displacement is then tangent to the path at P. The magnitude of the instantaneous velocity $\mathbf{v}$ at P is the instantaneous speed v at P, and its direction is tangent to the path at P, as shown in Figure 2-4(b). The instantaneous velocity may be written as

$$\boxed{\mathbf{v} = \lim_{\Delta t \to 0} \frac{\Delta \mathbf{s}}{\Delta t} = \frac{d\mathbf{s}}{dt}.} \tag{2-9}$$

Equations (2-8) and (2-9) are *vector* equations. Both sides of the equations contain vector quantities. All equations in physics must relate the same kinds of things. The two sides of the equation must not only

have the same dimensions but must also relate quantities of the same character.

We can visualize the meaning of the vector Equations (2-8) and (2-9) more easily by considering a case in which the displacement $\Delta\mathbf{s}$ is entirely in the x direction. Since the velocity vector is parallel to the displacement, the velocity vector must be in the x direction. Writing $\mathbf{v}_x$ for velocity in the x direction, and $\mathbf{x}$ for displacement in the x direction, we substitute in Equation (2-9) to obtain

$$\mathbf{v}_x = \lim_{\Delta t \to 0} \frac{\Delta \mathbf{x}}{\Delta t} = \frac{d\mathbf{x}}{dt}. \tag{2-10}$$

But now there can be no confusion about the direction of the displacement or about the direction of the velocity, since all directions are parallel to the x axis. Hence we may rewrite Equation (2-10) as a scalar equation in algebraic quantities.

$$v_x = \lim_{\Delta t \to 0} \frac{\Delta x}{\Delta t} = \frac{dx}{dt}. \tag{2-11}$$

By considering cases in which the displacement is entirely in the y direction, and in which the displacement is entirely in the z direction, we obtain

$$v_y = \frac{dy}{dt}, \tag{2-12}$$

and

$$v_z = \frac{dz}{dt}. \tag{2-13}$$

Any displacement in space may be thought of as the vector sum of its three component displacements, one in the x direction, one in the y direction, and one in the z direction. Hence Equations (2-11), (2-12), and (2-13) are true in the general case where the motion is in any direction, not just parallel to one of the coordinate axes. The velocity component in the x direction v_x depends only on the rate of change of the x coordinate of the position of the moving object. This conception of the separability of the components of the motion will greatly simplify the study of the motion of a particle, as we shall see when we study the motion of projectiles later in this chapter.

2-5 Acceleration

The discussion of motion with varying velocity can best be dealt with in a quantitative manner by the introduction of the concept of acceleration. *The acceleration of a body is defined as the change in its velocity divided by*

the time in which the change takes place. Just as in the case of velocity, we must distinguish between average acceleration and instantaneous acceleration. If the initial velocity of a body is $\mathbf{v}_1$ at a time t_1, and the final velocity is $\mathbf{v}_2$ at a time t_2, the average acceleration is, from the definition,

$$\bar{\mathbf{a}} = \frac{\mathbf{v}_2 - \mathbf{v}_1}{t_2 - t_1}. \tag{2-14}$$

The instantaneous acceleration is arrived at by examining the average acceleration obtained in a sequence of nested intervals converging on the point P where the acceleration is to be determined. Referring to Figure

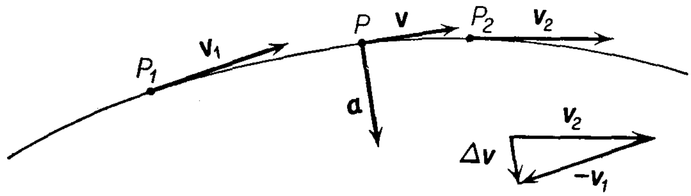


Fig. 2-5 Instantaneous acceleration **a** at a point P.

2-5, if the instantaneous velocity at point P_1 is given by $\mathbf{v}_1$, and the instantaneous velocity at P_2 is $\mathbf{v}_2$, the instantaneous acceleration at point P is given by

$$\boxed{\mathbf{a} = \lim_{\Delta t \to 0} \frac{\mathbf{v}_2 - \mathbf{v}_1}{t_2 - t_1} = \lim_{\Delta t \to 0} \frac{\Delta \mathbf{v}}{\Delta t} = \frac{d\mathbf{v}}{dt}.} \tag{2-15}$$

If the acceleration is constant, the average acceleration and the instantaneous acceleration are the same.

Let us consider a case in which the motion is entirely in the x direction. Writing a_x for the magnitude of the acceleration, v_x for the speed, and substituting in Equation (2-15), we find

$$a_x = \frac{dv_x}{dt}; \tag{2-16}$$

and for motions entirely in the y and z directions we obtain

$$a_y = \frac{dv_y}{dt}, \tag{2-17}$$

and

$$a_z = \frac{dv_z}{dt}. \tag{2-18}$$

Thus Equations (2-16), (2-17), and (2-18) are true for motion which takes

place in any direction and may be considered as the component forms of Equation (2-15). In the component forms the quantities appearing in the equations are all scalar algebraic quantities. The component forms of the equations are easier to use for the solution of most problems. The vector form is the easier way to remember the equations and to develop further theory.

As an illustration of accelerated motion, we can discuss the motion of an airplane under a variety of conditions. As it prepares to take off, the airplane moves along the runway with increasing speed; if this speed is changing at a constant rate, its acceleration is constant and is in the direction of its motion. This is the simplest type of accelerated motion. At the instant of take-off, the direction of the acceleration changes so that it will have a vertical component upward. After the plane has reached the desired altitude and speed and has leveled off in flight, it continues with constant velocity, that is, with zero acceleration. If, during the flight, the plane makes a turn at constant speed, its motion is again accelerated because the direction of motion is changing. When an airplane is preparing to land, it reduces its speed for safe landing. While the speed is being reduced, its motion is also accelerated; this time the direction of the acceleration is opposite to the direction of motion.

In this chapter we shall limit our discussion to motion with constant acceleration, leaving to later chapters the discussion of several cases of varying acceleration.

2-6 Straight-Line Motion under Constant Acceleration

The simplest type of motion with constant acceleration is that in which a body moves in a straight line with a speed which is increasing or decreasing at a *constant* rate. We may choose the direction of the x axis as along the direction of motion and rewrite Equation (2-14) as

$$\bar{a}_x = a_x = \frac{v_x - u_x}{t_f - t_i}, \tag{2-19}$$

in which v_x is the speed of the body in the x direction at the final time t_f, u_x is the speed of the body in the x direction at the initial time t_i, and a_x is the constant acceleration of the body in the x direction during this time interval.

Illustrative Example. An airplane approaching a landing field decreases its velocity from 250 mi/hr to 100 mi/hr in 20 sec. Find the acceleration.

Again we choose the x axis as the direction of motion. We set $u_x = 250$ mi/hr at time t_i, $v_x = 100$ mi/hr at $t_f = 20 \text{ sec} + t_i$, and we write

$$a_x = \frac{100 \text{ mi/hr} - 250 \text{ mi/hr}}{20 \text{ sec}} = \frac{-150 \text{ mi/hr}}{20 \text{ sec}} = -7.5 \frac{\text{mi}}{\text{hr sec}}.$$

Thus the speed of the airplane is decreasing at the rate of 7.5 mi/hr each second during the 20-sec interval. With the aid of the appropriate conversion factors, the above result may also be expressed as

$$a_x = -11 \frac{\text{ft/sec}}{\text{sec}} = -11 \frac{\text{ft}}{\text{sec}^2};$$

that is, the speed of the airplane is decreasing at the rate of 11 ft/sec each second during the 20-sec interval.

The question of positive or negative signs for the acceleration of a body, or for other quantities requiring some indication of their direction, is a matter of convenience. For most purposes it is preferable to adopt some consistent scheme of notation, working all problems in the same systematic way. This matter of approaching problems systematically is of great importance in physics and engineering. Problems which at first seem fiendishly difficult often yield to a persistent and systematic approach. The notation generally preferred is one in which the signs employed follow the usual right-handed rectangular coordinate system. The motion is thought of as occurring within a coordinate frame. The direction of the positive x axis is then the positive direction for the x components of the displacement, the velocity, and the acceleration. A motion described by a positive value of the velocity and a negative value of the acceleration is one in which the velocity is directed in the direction of increasing x, while the acceleration is in the opposite direction.

2-7 Equations of Motion for Constant Acceleration

When a body is moving in a straight line with constant acceleration a, we may derive its equations of motion most simply by the methods of the calculus. Dropping the subscript x from Equation (2-16), we have

$$\frac{dv}{dt} = a, \tag{2-16}$$

thus

$$dv = a\,dt,$$

which we may integrate as

$$\int dv = \int a\,dt,$$

yielding

$$v = at + u, \tag{2-20}$$

where u is a constant of integration. We note that at time $t = 0$ the speed v is equal to u. Thus u is the initial speed of the body.

From Equation (2-10), the definition of instantaneous velocity, we have

$$v = \frac{dx}{dt}, \tag{2-10}$$

and, substituting from Equation (2-10) into Equation (2-20) for v, we have

$$\frac{dx}{dt} = at + u,$$

or

$$\int dx = a\int t\,dt + \int u\,dt,$$

$$x = \tfrac{1}{2}at^2 + ut. \tag{2-21}$$

In Equation (2-21) we have set the initial position of the body at the origin $x = 0$ at time $t = 0$, thereby setting the constant of integration equal to zero. Equations (2-20) and (2-21) give the speed and position of the body as a function of time when the body is at the origin moving with speed u at the initial time $t = 0$, if the body is moving under constant acceleration a.

When a body is moving with constant acceleration in a straight line, its average speed is given by the average of its initial and final speeds. This may be shown by algebraic manipulation of Equations (2-20) and (2-21). We first factor the quantity t from the right-hand side of Equation (2-21) and write

$$x = (u + \tfrac{1}{2}at)t.$$

From Equation (2-20) we substitute

$$at = v - u$$

into the above equation to find

$$x = \frac{u + v}{2}\,t. \tag{2-22}$$

From the definition of the average speed as the distance traversed divided by the elapsed time

$$\bar{v} = \frac{x}{t},$$

we see that

$$\bar{v} = \frac{u + v}{2}; \tag{2-23}$$

that is, the average speed is the average of the initial and final speeds, for the case of constant acceleration.

A useful result may be obtained by eliminating the time t as a variable from Equations (2-20) and (2-22) to obtain an equation relating the initial speed u, the final speed v, and the distance traversed x. We may rewrite

Equation (2-20), by transposing the quantity u, as

$$v - u = at. \tag{2-24}$$

Both sides of Equation (2-22) may be multiplied by the quantity $2/t$ to obtain

$$v + u = \frac{2x}{t}\cdot \tag{2-25}$$

Multiplying the left-hand side of Equation (2-24) by the left-hand side of Equation (2-25) and setting the result equal to the product of the right-hand sides of the two equations yields

$$v^2 - u^2 = 2ax, \tag{2-26}$$

which is the result we have sought.

The principal results of the preceding section may be summarized in the following equations for the case of the motion of a particle with a constant acceleration a, with the particle starting at the origin with initial speed u at time $t = 0$.

$$v = u + at, \tag{2-27a}$$

$$x = ut + \tfrac{1}{2}at^2, \tag{2-27b}$$

$$v^2 = u^2 + 2ax, \tag{2-27c}$$

$$x = \bar{v}t, \tag{2-27d}$$

$$\bar{v} = \frac{u + v}{2}\cdot \tag{2-27e}$$

Equations (2-27) are repeated for emphasis, for they will be applied to problems and derivations many times throughout this text.

2-8 Freely Falling Bodies

One of the most common examples of motion with constant acceleration is that of a body which is dropped from any height and allowed to fall freely under the influence of gravity. By *free fall* we mean that such effects as air resistance (drag) or lift are assumed to be negligible. A falling leaf, a dandelion seed, and a glider dropped from some altitude are not freely falling bodies. A good approximation to a freely falling body may be obtained by dropping a round, heavy object. This is the motion that was first studied by Galileo. The results of many different experiments performed under many different conditions show that the acceleration of a freely falling body at any point near the earth's surface is a constant for that particular place and does not depend on the weight of the falling object.

The magnitude of the acceleration of gravity varies slightly with altitude, with latitude, and from point to point on the earth's surface having the same latitude and altitude. This is the subject of considerable geophysical exploration. If, at a certain place, the acceleration of gravity is slightly

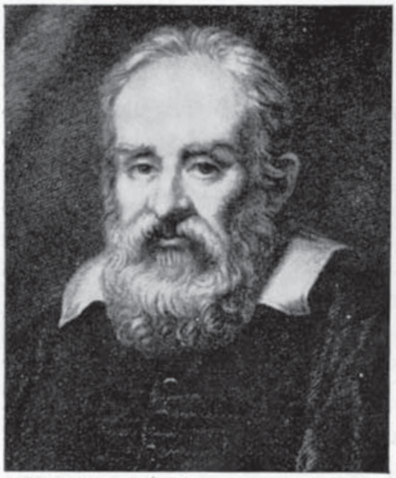


Fig. 2-6 Galileo Galilei (1564–1642). Discovered the laws of motion of freely falling bodies and of bodies moving along inclined planes. Constructed a telescope with which he observed the surface features of the moon and discovered four of the moons of Jupiter; his observations helped establish the validity of the heliocentric theory of the universe. (Courtesy of *Scripta Mathematica*.)

high or slightly low, the geophysicist uses this information to help find buried ore or oil bodies beneath the surface of the ground.

For the purposes of most calculations, the value of the acceleration of a freely falling body, designated by g, may be taken as

$$g = 32\,\frac{\text{ft}}{\text{sec}^2} = 980\,\frac{\text{cm}}{\text{sec}^2} = 9.8\,\frac{\text{m}}{\text{sec}^2}\cdot$$

The direction of the acceleration of gravity is vertically downward. We shall always use the symbol g to indicate a positive number, as above, and the direction appropriate for use in the sign conventions of a particular problem will always be explicitly shown as $+g$ or $-g$. For very accurate calculations the value of g appropriate to a particular place should be used. These values may be found in tables of physical constants. For reference, a few values of g at different latitudes and at sea level are given in Table 2-1.

A clearer understanding of the significance of acceleration can be obtained from the detailed consideration of the free fall of a baseball that is dropped from the top of a very tall building or from a cliff. When we say that a ball is dropped, we mean that its initial velocity is zero. Since it is accelerated downward at the rate of 32 ft/sec^2, at the end of the first second it will have acquired a velocity of 32 ft/sec downward. Its average velocity during the first second is 16 ft/sec, and the distance traveled during the time is 16 ft. At the end of the second second, its downward velocity will

TABLE 2-1 VALUES OF g

Latitude	g in cm/sec^2	g in ft/sec^2
0°	977.989	32.0862
30°	979.295	32.1290
45°	980.600	32.1719
60°	981.905	32.2147
90°	983.210	32.2575

have been increased by another 32 ft/sec to 64 ft/sec. At the end of the third second, it will again have acquired an additional velocity increment of 32 ft/sec so that its average velocity during the 3 sec is 48 ft/sec, and the distance traversed is 144 ft. Figure 2-7 shows the positions of the ball at 1-sec intervals and the corresponding velocities.

In our discussion of freely falling bodies, the effect of the air on the motion of a body through it was neglected. This discussion thus presents only a first approximation to the actual motion. In many cases this description is sufficiently accurate. However, when the velocity of the body is very great, such as the velocity of a bullet, or if the body is very small, such as a raindrop, or if the body presents a very large surface, such as a parachute, the resistance of the air plays an important part in determining the motion of the body.

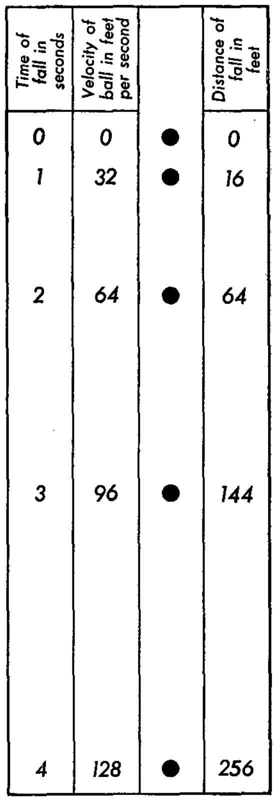


Fig. 2-7 The free fall of a baseball, showing the positions of the ball at intervals of 1 sec and the corresponding velocities.

Illustrative Example. Suppose that a ball is thrown vertically upward with an initial velocity of 80 ft/sec. Determine (a) how high it will go, (b) what velocity it will have as it moves down past its original point of projection, (c) its position 6 sec after it was thrown upward, and (d) the velocity with which it will be moving at this time.

Let us choose a set of coordinates with the origin O at the point of projection, and let us take the y axis as the line of motion (see Figure 2-8). The displacement from the origin will be measured by the y coordinate of the ball; it will be considered positive above the origin and negative below the origin. The acceleration is downward, and its magnitude is

g at all times and at all points of the path. Rewriting Equations (2-27b) and (2-27a) for motion in the y-direction, we get

$$y = u_y t + \tfrac{1}{2} a_y t^2,$$

$$v_y = u_y + a_y t.$$

(a) At the highest point of the path the ball will stop momentarily; this means that $v_y = 0$. Other known quantities are $a_y = -g$ and $u_y = 80$ ft/sec.

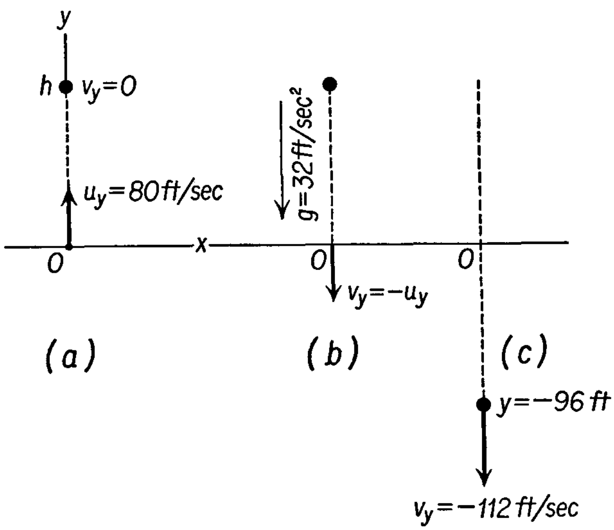


Fig. 2-8 (a) A ball thrown upward reaches a height h at which $v = 0$. (b) On its return journey it passes the origin with a speed equal to its initial speed but in the opposite direction. (c) Position and velocity of the ball 6 sec after it started its motion.

The unknown quantities are y and t. Having two equations in the two unknowns, the problem may be easily solved by the methods of algebra. Making the necessary substitutions,

$$y = 80 \frac{\text{ft}}{\text{sec}} \times t - \tfrac{1}{2} \times 32 \frac{\text{ft}}{\text{sec}^2} \times t^2,$$

$$0 = 80 \frac{\text{ft}}{\text{sec}} - 32 \frac{\text{ft}}{\text{sec}} \times t.$$

Solving these equations, we find $t = 2.5$ sec and $y = 100$ ft. Thus the highest point of the path is 100 ft above the initial point of projection.

(b) When the ball falls down and passes its original point, its position is given by $y = 0$. We apply the equations above once again. This time the known quantities are $y = 0$, $u_y = 80$ ft/sec, $a_y = -32$ ft/sec^2. The unknown quantities are v_y and t, and once again we have two equations in two unknowns. Making the

necessary substitutions,

$$0 = 80\,\frac{\text{ft}}{\text{sec}} \times t - \tfrac{1}{2} \times 32\,\frac{\text{ft}}{\text{sec}^2} \times t^2,$$

$$v_y = 80\,\frac{\text{ft}}{\text{sec}} - 32\,\frac{\text{ft}}{\text{sec}^2} \times t.$$

From the first equation the solutions are $t = 0$ and $t = 5$ sec; that is, the ball is at $y = 0$ at both of these times. The first of the solutions is trivial; the second one shows that the ball will be back at its starting position 5 sec after it is projected. Substituting the value into the second equation, we find $v_y = -80$ ft/sec. Thus the ball returns to its initial position with exactly the same speed it had when it started, but in the opposite direction.

(c) and (d) To find the position and velocity of the ball after 6 sec of motion, we again substitute appropriate values into our two equations. This time the known quantities are $t = 6$ sec, $u_y = 80$ ft/sec, $a_y = -32$ ft/sec^2, and the unknown quantities are y and v_y. Once again we have two equations in two unknowns, and a complete algebraic solution is possible. Substituting, we find that

$$y = 80\,\frac{\text{ft}}{\text{sec}} \times 6\text{ sec} - \tfrac{1}{2} \times 32\,\frac{\text{ft}}{\text{sec}^2} \times (6\text{ sec})^2,$$

$$v_y = 80\,\frac{\text{ft}}{\text{sec}} - 32\,\frac{\text{ft}}{\text{sec}^2} \times 6\text{ sec}.$$

The first equation immediately yields $y = -96$ ft, and the second equation gives $v_y = -112$ ft/sec. This is the same speed that a body would acquire if it fell from the highest point in the path, a distance of 196 ft.

2-9 Motion of a Projectile

The motion of a projectile after it leaves the muzzle of the gun is a special case of a freely falling body in which the initial velocity of the projectile is at any arbitrary direction to the vertical. We shall again limit our discussion to the ideal case in which the air resistance is neglected. Suppose that a bullet is fired horizontally with an initial velocity **u**, as illustrated in Figure 2-9. Once the bullet leaves the muzzle of the gun, its acceleration is vertically downward and is equal to **g** (directed vertically downward). This means that, in addition to its horizontal motion with velocity **u**, the bullet will acquire an additional velocity vertically downward equal to $\mathbf{g}t$; that is, the downward velocity will increase with the time just as if it were dropped. In fact, if a second bullet were dropped at the same time that the first bullet was fired, both would reach the ground at exactly the same time. The actual velocity **v** at any instant will be the vector sum of these two velocities; that is,

$$\mathbf{v} = \mathbf{u} + \mathbf{g}t, \tag{2-28}$$

as shown in Figure 2-9.

To determine the path of the bullet, let us choose a set of rectangular coordinates with the origin O situated at the muzzle of the gun and choose the x axis in the horizontal direction. Remembering that there is no acceleration in the x direction, we find from Equation (2-27b) that

$$x = ut; \tag{2-29}$$

that is, the motion of the projectile in the x direction is one with constant velocity. Since the initial velocity has no component in the y direction

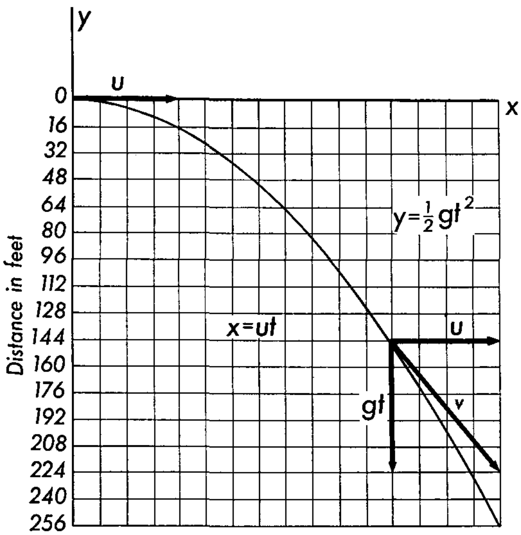


Fig. 2-9 Path of a bullet fired horizontally is a parabola.

$u_y = 0$. The acceleration is entirely in the y direction with $a_y = -g$. We rewrite Equation (2-27b) with the symbol y replacing each x to get the general equation appropriate to vertical motion, and substitute values appropriate to the problem to find

$$y = -\tfrac{1}{2}gt^2. \tag{2-30}$$

These values of x and y are plotted in Figure 2-9, and the curve obtained shows the path of the bullet. To obtain the equation of the path, we eliminate t from the two equations displayed above to obtain

$$y = \frac{-g}{2u^2}x^2. \tag{2-31}$$

This is the equation for the path, which is a parabola.

Illustrative Example. A bomber flying eastward with a velocity of 480 mi/hr drops a bomb from an elevation of 1,600 ft. Assuming that we can neglect

air resistance, determine where the bomb will land and how long it will take to get there.

Since the bomb was in the airplane until the instant it was released, its initial velocity is the same as that of the airplane; that is, $u_x = 480$ mi/hr = 704 ft/sec. We can obtain the time of fall for the bomb from Equation (2-30). Substituting $y = -1{,}600$ ft, we get

$$-1{,}600 \text{ ft} = -\tfrac{1}{2} \times 32 \frac{\text{ft}}{\text{sec}^2} \times t^2,$$

from which

$$t = 10 \text{ sec}.$$

Using Equation (2-29), we can find where it will strike the ground 10 sec after it was released:

$$x = 704 \frac{\text{ft}}{\text{sec}} \times 10 \text{ sec} = 7{,}040 \text{ ft}.$$

The bomb will strike at a distance of 7,040 ft east of the point at which it was released. If the plane continues to move with a velocity of 480 mi/hr after dropping the bomb, the latter will strike the target when the plane is directly overhead.

The velocity with which the bomb will strike the ground will be the vector sum of the horizontal velocity

$$u = 704 \frac{\text{ft}}{\text{sec}}$$

and the vertical velocity

$$gt = 320 \text{ ft/sec}.$$

Hence

$$v = \sqrt{(704)^2 + (320)^2} \frac{\text{ft}}{\text{sec}} = 772 \frac{\text{ft}}{\text{sec}}$$

directed at an angle of about 22° with the horizontal.

While a projectile may be fired in any direction, the only acceleration it experiences is vertical. The projectile follows a parabolic path; hence the motion lies entirely within a plane determined by the vertical and the direction of the barrel of the gun. We can describe the motion most conveniently in two dimensions, considering the y direction as vertical and the x direction as the direction of the horizontal projection of the gun barrel. A convenient origin of coordinates is the mouth of the gun barrel. The path of the projectile is shown in Figure 2-10. If the projectile is fired with an initial velocity u, at an angle of elevation θ with the horizontal, the x component of the initial velocity vector will be

$$u_x = u \cos \theta,$$

and the y component of the initial velocity vector will be

$$u_y = u \sin \theta.$$

The motion may be completely described from Equations (2-27a) and

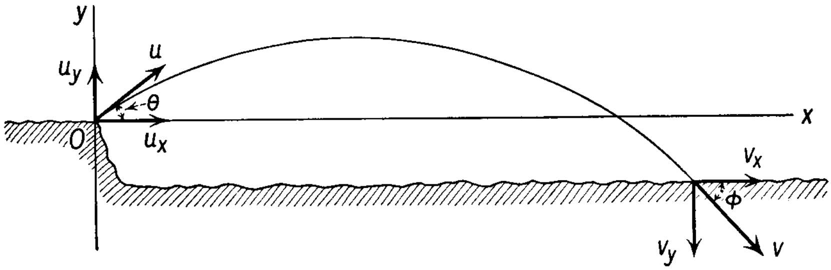


Fig. 2-10 Projectile fired from a cliff.

(2-27b), giving the x and y positions and velocity as functions of time. Rewriting these equations under the conditions of the problem, with $u_x = u\cos\theta$, $u_y = u\sin\theta$, $a_x = 0$, $a_y = -g$, we have

x position: $$x = ut\cos\theta; \tag{2-32a}$$

x velocity: $$v_x = u\cos\theta; \tag{2-32b}$$

y position: $$y = ut\sin\theta - \tfrac{1}{2}gt^2; \tag{2-32c}$$

y velocity: $$v_y = u\sin\theta - gt. \tag{2-32d}$$

To find the magnitude and direction of the velocity vector **v** at the time **t**, we utilize the techniques of Chapter 1:

magnitude of **v**: $$v = (v_x^2 + v_y^2)^{1/2}; \tag{2-32e}$$

direction of **v**: $$\phi = \text{arc tan}\,\frac{v_y}{v_x}; \tag{2-32f}$$

where ϕ is the angle between the velocity vector and the positive x axis.

Illustrative Example. A gun located on a cliff 160 ft high fires a shell with a muzzle velocity of 1,600 ft/sec at an angle of elevation of 37°, as in Figure 2-10. Find the time of flight, the horizontal distance the shell will travel, and the velocity with which the shell will strike the ground.

From the statement of the problem, the shell will strike the ground at a point where $y = -160$ ft. Other known quantities are $u = 1{,}600$ ft/sec, $\theta = 37°$. The unknown quantities are v_x, v_y, x, and t. Since there are four unknowns, we require four independent equations for the solution of the problem. Equations (2–32a to d) fulfill the requirements of the problem. With four equations and four unknowns the problem has been reduced to algebra. Putting numerical values in Equation (2-32a) yields

$$x = 1{,}600\,\frac{\text{ft}}{\text{sec}} \times 0.8 \times t;$$

Equation (2-32b) yields

$$v_x = 1{,}600\,\frac{\text{ft}}{\text{sec}} \times 0.8 = 1{,}280\,\frac{\text{ft}}{\text{sec}};$$

Equation (2-32c) yields

$$-160\text{ ft} = 1{,}600\,\frac{\text{ft}}{\text{sec}} \times 0.6t - \tfrac{1}{2} \times 32\,\frac{\text{ft}}{\text{sec}^2} \times t^2;$$

and Equation (2-32d) yields

$$v_y = 1{,}600\,\frac{\text{ft}}{\text{sec}} \times 0.6 - 32\,\frac{\text{ft}}{\text{sec}^2} \times t.$$

Solving Equation (2-32c) for t, we get

$$t^2 - 60t - 10 = 0,$$

so that

$$t = \frac{60 \pm \sqrt{3{,}600 + 40}}{2},$$

from which

$$t = 60.2\text{ sec}.$$

The negative value of t obtained in this solution has been discarded as physically meaningless. With this result, the other parts of the problem yield

$$x = 1{,}600\,\frac{\text{ft}}{\text{sec}} \times 0.8 \times 60.2\text{ sec} = 77{,}000\text{ ft};$$

$$v_y = 1{,}600\,\frac{\text{ft}}{\text{sec}} \times 0.6 - 32\,\frac{\text{ft}}{\text{sec}^2} \times 60.2\text{ sec} = -966\,\frac{\text{ft}}{\text{sec}}\cdot$$

From Equation (2-32e) we find

$$\begin{aligned} v^2 = v_x^2 + v_y^2 &= \left(1{,}280\,\frac{\text{ft}}{\text{sec}}\right)^2 + \left(-966\,\frac{\text{ft}}{\text{sec}}\right)^2 \\ &= (1.28 \times 10^3)^2 + (0.966 \times 10^3)^2 \\ &= 2.57 \times 10^6; \end{aligned}$$

$$v = 1.60 \times 10^3\,\frac{\text{ft}}{\text{sec}} = 1{,}600\,\frac{\text{ft}}{\text{sec}}\cdot$$

From Equation (2-32f) we find

$$\begin{aligned} \phi &= \arctan\frac{v_y}{v_x} = \arctan\frac{-966}{1{,}280} \\ &= \arctan(-0.755) \\ &= -37.1°. \end{aligned}$$

The magnitude of **v** and its direction could have been predicted either from the properties of the parabola or from the fact that the shell is a freely falling body once it leaves the gun.

Problems

2-1. A car is driven over a measured mile in 1.5 min. Determine the speed of the car (a) in miles per hour and (b) in feet per second.

2-2. A river steamer can travel at the rate of 15 mi/hr in still water. How

long will the trip between two cities 60 mi apart take (a) downstream and (b) upstream, if the river current is 3 mi/hr? (c) What will be the average speed for the round trip?

2-3. A ferryboat which can sail at the rate of 8 mi/hr in still water travels straight across a river $\frac{1}{2}$ mi wide in which there is a current of 2 mi/hr. (a) What is the velocity of the ferryboat with respect to the shore? (b) How long does a trip take?

2-4. The first three runners in a 100-yd race were clocked in 9.5 sec, 10.0 sec, and 10.5 sec, respectively. (a) What was the average speed of each runner and (b) how far apart were the first and last runners when the first one reached the finish line?

2-5. An airplane heads due north with a velocity of 250 mi/hr. A west wind is blowing with a velocity of 40 mi/hr. What is the velocity of the airplane relative to the ground?

2-6. An airplane whose normal speed in still air is 260 mi/hr must travel due east. (a) What course must the aviator set for the plane when there is a steady southwest wind of 40 mi/hr? (b) How long will it take to travel 850 mi?

2-7. An automobile starting from rest acquires a speed of 40 mi/hr in 12 sec. What is its average acceleration?

2-8. How long will it take for a car, starting from rest, to acquire a speed of 60 mi/hr if its acceleration is 12 ft/sec^2?

2-9. The brakes are applied to the wheels of a locomotive when it is traveling at 70 mi/hr. It comes to rest 24 sec after the brakes are applied. What is its average acceleration?

2-10. An automobile which is traveling at a speed of 55 mi/hr must be brought to a stop within 150 ft. What is the minimum acceleration that must be given to the car to accomplish this?

2-11. An airplane taking off on a runway 1,200 ft long must acquire a speed of 80 mi/hr to get safely into the air. (a) What is the minimum safe acceleration for this airplane? (b) How long will it take for the airplane to acquire this speed when so accelerated?

2-12. A car approaching a turn in the road has its speed decreased from 50 mi/hr to 30 mi/hr while traversing a distance of 120 ft. (a) What was its acceleration and (b) how long did it take to traverse this distance?

2-13. A boy drops a stone from a bridge 80 ft above the water. (a) With what speed did the stone strike the water? (b) With what speed would the stone have struck the water if it had been thrown down with a speed of 24 ft/sec?

2-14. A boy throws a ball vertically upward and catches it 1 sec later. (a) How high up did the ball go? (b) With what speed was it thrown upward?

2-15. A boy throws a stone horizontally with a speed of 30 ft/sec from a cliff 256 ft high. (a) How long will it take the stone to strike the ground? (b) Where will the stone land? (c) With what velocity will the stone strike the ground?

2-16. A small block starting from rest takes 5 sec to slide down an inclined plane 80 cm long. (a) What was its acceleration and (b) with what speed did it reach the bottom of the incline?

2-17. Two horizontal wires are placed parallel to each other 100 cm apart, one directly above the other. A falling ball is clocked as it passes each of these

wires. If the time elapsed is 0.20 sec, determine the speed the ball had when it passed each wire.

2-18. Fighter planes fly at 35,000 ft elevation. What must be the muzzle velocity of an antiaircraft shell to reach this height, neglecting air resistance?

2-19. A rifle fires a bullet with a speed of 30,000 cm/sec. If the elevation of the rifle is 30° with the horizontal, determine (a) the range of the bullet on horizontal ground and (b) the velocity of the bullet when it reaches the ground.

2-20. A projectile is fired vertically upward with an initial velocity of 1,800 ft/sec. (a) How high does it rise? (b) What velocity will it have 5 sec after leaving the gun? (c) What is its altitude 5 sec after leaving the gun?

2-21. A car moving with a speed of 30 mi/hr reaches the top of a hill. As it goes down the hill, its speed increases to 45 mi/hr in 1.5 min. (a) What is the acceleration of the car and (b) what distance does it travel in this time?

2-22. A stone thrown horizontally from a hill takes 6 sec to reach the ground. Determine, in meters, the height of the hill.

2-23. A falling stone is seen to pass a window 2 m high in 0.3 sec. (a) Determine the average speed of the stone. (b) Determine the speed with which it reaches the level of the top of the window. (c) Determine the height above this point from which it fell.

2-24. The distance between two stop lights on a cross-town street is 800 ft. If the acceleration of a certain car, both positive and negative, is kept at 6 ft/sec^2, and if the speed limit on this street is 30 mi/hr, determine the minimum time to traverse this distance.

2-25. A ball is thrown a distance of 65 ft in 1.2 sec. Assuming that it was caught at the same level as it was thrown, (a) determine how high the ball rose in its path of motion. (b) With what velocity was the ball thrown?

2-26. A gun fires a shell with a velocity of 600 m/sec at an angle of 45° with the horizontal. Neglecting air resistance, (a) determine the range of this gun, (b) determine the maximum height reached by the shell, and (c) determine the time of flight of this shell on level ground.

2-27. Derive the equation for the range of a projectile fired on level ground, $R = \dfrac{u^2 \sin 2\theta}{g}$, where R is the range, θ is the angle of elevation, and u is the initial velocity. Show that the maximum range is achieved when $\theta = 45°$.

2-28. The x coordinate of an object moving along the x axis is given by the equation $x = 3 - 5t + 12t^2$ ft. Find the corresponding equation for the velocity and acceleration of the object at any time t.

2-29. A body moving in space has its motion described by the equations $x = 12t + 15$, $y = 6t^2$ where the distances are in meters and the time is given in seconds. Find the magnitude and direction of the velocity and the acceleration when $t = 3$ sec.

2-30. A ball is thrown toward a building 50 ft distant at a speed of 100 ft/sec. At what angle must it be thrown if it is to pass through a window 42 ft from the ground?

2-31. A railroad car is moving due north at a speed of 60 mi/hr. A ball is thrown from the window due east at an angle of elevation of 30° and a speed of 40 ft/sec. (a) Find the time at which it strikes the ground 10 ft below the window

of the car. (b) How far east of the track does the ball land? (c) How far north of the point of projection does the ball land?

2-32. By differentiation with respect to time, show that the equation

$$x = x_i + ut + \tfrac{1}{2}at^2$$

describes the position of a particle whose initial position is x_i and whose initial speed is u moving along the x axis with constant acceleration a.

2-33. When a balloon is at a height of 6400 ft and rising at a speed of 32 ft/sec, a stone is thrown vertically out of the balloon. The stone hits the ground directly below in 20 sec. (a) What was the initial velocity of the stone relative to the balloon? (b) Relative to the ground?

2-34. Motorist A, starting from rest, accelerates at a rate of 6 ft/sec^2. At the same time that A begins, motorist B, starting from rest at a point 100 ft ahead of A, accelerates at a rate of 4 ft/sec^2. (a) How far does motorist A travel before they meet? (b) At the instant they meet each motorist decelerates at the rate of 5 ft/sec^2 until his car comes to rest. How far apart are they when they have stopped?

2-35. A train is moving with uniform speed along a level road. A man on the observation platform drops a ball. What is the path of the ball as observed (a) by the man on the train and (b) by another person standing at a short distance from the tracks?

2-36. In a laboratory experiment an air rifle is clamped in position and aimed by sighting along the barrel. The target is released just as the bullet leaves the muzzle of the rifle. Show that the bullet will always hit the target.

2-37. If there is no wind, raindrops fall vertically with uniform speed. A man driving a car on a windless rainy day observes that the tracks left by the raindrops on the side windows are all inclined at the same angle. Show how the vertical speed of the raindrops can be determined from the inclination of the tracks and the reading of the speedometer.

2-38. Show that the speeds of a projectile are the same at any two points in its path which are at the same elevation.

2-39. A boy seated in a rapidly moving railroad car tosses a ball up into the air. Will the ball come down in front of him; behind him; into his hand? What will happen when the car is accelerating in the forward direction? Going round a curve?

3

The Equilibrium of a Particle

3-1 Equilibrium

A particle which remains at rest or in uniform motion with respect to its frame of reference is said to be in *equilibrium* in that frame. Centuries ago it was recognized that the state of rest was a *natural* state of things, for it was observed that objects set in motion on the surface of the earth tended to come to rest. The maintenance of any horizontal motion on earth was thought to require the continued exercise of a force, hence to be a *violent* motion, while vertical motion like that of a falling body was thought to be *natural*. In heavenly bodies circular motion was thought to be natural. That uniform motion in a straight line was a universal equilibrium condition, a natural state of things, was not recognized until the work of Galileo (1564–1642) and Newton (1642–1727), which represented a very significant contribution to the study of mechanics and to our understanding of nature.

Newton summarized his conception of motion in three principles, which are today called *Newton's laws of motion,* the first of which may be stated as follows: *A body at rest will remain at rest, and a body in motion will continue in motion with constant speed in a straight line, as long as no net force acts upon the body.*

Although, as the result of much repetition, Newton's first law may today seem to be another trite statement, the result of simple common sense, it was indeed a very startling conception. No one of us has ever seen an object which moved with constant speed in a straight line for an infinite length of time either on the earth or in the heavens. Nevertheless the Newtonian formulation of the conditions of equilibrium has proved itself invaluable in our understanding of nature and is universally accepted as the basis for the formulation of an important division of mechanics. The experimental validity of the Newtonian formulation of equilibrium is reestablished each time a new structure is erected, each time an airplane flies.

3-2 Equilibrium of a Particle

According to Newton's first law, a particle is said to be in equilibrium if there is no net force acting on it. This does not mean that no forces act on the particle, but rather that the resultant of all the forces which do act on the particle is zero. The direction in which a force acts is an important fact needed for its specification. Force is, therefore, a vector quantity, and the resultant of the forces must be obtained by vector methods. If we label the forces acting on the particle by **A**, **B**, **C**, . . . , the condition for the equilibrium of a particle may be written in the form of an equation as

$$\boxed{\mathbf{R} = \mathbf{A} + \mathbf{B} + \mathbf{C} + \cdots = 0,} \tag{3-1}$$

where **R** is the resultant of the forces acting on the particle.

We have already seen that a vector equation may be resolved into three independent scalar equations, one for each of three mutually perpendicular directions, so that Equation (3-1) may be written as

$$R_x = 0 = A_x + B_x + C_x + \cdots, \tag{3-2a}$$

$$R_y = 0 = A_y + B_y + C_y + \cdots, \tag{3-2b}$$

$$R_z = 0 = A_z + B_z + C_z + \cdots, \tag{3-2c}$$

in which A_x, B_x, C_x, . . . are the x components of the forces **A**, **B**, **C**, . . . , respectively, and A_y, B_y, C_y, . . . and A_z, B_z, C_z, . . . are the y and z components of these forces, respectively.

We shall usually restrict our discussion to the equilibrium of bodies which are acted upon by forces all of which are in one plane. In such cases Equation (3-2c) is redundant, and the conditions for the equilibrium of a particle become

$$\boxed{\begin{aligned} R_x &= 0, \\ R_y &= 0. \end{aligned}} \qquad \begin{aligned} &(3\text{-}2a) \\ &(3\text{-}2b) \end{aligned}$$

The word *particle* has been used rather loosely to imply a small body on which the forces act *concurrently;* that is, all forces are directed toward a single point. At times the statements relating to the statics of a particle will be applied to larger bodies when these move in *translational* motion; that is, when there is no rotation of the body itself.

When we examine the equilibrium of a particle, we must be careful to isolate the particle in our minds and to examine the forces exerted on the particle by each object capable of exerting a force on it. We shall see that the conditions for the equilibrium of a particle will enable us to determine

the forces exerted by many structural elements which make up a complicated assembly and thereby provide information essential to the design of complex structures.

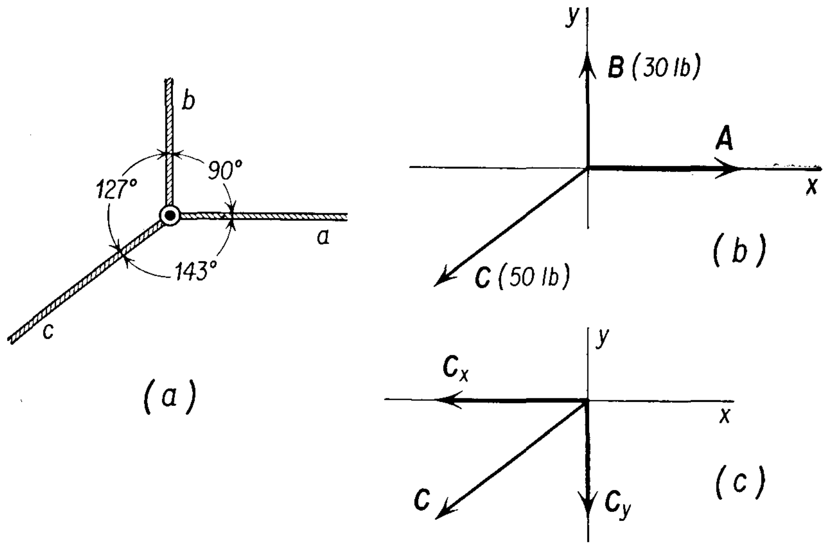


Fig. 3-1

Illustrative Example. Three strings are tied to a small ring (here considered to be a particle), as shown in Figure 3-1(a). The string b exerts a force of 30 lb upon the ring; the string c exerts a force of 50 lb upon the ring. What force must be exerted by the string a if the ring is to remain in equilibrium?

From experience we know that a taut string exerts a force directed along its own length. In Figure 3-1(b), the ring is placed at the origin of an x-y coordinate system, and the strings are replaced by the forces that they exert on the ring; the force of the string b is represented by the symbol **B**, and the force of string c is represented by the symbol **C**. While forces **B** and **C** are drawn to scale, the force **A** is not, since its value is not known.

To apply Equations (3-2) to the solution of the problem, we must resolve the force **C** into its x and y components C_x and C_y. From Figure 3-1(c) we see that

$$C_x = -C \cos 37° = -50 \text{ lb} \times 0.8 = -40 \text{ lb},$$

$$C_y = -C \cos 53° = -50 \text{ lb} \times 0.6 = -30 \text{ lb},$$

and, substituting numerical values into Equations (3-2), we have

$$R_x = A - 40 \text{ lb} = 0,$$

$$R_y = +30 \text{ lb} - 30 \text{ lb} = 0.$$

Thus the ring is in equilibrium in the y direction under the action of the applied forces and will be in equilibrium in the x direction if $R_x = 0$; that is, if $A = 40$ lb.

3-3 Newton's Third Law

Newton's third law of motion states that *whenever one body exerts a force on another, the second body exerts a force equal in magnitude and opposite in direction on the first body.* This law is sometimes called the *law of action and reaction.* In the solution of problems in statics, this concept is applied extensively. We focus our attention first on the equilibrium of one point in

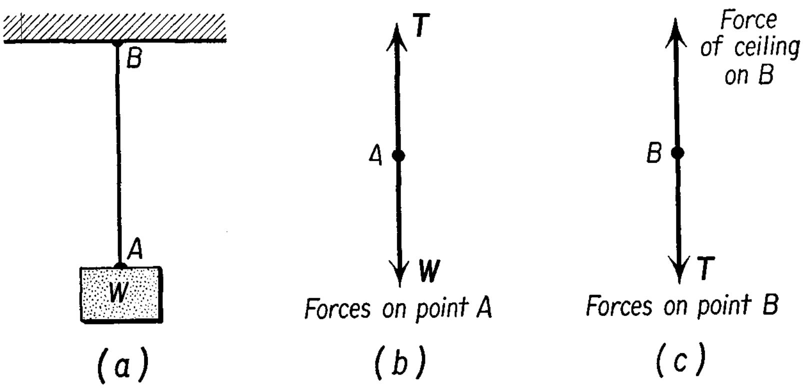


Fig. 3-2

the structure, then on a second, and a third, and so on, to find all the forces exerted on or by each member of the structure. If we pull against a rope, our hands exert a force on the rope, but the rope exerts an equal and opposite force on our hands. When a packing crate rests on the floor, the crate exerts a force on the floor and the floor exerts an equal and opposite force on the crate.

Consider the forces associated with a hanging weight. If a body of weight W is supported by a rope fastened to the ceiling of a room, as in Figure 3-2(a), we can determine the tension in the rope by considering the equilibrium of a particle of the rope at point A. To do this we draw a second diagram called a *force diagram,* to distinguish it from the sketch illustrating the appearance of the system which we call the *space diagram.* The forces on A are the tension $\mathbf{T}$ in the rope and the weight of the body, as shown in the force diagram, Figure 3-2(b). Since the force of gravity is vertically downward, the rope must be in the vertical direction. The force $\mathbf{T}$ at A must be equal in magnitude to $\mathbf{W}$ but opposite in direction.

To determine the force that the ceiling must exert on the rope, we apply the results obtained from examining the equilibrium of point A to the examination of the equilibrium of an element of the rope at point B. The forces on this element are the tension in the rope $\mathbf{T}$ which at B is directed

vertically downward, and the force of the ceiling of unknown magnitude and direction; see Figure 3-2(c). Since the element at B is in equilibrium, the sum of all forces acting on B must be zero, and the force exerted by the ceiling must be vertically upward, and in magnitude equal to T, hence equal to W.

3-4 Solution of Simple Problems in Statics

Illustrative Example. A *boom,* or a *strut,* whose weight can be neglected, is a typical example of a pinned beam; that is, the beam is connected to the other part of the structure by means of a strong pin or rod which passes freely through it. A beam at rest which is pinned at one end can exert a force either in tension or in compression (that is, a push or a pull), but this force too can only be exerted in a direction parallel to the beam, for if the beam is pinned and can rotate freely,

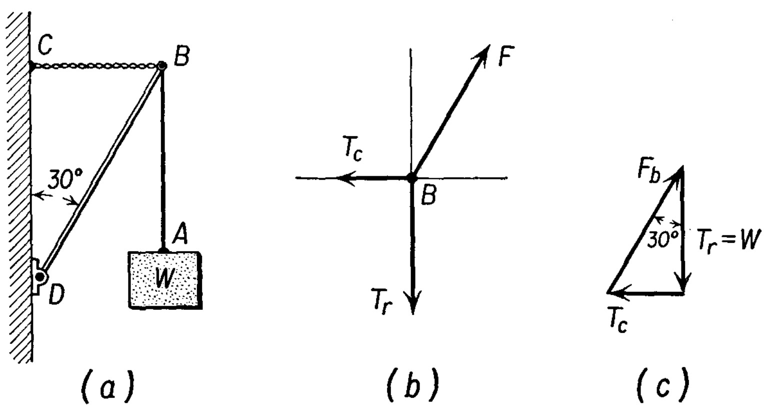


Fig. 3-3

any force perpendicular to its own length would cause it to rotate. Suppose a weight $\mathbf{W}$ of 60 lb hangs from a rope supported at the outer end of a boom of negligible weight, as in Figure 3-3(a). The boom is pinned at its lower end and is supported at its upper end by a horizontal chain fastened to the wall. To determine the forces exerted by each member, we first isolate a small segment of the rope near point A; as we have already seen, the tension in the rope $\mathbf{T}_r$ must be equal in magnitude to the weight W. To examine the equilibrium at point B, we draw a force diagram as in Figure 3-3(b), with the forces acting on the point B drawn as though they radiate out from a common origin. Since both the rope and chain can exert only tension forces along their own lengths, we label these forces $\mathbf{T}_r$ and $\mathbf{T}_c$ (for the tension in the chain), respectively, and draw them in appropriate directions. A pinned beam can exert a force only along its own length. The beam is in compression, and the force exerted by the beam must be a thrust away from the wall. The force exerted by the beam is $\mathbf{F}$, as indicated in the

figure. Since the point B is in equilibrium, the vector sum of all these forces acting on B must be zero.

Solving Equation (3-1) by the polygon method, we observe that the three force vectors, added in sequence head to tail, must form a closed triangle, for the resultant of the three forces, the vector drawn from the tail of the first to the head of the last, must be of length zero. The force vectors are parallel to their respective structural members; appropriate angles may therefore be identified on the force diagram from the space diagram; thus $\mathbf{F}$ makes an angle of 30° with the vertical force $\mathbf{T}_r$, and the vector $\mathbf{T}_r$ is perpendicular to $\mathbf{T}_c$. With this information the methods of trigonometry enable us to solve the problem. From Figure 3-3(c) we have

$$\frac{T_r}{F} = \cos 30°;$$

thus

$$F = \frac{T_r}{\cos 30°} = \frac{60\text{ lb}}{0.866} = 69.4\text{ lb}.$$

Furthermore,

$$\frac{T_c}{T_r} = \tan 30°;$$

thus

$$T_c = T_r \tan 30° = 60\text{ lb} \times 0.577 = 34.6\text{ lb}.$$

We have found the forces exerted by the three structural elements and have therefore obtained a complete solution to the problem. Since, by Newton's third law, the force exerted by each of these elements is equal and opposite to the force exerted on it, knowledge of the strength of materials enables the engineer to design a suitable structure.

Illustrative Example. A string is passed over two pulleys, and weights of 30 lb and 40 lb are hung from the ends. When a weight of 50 lb is hung on the string anywhere between the two pulleys, it is found that the angle made by the two parts of the string supporting the 50-lb weight is 90°, as shown in Figure 3-4(a), no matter where the weight is placed. The angle does not change when the pulleys are raised or lowered with respect to each other. Explain.

Examination of the equilibrium of the element of string at point A shows that the tension in the string at this point must be 30 lb. *The tension in the string is not altered when the string passes over a frictionless pulley.* A frictionless pulley is one in which the bearing of the pulley is perfectly smooth, although the surface of the pulley wheel may be quite rough. The tension in the string $\overline{BC}$ must be 30 lb. Similarly, the tension in the string $\overline{CD}$ must be 40 lb. At the point C there is an abrupt change in the direction of the string, as there must be when the tension is different in two parts of a string.

Consider the equilibrium of a particle at the point C. Three forces act on point C: namely, the two tensions in the strings and the weight of 50 lb. Since the point C is in equilibrium, the vector sum of these forces must equal zero. The force diagram is shown in Figure 3-4(b). The force polygon, drawn as Figure 3-4(c), is a 3–4–5 triangle, hence a right triangle, with the right angle between the 30-lb vector and the 40-lb vector. The directions $\overline{BC}$ and $\overline{CD}$ are parallel to the

30-lb and the 40-lb vectors, respectively. By a theorem in geometry which states that two angles whose sides are mutually perpendicular are either equal or supplementary, the angle BCD must be 90°.

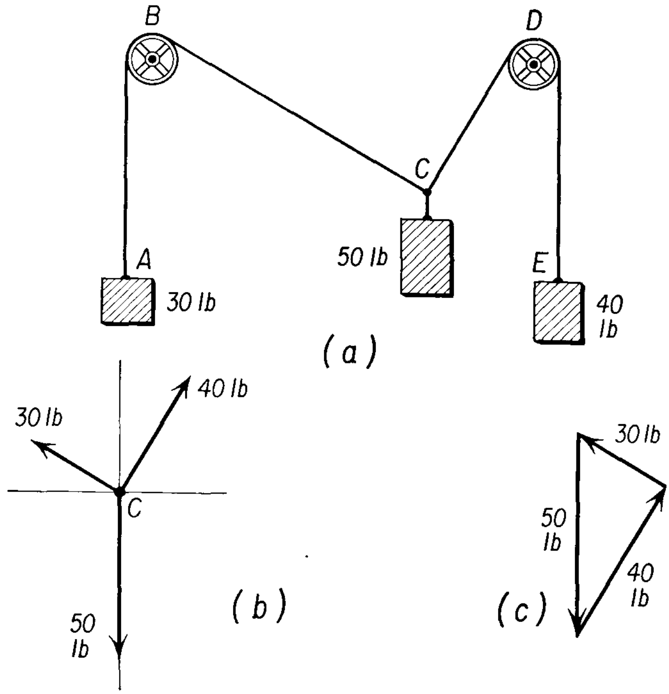


Fig. 3-4

We note again that the force diagram and the space diagram are not drawn to the same scale; distances on the force diagram do not necessarily transfer to the space diagram. Angles from the space diagram may be identified on the force diagram, because structural members often exert forces which bear a simple relationship to the directions of these members.

3-5 Frictional Forces

It is common experience that moving objects on the surface of the earth do not continue to move in a straight line with uniform speed, and that a force must be applied to maintain uniform motion. As part of our attempt to develop a consistent picture of nature, we have developed the concept of friction. The frictional force is conceived as a force that opposes the motion; this force must be equal in magnitude to the applied force required to keep a body sliding over a surface with uniform speed.

Friction between solid bodies can be classified into two types, *sliding friction* and *rolling friction.* Sliding friction occurs whenever the surface of one body slides over another; this kind of friction exists between the brake lining and the brake drum of the braking mechanism on the wheels of a car. Rolling friction exists between the wheels of a car and the road when no slippage occurs. It is common experience that considerably more effort is required to start a heavy object sliding across a floor than is needed to keep it moving once it has been started. We may thus distinguish between the force of *static* friction required to start an object in motion and the force of *kinetic* friction required to keep an object in motion, for they have different magnitudes. The results of many experiments show that, to a good approximation, the magnitude of the frictional force does not depend on the area of contact between the two surfaces but depends upon the nature of the surfaces and upon the force pressing them together. Calling the magnitude of the frictional force F_r and the magnitude of the normal force with which one surface presses against another N, we write

$$\boxed{F_r = fN,} \tag{3-3}$$

were f is a dimensionless constant of proportionality called the *coefficient of friction.* Some typical values of the coefficient of kinetic or sliding friction are shown in Table 3-1. While values in the table are all less than 1, there is no fundamental reason why the coefficient of friction should not have a value greater than 1. The coefficient of static friction is higher than the coefficient of kinetic friction.

TABLE 3-1 COEFFICIENTS OF SLIDING OR KINETIC FRICTION

Materials	
Wood on wood, dry	0.25–0.50
Metal on oak, dry	0.50
Leather on metal, dry	0.56
Metal on metal, dry	0.15–0.20
Steel on agate, dry	0.20
Masonry on clay, dry	0.51

The frictional force always acts in such a direction as to oppose the motion of one surface relative to another. When an object is in motion, the force of kinetic friction given by Equation (3-3) is always present. This is not true for static friction. The coefficient of static friction gives the maximum value of the frictional force—the force which must be applied to start the motion. As long as the object is at rest, the frictional

force may take on any value (up to that maximum) necessary to fulfill the conditions of equilibrium (see the second *Illustrative Example* of Section 3-6 and Figure 3-6).

In screw-thread fastening devices, the effect of tightening a bolt against the work, or against a spring washer, is to increase the normal force with which the thread surfaces of the nut bear against the thread surfaces of the bolt. This increases the frictional force between the nut and bolt, as shown by Equation (3-3), and helps keep the nut from untwisting.

Illustrative Example. A steel block weighing 175 lb is pulled horizontally with uniform speed over another steel block. If the coefficient of kinetic friction between the two surfaces is 0.20, determine the force of friction between them.

The force pressing the two surfaces together is the weight of the upper block; hence $N = 175$ lb. From Equation (3-3), the force of friction is

$$F_r = fN,$$

$$F_r = 0.2 \times 175 \text{ lb} = 35 \text{ lb}.$$

3-6 Body on an Inclined Plane

When a body rests upon an inclined plane, it is sometimes convenient to resolve the forces acting on the body into components parallel and perpendicular to the plane. The force exerted by a plane surface may similarly be resolved into a component parallel to the plane, which is called the frictional force, and a component perpendicular to the plane, called the normal force. In describing a surface by the word *smooth*, we imply that the surface is frictionless and is capable of exerting only a force normal to itself. When it is called a *rough* surface, we imply that it is capable of exerting a force in any outward direction, and the component of that force parallel to the surface is the frictional force.

Illustrative Example. A weight of 80 lb rests on a smooth plane which is inclined at an angle of 37° with the horizontal, as shown in Figure 3-5(a). What is the magnitude of the horizontal force **F** which will keep the block from sliding down the plane?

The forces acting on the body are shown in Figure 3-5(b); they are the horizontal forces **F**, the force of gravity **W**, and the force of the smooth plane **N**, which must be normal to the plane. The vector sum of **W**, **F**, and **N** is zero, hence these vectors must form a closed triangle, as shown in Figure 3-5(c). The angle between **W** and **N** in Figure 3-5(c) is equal to the angle made by the inclined plane with the horizontal and is 37°. Since a 37° right triangle is approximately a 3–4–5 right triangle, and $W = 80$ lb, we see that $F = 60$ lb and $N = 100$ lb. Thus a horizontal force of 60 lb will keep a body of weight 80 lb from sliding down a smooth inclined plane. According to Newton's first law, such a force will also keep the body sliding up or down the plane at uniform speed once the body

has achieved that speed, for under the application of a 60-lb horizontal force the vector sum of the applied forces is equal to zero.

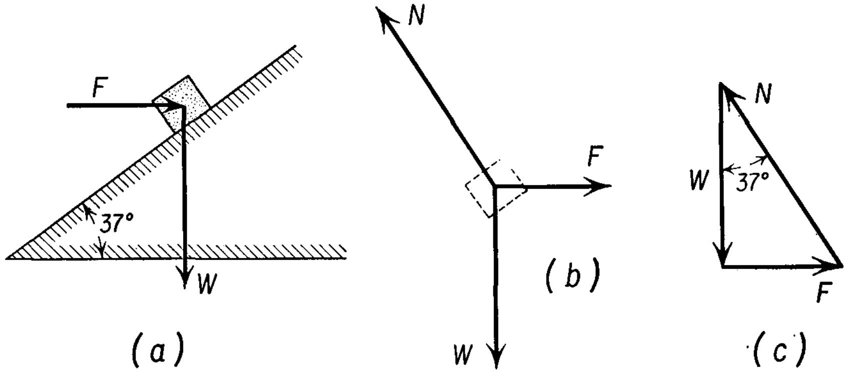


Fig. 3-5

Illustrative Example. A steel block weighing 100 lb rests upon a plank which is inclined at an angle θ of 30° with the horizontal, as shown in Figure 3-6(a). The coefficient of static friction is 0.8. What is the frictional force between the block and the plank?

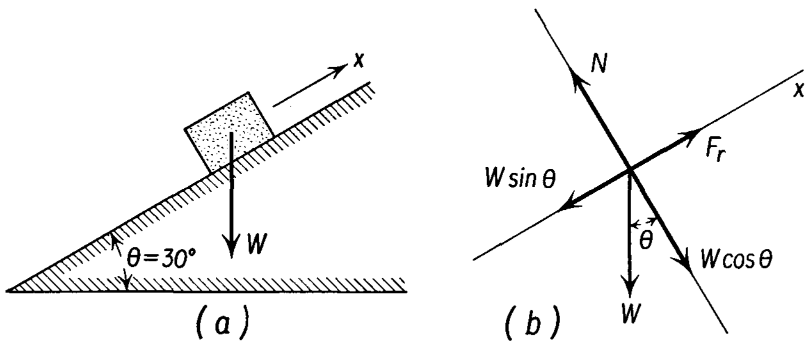


Fig. 3-6

It is convenient to choose the direction of the x axis as parallel to the plank and the direction of the y axis as perpendicular to the plank. The forces acting on the block are shown in Figure 3-6(b); they are the gravitational force $\mathbf{W} = 100$ lb acting vertically downward, the force exerted by the plank on the block which is resolved into the force $\mathbf{N}$ normal to the plank, and the frictional force $\mathbf{F}_r$ parallel to the plank. We choose the direction of $\mathbf{F}_r$ in order to oppose the tendency of the block to slide down the plank. The weight of the block is also resolved into x and y components; the y component is $W \cos \theta = -86.6$ lb, and the x component is $W \sin \theta = -50$ lb. Applying Equation (3-2b) for equilibrium in the y

direction, we have

$$N - 86.6 \text{ lb} = 0,$$
$$N = 86.6 \text{ lb}.$$

From Equation (3-3) the maximum value of the frictional force is

$$F_r\,(\text{max}) = 0.8 \times 86.6 = 69.3 \text{ lb}.$$

But from Equation (3-2a) the conditions for the equilibrium of the block require the magnitude of the frictional force given by

$$F_r - 50 \text{ lb} = 0,$$

so that

$$F_r = 50 \text{ lb}.$$

Hence the frictional force is 50 lb, which is less than the maximum value obtained from the coefficient of static friction.

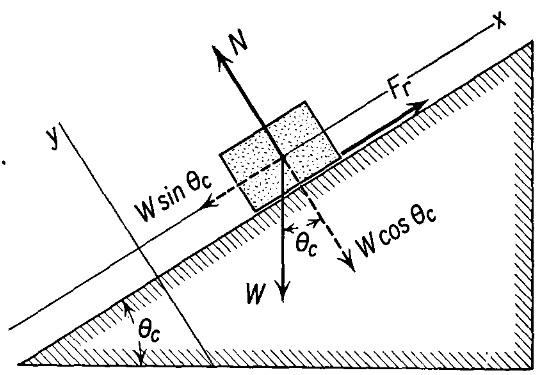


Fig. 3-7 Angle of repose on a rough inclined plane.

If the coefficient of kinetic friction between a block of weight W and a plane is given by f, it is interesting to consider the angle of inclination of the plane with the horizontal at which the block will continue to slide down the plane with uniform speed. Let us call this angle θ_c.

Once started in motion down the plane, the block will slide down with uniform speed if the sum of the forces parallel to the plane is equal to zero. Referring to Figure 3-7, the magnitude of the y component of the weight of the block is given by $W \cos \theta_c$, and this is equal to the magnitude of the normal force N. The frictional force is given by

$$F_r = fN = fW \cos \theta_c.$$

Since the sum of the x components of the forces must be equal to zero, we have

$$-W \sin \theta_c + F_r = 0,$$

and, substituting for F_r, we get

$$W \sin \theta_c = fW \cos \theta_c,$$

so that

$$f = \tan \theta_c. \tag{3-4}$$

The coefficient of sliding friction is given by the tangent of the angle at which the block, if started, will slide down the plane with uniform speed. At any angle slightly less than $\theta_c = \text{arc} \tan f$, the block will come to rest. The angle θ_c is called the *angle of repose*. Considerations similar to these illustrate why a pile of coal has uniformly sloping sides, and why some materials will stand in steeper piles than others. A knowledge of the *angle of repose*, the angle of elevation of the surface of the pile of granular materials, is of practical value in the design of appropriate storage bins.

Problems

3-1. A body weighing 15 lb hangs from one end of a vertical cord. What is the tension in the cord?

3-2. A body weighing 35 lb is hung from the center of a cord. The angle between the two parts of the cord is 120°. Determine the tension in the cord.

3-3. A body weighing 120 lb hangs from a cord which is attached to the ceiling. A horizontal force pushes the body out so that the cord makes an angle of 30° with the vertical. Determine (a) the magnitude of the horizontal force and (b) of the tension in the cord.

3-4. A rope 20 ft long has its ends fastened to the tops of two poles 16 ft apart. A weight of 240 lb hangs 8 ft from one end of the rope. Determine the tension in each section of the rope.

3-5. In order to pull a car out of a rut, a man ties a rope around a tree and attaches the other end to the front bumper of the car. The man then pulls on the middle of the rope in a direction at right angles to the line from the tree to the car. (a) Determine the tension in the rope if the man exerts a force of 80 lb when the angle between the two parts of the rope is 160°. (b) What force does the rope exert on the car?

3-6. A box weighing 70 lb is held up by two ropes, one of which makes an angle of 30° with the vertical while the other makes an angle of 60° with the vertical. Find the tension in each rope.

3-7. A boom in the form of a uniform pole, whose weight may be neglected, is hinged at its lower end. The boom is held at an angle of 60° with the ground by means of a horizontal cable attached to its upper end. (a) Determine the tension in the cable when a load of 1,000 lb is attached to the upper end. (b) Determine the thrust exerted by the boom.

3-8. One end A of a rigid, light, horizontal bar is attached to a wall, while the other end C is supported by a rope which is attached to a point D on the wall directly above A, as shown in Figure 3-8. The length of the bar AC is 12 ft, and the length of the rope CD is 13 ft. Determine (a) the tension in the rope and (b) the thrust exerted by the bar when a weight of 3,000 lb is hung from C.

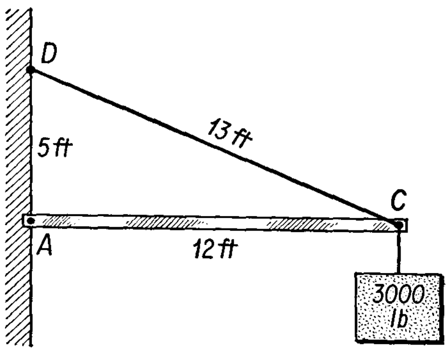


Fig. 3-8

3-9. A car weighing 3,500 lb is on a hill which rises 5 ft for every 100 ft of length. Determine the component of the weight which acts parallel to the hill.

3-10. A crate weighing 150 lb is held on a smooth inclined plane by means of a rope tied to this crate and to the top of the plane. If the inclination of the plane to the horizontal is 30°, (a) determine the tension in the rope and (b) the push of the plane against the crate. (c) What will be the tension in the rope if it is used to pull the crate up the plane at a uniform rate? (d) Determine the tension in the rope if the crate is allowed to slide down the plane at a uniform rate.

3-11. A barrel weighing 120 lb is held on a smooth inclined plane by means of a force applied horizontally. The inclination of the plane is 37°. Determine (a) the magnitude of the horizontal force and (b) the push of the plane.

3-12. A block weighing 2 lb is dragged along a rough, level floor at uniform speed by a rope which makes an angle of 30° with the floor. If the coefficient of kinetic friction between the floor and the block is 0.3, find the tension in the rope.

3-13. In Problem 3-10 the coefficient of static friction between the plane and the crate is 0.2 and the coefficient of kinetic friction is 0.1. Recalculate your answers to parts (a), (b), (c), and (d).

3-14. (a) A block weighing 50 lb rests on a horizontal plane. Find the frictional force between the block and the plane. The coefficient of static sliding friction is 0.8. (b) What is the frictional force between the block and the plane when the plane is inclined at an angle of 30° with the horizontal?

3-15. A box weighing 100 lb is pushed at constant speed, up a rough plane inclined at an angle of 37° with the horizontal by a steady horizontal force of 85 lb. (a) Find the frictional force and (b) the coefficient of kinetic sliding friction between the block and the plane. (c) Find the horizontal force which must be applied to lower the block down the plane.

3-16. A steeple jack sits in a chair which is fastened to a long rope. The rope is passed over a pulley fixed at the top of the steeple and hangs down within reach of the steeple jack. If the steeple jack and chair weigh 150 lb, with what force must he pull on the free end of the rope to raise himself at a steady rate? Neglect the weight of the rope.

4

Statics of a Rigid Body

4-1 The Concept of a Rigid Body

In the preceding chapter we observed that a particle would remain in equilibrium, in a state of rest, or in a state of uniform motion in a straight line when the resultant of all the forces acting on it was equal to zero. This condition for equilibrium was extended to larger bodies under either of two possible conditions: If the forces acting on the body were concurrent, that is, if they were directed toward a single point, the body could be treated as if it were a particle; or if the body moved with uniform translational motion in which every particle of the body moved in the same fixed direction with uniform speed, the whole body could be treated as though it were a particle.

Many of the problems of the equilibrium of extended bodies do not fulfill these conditions. The forces acting on the body do not pass through a single point, and the motion of the body is not one of uniform translational motion but may include rotation as well. The motion of a body is often quite complicated, as in the case of a spiraling football. The ball is generally thrown so that it spins about its longer axis, but, in addition to its spinning motion, the axis of rotation itself rotates, and the ball has a general translational projectilelike motion superimposed upon the rotational motions.

We shall restrict ourselves to the study of rotation about a fixed axis and shall devote our attention first to the case of equilibrium of a body with respect to rotation about a fixed axis.

While all material bodies deform somewhat under the action of applied forces, it is convenient to think of them as nondeforming, or as rigid; we shall define a *rigid body as one in which all dimensions remain the same, regardless of the nature of the applied forces.* With this concept the statics of material bodies can be greatly simplified, for, instead of having to study the body as though it were a vast collection of particles to which the conditions of equilibrium must be applied to one particle at a time, the entire body

may be treated as a single object, and its equilibrium may be studied through the introduction of a new concept called *torque.*

4-2 Moment of a Force; Torque

The effect of a force in producing rotation is determined by two factors, (a) the force itself and (b) the distance of the line of action of the force from some line considered as an *axis of rotation.* Suppose that a force **F** acts on a rigid body, as shown in Figure 4-1; its line of action is collinear with the

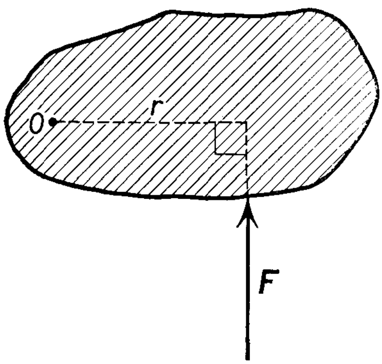


Fig. 4-1 Torque produced by a force **F** whose line of action is at a distance r from the axis through O perpendicular to the plane of the paper is $G = Fr$.

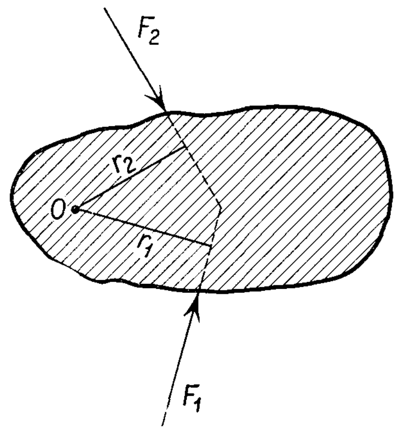


Fig. 4-2

vector **F**. Imagine an axis through point O perpendicular to the plane of the paper such that the distance from O to the line of action of the force **F** is r. The effect of the force in producing rotation about the axis through O, called the *moment of the force*, or the *torque*, is defined as the *product of the force and the perpendicular distance from the axis to the line of action of the force.* If G represents the magnitude of the torque, then

$$\boxed{G = Fr.} \tag{4-1}$$

As we view Figure 4-1, the torque will tend to produce a rotation of the body in a counterclockwise direction about an axis through O; the torque G is said to be in a counterclockwise direction. Figure 4-2 shows a rigid body acted upon by two forces $\mathbf{F}_1$ and $\mathbf{F}_2$ at distances r_1 and r_2, respectively,

from an axis through O perpendicular to the plane of the paper. The torque produced by $\mathbf{F}_1$ about O is F_1r_1 in a counterclockwise direction; the torque produced by $\mathbf{F}_2$ about O is F_2r_2 in a clockwise direction. By convention *a torque in a counterclockwise direction is usually called positive,* and one in a *clockwise direction is usually called negative.* Thus the total torque produced by these forces about O as an axis is

$$G = F_1r_1 - F_2r_2.$$

Whenever the torque produced by a force about a particular axis is to be determined, it is essential to find the *perpendicular distance* from the axis to the line of action of the force. In Figure 4-3 the force $\mathbf{F}$ is applied at the point E on the rim of a disk. To find the torque about an axis perpendicular

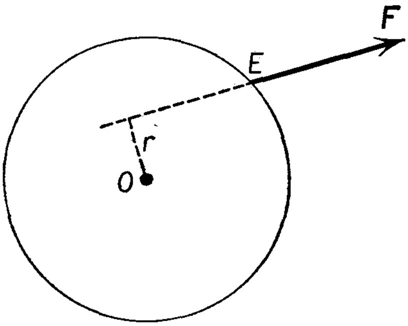


Fig. 4-3 Force $\mathbf{F}$ applied at point E produces torque $-Fr$ about an axis through O perpendicular to the plane of the paper.

to the plane of the paper through O at the center of the disk, it is necessary to extend the line of action of the force $\mathbf{F}$ as shown by the dotted line, and then to drop a perpendicular from O onto this line to obtain the perpendicular distance r. The torque of $\mathbf{F}$ about the axis through O is $-Fr$, the minus sign indicating that it acts in a clockwise direction.

The units used for expressing a torque must be those appropriate for the product of a force and a distance. Thus pound feet (lb ft), newton meters (nt m), and dyne centimeters (dyne cm) are the appropriate units of torque in the British gravitational, the mks, and the cgs systems of units, respectively.

4-3 Vector Representation of Torque

Only coplanar forces were considered in the above discussion; the axis about which the moments of the forces were taken was always at right angles to the plane containing the forces. In this simple case the direction of rotation, and hence the direction of the torque, was specified either as clockwise or counterclockwise. In the more general case where the forces are not coplanar and the axis of rotation may have any arbitrary direction,

it is necessary to have a more general method of representing torque as a vector.

As we have already seen, conventional rectangular coordinate systems are *right-handed;* that is, if the fingers of the right hand are pointed in the direction of the x axis and the fingers are bent so that they point toward the direction of the y axis, the outstretched thumb will point in the direction of the z axis. The disposition of the fingers and thumb of the right hand are commonly used to represent vector quantities involving rotation. If the fingers of the right hand were used to grasp the disk illustrated in Figure 4-4, with the fingers pointing in the direction of the rotation which the force at A might produce, the extended thumb would point in the direction of the axis of rotation. To represent the torque produced by the force **F** at A by

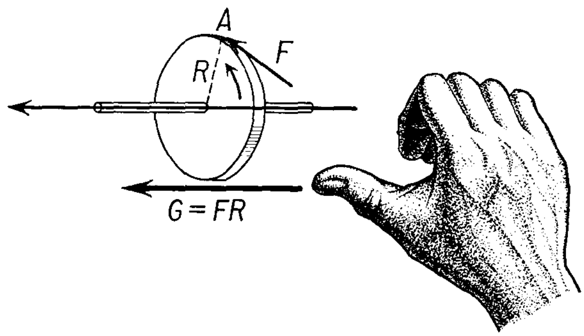


Fig. 4-4 The right hand rule: if the fingers of the right hand follow the direction of rotation, the thumb will point in the direction in which the arrow should be drawn along the axis of rotation.

a vector, we would draw a vector of magnitude given by $G = R \times F$ pointing along the line of the axis of rotation to the left. Conversely, if the torque vector were given as to the left, then, pointing the right thumb in the direction of the vector, the curled fingers of the right hand would point in the direction of rotation the torque would tend to produce.

4-4 Equilibrium of a Rigid Body

When a rigid body remains at rest under the action of a system of forces, the body is said to be in equilibrium. In addition, under certain special conditions a body may be in equilibrium even when it is in motion. For example, a rigid body is in equilibrium if it moves in such a way that every particle in the body moves with uniform speed in a straight line. Another type of equilibrium is that of a wheel rotating about its axis with uniform angular speed. For a rigid body to remain in equilibrium when acted upon by a set of forces, two conditions must be satisfied:

(a) *The vector sum of all the forces acting on the body must be zero.* This condition assures that there will be no change in the state of the translational motion. Writing the condition in the form of an equation, we have

$$\boxed{\mathbf{F}_1 + \mathbf{F}_2 + \cdots + \mathbf{F}_n = \sum_{i=1}^{n} \mathbf{F}_i = 0.} \tag{4-2}$$

We note that this is the same as the condition for the equilibrium of a particle.

(b) *The vector sum of all the torques acting on the body about any axis must be zero.* In dealing with two-dimensional problems, this is equivalent to saying that the sum of the clockwise torques about any axis must equal the sum of the counterclockwise torques about the same axis. Writing this condition in the form of an equation, we have

$$\boxed{\mathbf{G}_1 + \mathbf{G}_2 + \cdots + \mathbf{G}_n = \sum_{i=1}^{n} \mathbf{G}_i = 0.} \tag{4-3}$$

This condition on the torques, that the sum of the torques must equal zero, is a new condition for equilibrium applicable to a rigid body which was not pertinent to the equilibrium of a particle, for all the forces acting on a particle had to intersect in that particle. The forces acting on a rigid body do not generally act on a single point in the body and consequently will give rise to rotational motion unless Equation (4-3) is fulfilled.

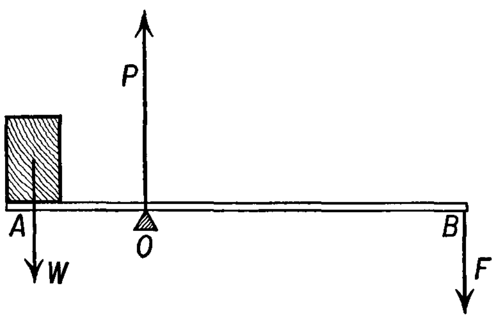


Fig. 4-5 Lever in equilibrium.

Illustrative Example. Let us analyze the forces associated with the operation of a lever. Essentially, a lever consists of a rigid bar AB, as in Figure 4-5, capable of rotating about a point of support O, called the *fulcrum,* which defines the axis of rotation. Suppose a weight **W** is placed at the end A and that some vertical force **F** is applied downward at the end B to keep the lever in equilibrium in a

horizontal position. Applying Equation (4-2) to the equilibrium of the bar AB, since the forces $\mathbf{W}$ and $\mathbf{F}$ are both in the y direction, the only other possible force, the force exerted by the fulcrum at O, must also be in the y direction. Calling this force $\mathbf{P}$, the vector equation for the forces must be

$$\mathbf{W} + \mathbf{F} + \mathbf{P} = 0,$$

and rewriting the equation with the symbols W, P, and F representing the magnitudes of the three forces, their directions being taken from the directions of the arrows on the figure, we have

$$-W + P - F = 0;$$

hence

$$P = W + F.$$

To apply the second condition for equilibrium, let us take moments of the forces about the point O with respect to an axis pointing normally out of the paper. If we consider O as the origin of a coordinate system with the positive x axis pointing toward the right to the point B, the positive y direction as the direction given by the vector $\mathbf{P}$, then the positive z direction points normally out of the paper toward the reader, as given by the right-hand convention. The moment of $\mathbf{W}$ about O is $+W \times \overline{AO}$, since the rotation which would be generated by $\mathbf{W}$ would be counterclockwise, and the torque vector would point in the positive z direction. The moment of $\mathbf{F}$ about O is $-F \times \overline{OB}$, since this is clockwise; the moment of $\mathbf{P}$ about O is zero. All the torques are in the z direction, and we apply the conditions for equilibrium in the form of Equation (4-3)

$$\sum G = W \times \overline{AO} - F \times \overline{OB} = 0,$$

from which

$$W \times \overline{AO} = F \times \overline{OB},$$

so that

$$\frac{W}{F} = \frac{\overline{OB}}{\overline{AO}}.$$

The distances $\overline{AO}$ and $\overline{OB}$ are called the *lever arms* of the respective forces $\mathbf{W}$ and $\mathbf{F}$. Thus, in the case of a lever, W and F are in the inverse ratio of their lever arms. By placing the fulcrum closer to $\mathbf{W}$, we shall now need a smaller force $\mathbf{F}$ to lift $\mathbf{W}$. The fulcrum may be placed at any point along the bar, and the positions of $\mathbf{W}$ and $\mathbf{F}$ may be moved around to get almost any desired result consistent with the approximation that the bar remains a rigid body. Many common tools are applications of the principle of the lever, as may be seen from an analysis of the use of the shovel, crowbar, tongs, wrench, tweezers, pliers, scissors, chain tightener, nail puller, and nutcracker.

Illustrative Example. A strong steel bar 5 ft long is supported at its two ends A and B, as shown in Figure 4-6. A weight of 160 lb is placed 2 ft from end A. Neglecting the weight of the bar, determine the forces exerted by the supports.

The forces acting on the steel bar are shown in Figure 4-6. The forces exerted by the supports are shown as $\mathbf{F}_A$ and $\mathbf{F}_B$. From the first condition of equilibrium, we get

$$F_A + F_B - 160\text{ lb} = 0.$$

In applying the second condition for equilibrium, we are at liberty to choose any axis of rotation. Let us choose an axis through the point A directed normally out of the paper. Following the previous example, we call this the positive z direction. The sum of the moments of all the forces about A is zero, yielding

$$F_A \times 0 - 160 \text{ lb} \times 2 \text{ ft} + F_B \times 5 \text{ ft} = 0,$$

from which

$$F_B = 64 \text{ lb}.$$

Substituting this back into the first equation gives us

$$F_A = 96 \text{ lb}.$$

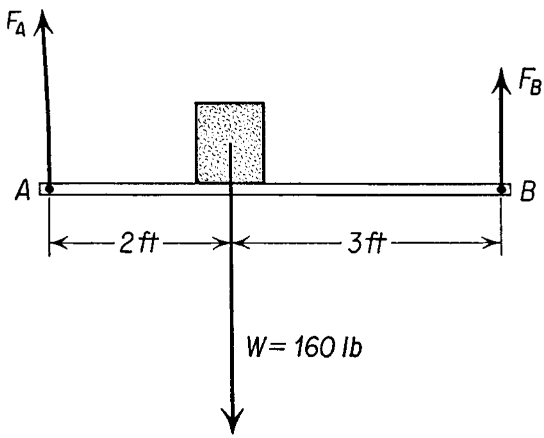


Fig. 4-6

This example really represents the solution of a great many problems in statics. If the line AB represents a simple bridge, then F_A and F_B represent the forces exerted by the bridge piers, and we have solved the problem of the loads borne by piers under one particular load distribution. If the line AB represents the bed of a truck, as it well might with the substitution of somewhat different numbers for the distance and weight, then **W** might represent the weight of the engine, and the two forces might represent the load borne by the front and rear tires.

Illustrative Example. A rod 8 ft long, and considered to be weightless, is pinned to a wall at one end, as shown in Figure 4-7(a). To support the rod horizontally a cord 10 ft long is fastened to the outer end of the rod and to the wall a distance of 6 ft above the pin. A 64-lb weight W is hung from the rod a distance of 3 ft from the pinned end. Find the tension in the cord and the force exerted by the pin on the rod.

We observe that we are here concerned with the equilibrium of a rigid body, namely the rod. From the dimensions given, the space figure is a 3–4–5 right triangle, and the angle ACD is 37°. Let us isolate the rod AC and label all the forces acting on it as shown in Figure 4-7(b). Since we know neither the magnitude nor the direction of the force exerted by the pin at A, we label the com-

ponents of this force $\mathbf{A}_x$ and $\mathbf{A}_y$, and draw them in the directions we expect these forces to act. Although we know the direction of the tension in the cord, it is more convenient to work in terms of the components of the tension $\mathbf{T}_x$ and $\mathbf{T}_y$. The forces on the rod are then A_x, A_y, W, T_x, and T_y, where these symbols in italics once again represent the magnitudes of the forces, the directions being

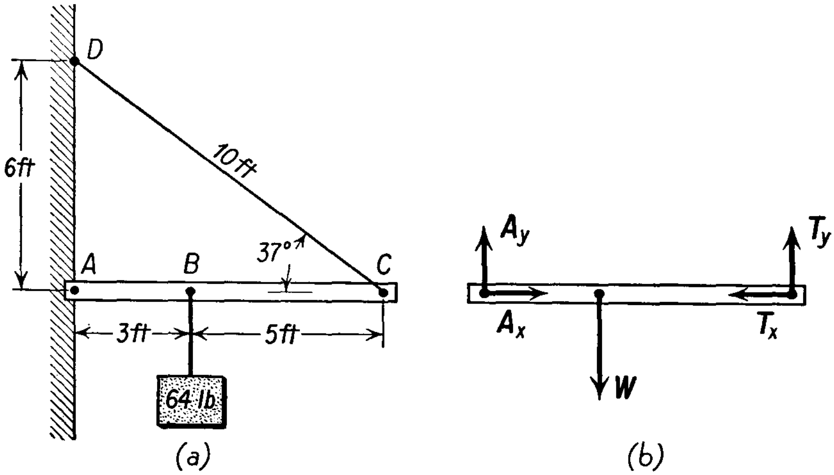


Fig. 4-7

given in the diagram. Following such a procedure, if one of the forces proves to have a negative value on solution of the problem, the direction of the particular force will be opposite to that shown in the figure.

We apply the component form of Equation (4-2) for the translational equilibrium of a rigid body:

$$\sum F_x = A_x - T_x = 0; \tag{a}$$

$$\sum F_y = A_y - W + T_y = 0. \tag{b}$$

Since T_x and T_y are components of a force $\mathbf{T}$, we may write

$$\frac{T_y}{T_x} = \tan 37° = \tfrac{3}{4}. \tag{c}$$

At this stage we have three equations in four unknowns, A_x, A_y, T_x, and T_y, and we need an additional relationship among these quantities to obtain a solution to the problem.

The second condition for equilibrium, Equation (4-3), provides the necessary relationship. Once again the positive z direction is taken as pointing out of the paper. The axis of rotation will be taken in the z direction, and the location of the axis of rotation will be chosen through the pin at A. The line of action of the forces $\mathbf{A}_x$, $\mathbf{A}_y$, and $\mathbf{T}_x$, all pass through the point A; hence these forces produce zero torque about an axis through A. It was for this reason that the point A was chosen as the location of the axis of rotation, and not because the pin was located

at A. The point C would have been an equally good choice for the location of the axis of rotation.

Substituting in Equation (4-2) for the torques about an axis through A, we obtain

$$\sum G_A = 0 = A_y \times 0 \text{ ft} + A_x \times 0 \text{ ft} - 64 \text{ lb} \times 3 \text{ ft} + T_x \times 0 \text{ ft} + T_y \times 8 \text{ ft},$$

from which

$$64 \times 3 \text{ lb ft} = 8 \times T_y \text{ ft};$$

hence

$$T_y = 24 \text{ lb}. \tag{d}$$

With this result the entire problem is reduced to algebra. From Equation (c) we get

$$T_x = \frac{T_y}{\tan 37^\circ} = \frac{24 \text{ lb}}{0.75} = 32 \text{ lb}. \tag{e}$$

From Equations (b) and (d) we find

$$A_y - 64 \text{ lb} + 24 \text{ lb} = 0,$$

so that

$$A_y = 40 \text{ lb}.$$

From Equations (a) and (e) we find that

$$A_x = T_x = 32 \text{ lb}.$$

Hence the tension in the rope T is of magnitude

$$T = (T_x^2 + T_y^2)^{1/2} = [(32)^2 + (24)^2]^{1/2} = 40 \text{ lb}.$$

The direction of **T** is known from the statement of the problem. The magnitude of the force on the pin A is given as

$$A = (A_x^2 + A_y^2)^{1/2} = [(32)^2 + (40)^2]^{1/2} = 51.2 \text{ lb};$$

the direction of the force **A** can be expressed in terms of the angle θ that it makes with the rod considered as the x axis; thus

$$\theta = \text{arc tan}\,\frac{A_y}{A_x} = \text{arc tan}\,\frac{40}{32} = 51.4^\circ.$$

4-5 Center of Gravity

In all our previous discussions in which it was necessary to consider the weight of a body, we represented it by a single force **W** downward. Actually, the earth exerts a force of attraction on each particle of the body; the weight of the body is the resultant of all the forces which act on all the particles of the body. We ask whether it is possible to think of an extended distribution of matter as though all its weight were concentrated at a single point in space. A plumb bob, a weight hung on the end of a string, represents an approximation of a particle. When a plumb bob is suspended, the weight hangs directly beneath the point of support. From an experi-

mental viewpoint, if there is a single point associated with an extended object where all the weight appears to be concentrated, this point should always come to rest beneath the point of support, no matter how the object is suspended. If an extended object is suspended from first one, then another, of several different points of support, the vertical lines through these points always intersect in a single point called the *center of gravity*. A single upward force of magnitude equal to the weight of the body will be sufficient to produce equilibrium if this force is applied at the center of gravity, regardless of the orientation of the body.

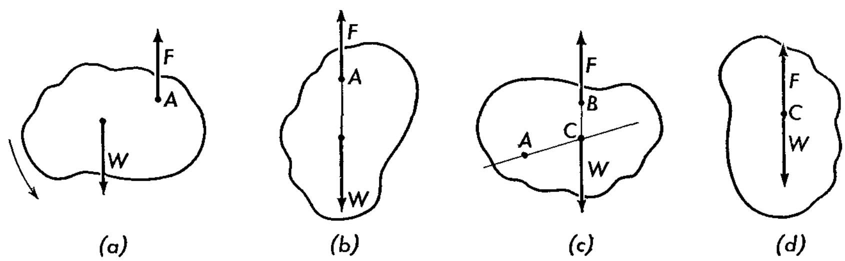


Fig. 4-8 Method of determining the position of the center of gravity of a body.

Suppose the body shown in Figure 4-8(a) is supported by a vertical force **F** at A, equal in magnitude to the weight of the body **W**, shown acting through the center of gravity. Considering an axis of rotation through A, the force **W** generates a torque which tends to rotate the body in the counterclockwise direction. The sum of the torques is not zero, and the body is not in equilibrium. Only when the center of gravity lies directly beneath the point of support, as in Figure 4-8(b), are the two conditions for equilibrium fulfilled. If the body is now supported at some other point B, the body will once again come to equilibrium, with its center of gravity beneath the point of support. The vertical line drawn through A when the body was in the position given in Figure 4-8(b) and the vertical line drawn through the second point of support B shown in Figure 4-8(c) intersect in the center of gravity C. Finally, when the body is supported at its center of gravity, the resultant of the force of support **F** and the force of gravity **W** is zero and therefore generates no torque about any point of support or about any other possible axes of rotation. Hence the body is in equilibrium in any orientation when it is supported at the center of gravity. The center of gravity is the balance point of the body.

If a body is homogeneous, that is, made of the same material throughout, and of simple geometric shape, such as a rectangular stick or a disk, a square plate or a sphere, the center of gravity lies at the geometrical center of the body. The center of gravity need not always lie at a place where

any of the matter of the body is located. For example, the center of gravity of a hollow ball lies at the center of the ball, and the center of gravity of a bottle lies somewhere within the bottle. Nevertheless, the location of the center of gravity is rigidly fixed to the body and cannot be moved without altering the body to which it is "attached."

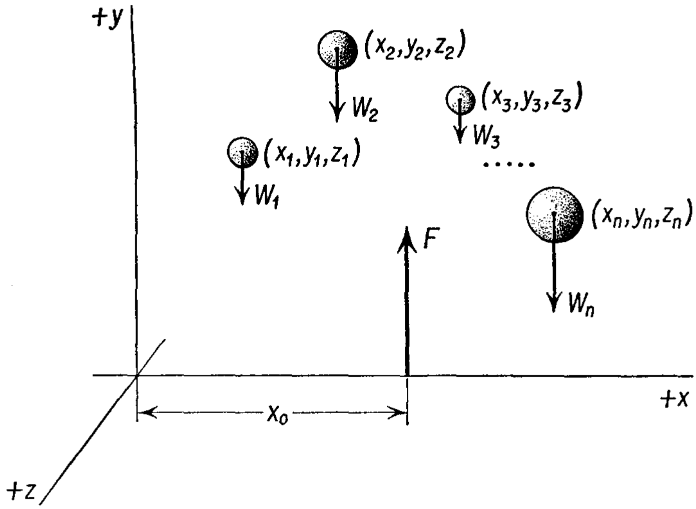


Fig. 4-9 The single force **F** acting through the center of gravity of the system of particles will support the system in equilibrium.

The location of the center of gravity of a distribution of particles may easily be calculated from the conditions of equilibrium for a rigid body. Consider a collection of n particles, each of which has weight W_i where $i = 1, 2, 3, \ldots n$, and is located at coordinates (x_i, y_i, z_i), as shown in Figure 4-9. To find the coordinates of the center of gravity, we imagine that these weights are attached to a rigid weightless framework, and we seek the location of a single force **F** which will support the system in equilibrium. The equilibrium for translational motion will be assured if **F** satisfies the first condition for equilibrium. Thus, summing the forces as shown in the figure,

$$\sum F_y = +F - W_1 - W_2 - W_3 - \cdots - W_n = 0,$$

from which

$$F = W_1 + W_2 + W_3 + \cdots + W_n = \sum W_i.$$

To satisfy the second condition for equilibrium, the sum of the torques acting on the system about any axis must be zero. We choose an axis of rotation directed along the z axis, passing through the origin. Each of the

forces $\mathbf{W}_i$ is acting in the $-y$ direction, while the force $\mathbf{F}$ is acting in the $+y$ direction through an unknown point whose coordinates may be taken as (x_0, y_0, z_0). The moment arm of the force $\mathbf{F}$ about the chosen axis is given by x_0, while the moment arm of a force $\mathbf{W}_i$ is given by its x coordinate x_i. Applying Equation (4-3) for determining the z components of the torque, we find

$$\sum G_z = +Fx_0 - W_1x_1 - W_2x_2 - W_3x_3 - \cdots - W_nx_n = 0;$$

thus
$$x_0 = \frac{W_1x_1 + W_2x_2 + W_3x_3 + \cdots + W_nx_n}{F},$$

so that
$$\boxed{x_0 = \frac{\sum W_ix_i}{\sum W_i}.} \tag{4-4a}$$

By reorienting the system so that the x axis is vertically upward, we can find the y coordinate of the center of gravity

$$\boxed{y_0 = \frac{\sum W_iy_i}{\sum W_i},} \tag{4-4b}$$

and in one additional reorientation we obtain

$$\boxed{z_0 = \frac{\sum W_iz_i}{\sum W_i}.} \tag{4-4c}$$

A distribution of matter not made up of point particles can be imagined to be divided into pieces of simple geometric shapes. Each of these may be replaced by a point particle of the same weight located at its center of gravity, and the location of the center of gravity of the body may then be calculated from Equations (4-4).

Illustrative Example. Find the location of the center of gravity of a carpenter's square made of sheet steel. The body dimensions of the rule are 24 in. $\times$ 2 in., and the dimensions of the tongue are 16 in. $\times$ $1\frac{1}{2}$ in. The square, laid onto a coordinate system, is illustrated in Figure 4-10. Suppose the square is made of material weighing σ (sigma) lb/in.^{2}. We divide the square up into two simple rectangles—a body section of dimensions 24 in. $\times$ 2 in.^{2} and a tongue section of dimensions 14 in. $\times$ $1\frac{1}{2}$ in.^{2}, as shown in the figure. The center of gravity of each of these sections is located at the center of that section. Thus we may imagine the body section whose cross-sectional area is 48 in.^{2} to be replaced

by a particle weighing 48σ lb located at the point whose (x, y, z) coordinates are given by (12, 1, 0). Similarly, the tongue section may be replaced by a particle

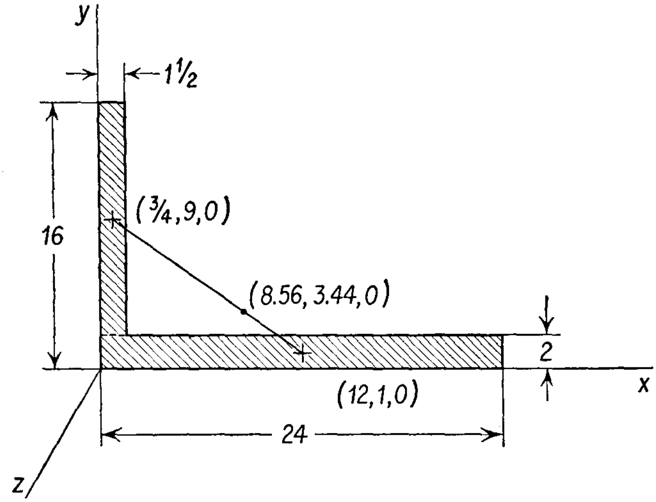


Fig. 4-10

weighing 21σ lb located at a point whose coordinates are $(\frac{3}{4}, 9, 0)$. For the case of two point particles, Equations (4-4) reduce to

$$\begin{aligned} x_0 &= \frac{W_1x_1 + W_2x_2}{W_1 + W_2} \\ &= \frac{48\sigma \text{ lb} \times 12 \text{ in.} + 21\sigma \text{ lb} \times \frac{3}{4} \text{ in.}}{48\sigma \text{ lb} + 21\sigma \text{ lb}} \\ &= 8.56 \text{ in.}, \\ y_0 &= \frac{W_1y_1 + W_2y_2}{W_1 + W_2} \\ &= \frac{48\sigma \text{ lb} \times 1 \text{ in.} + 21\sigma \text{ lb} \times 9 \text{ in.}}{48\sigma \text{ lb} + 21\sigma \text{ lb}} \\ &= 3.44 \text{ in.}, \end{aligned}$$

and, since the figure may be thought to be in the x-y plane,

$$z_0 = 0.$$

Thus the coordinates of the center of gravity have been obtained. As shown in the figure, the center of gravity of the system lies along the line joining the centers of gravity of the base and the tongue of the square.

We may represent the procedure for finding the center of gravity of an extended body in the form of an integral by replacing the summation

signs in Equations (4-4) by integral signs. Thus we have

$$x_0 = \frac{\int x\,dw}{\int dw} = \frac{\int x\,dw}{W}, \tag{4-5a}$$

$$y_0 = \frac{\int y\,dw}{W}, \tag{4-5b}$$

$$z_0 = \frac{\int z\,dw}{W}, \tag{4-5c}$$

where dw is the weight of a small volume element of the body located at coordinates x, y, z, and the total weight of the body is represented by W.

4-6 Discussion and Further Examples

The problems of statics vary greatly in difficulty, but if they are soluble at all they are soluble by the methods and principles developed in this chapter. The two fundamental principles which govern the equilibrium of a rigid body, and which govern the equilibrium of a particle in the limiting case that the rigid body is composed of a single particle, are: *The vector sum of all the forces acting on the body must be zero. The vector sum of all the torques about any axis acting on the body must be zero.* Written in equation form, these two statements are

$$\boxed{\begin{aligned} \sum F &= 0; \\ \sum G &= 0. \end{aligned}} \qquad \begin{aligned} &(4\text{-}6a) \\ &(4\text{-}6b) \end{aligned}$$

These two equations, in extremely concise form, represent our entire knowledge of the forces exerted by and on structural elements and form the analytical foundation upon which all structures are built. While, in general, equilibrium is interpreted to mean a state of rest with respect to the earth, it must be recognized that rest and uniform motion in a straight line are equivalent conditions, according to Newton's first law of motion. Thus it is that the very same equations apply to the equilibrium of a structure moving with uniform speed, and the analytic procedures which apply to the construction of a crane or a bridge may also be used in the design of an airplane.

Illustrative Example. A wagon wheel 26 in. in diameter and weighing 10 lb rests against a square curb 8 in. high, as shown in Figure 4-11. What horizontal force applied to the axle is necessary to push the wheel over the curb?

The wagon wheel will start to rise when the supporting force exerted by the roadway on the wheel is zero. At that time the forces acting on the wheel, as shown in Figure 4-11(b), are the unknown horizontal force **H**, the force of gravity **W** acting at the center of gravity of the wheel, and the force of the curb **P** against

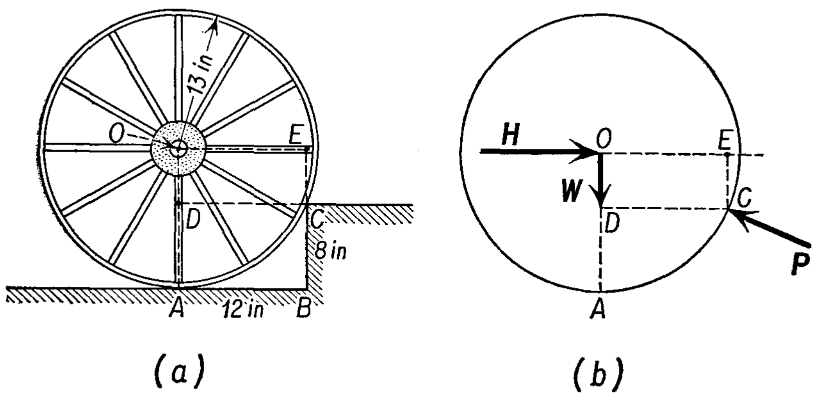


Fig. 4-11

the wheel. Let us choose an axis of rotation normal to the plane of the paper at the curb C. The moment arm of the force **W** is the distance $\overline{DC}$, 12 in. The moment arm of the force **H** is $\overline{EC} = \overline{OD} = 5$ in. Applying the torque condition for equilibrium, we know that the sum of the torques G_C about an axis normal to the plane of the paper through C is equal to zero; or

$$\sum G_C = 0 = W \times \overline{DC} - H \times \overline{EC},$$

or

$$10 \text{ lb} \times 12 \text{ in.} - H \times 5 \text{ in.} = 0,$$

so that

$$H = 24 \text{ lb}.$$

Illustrative Example. A ladder 26 ft long and weighing 30 lb leans against a smooth wall 24 ft from the ground and rests on a rough floor 10 ft from the wall. A man weighing 200 lb climbs 20 ft up the ladder before the ladder starts to slip [see Figure 4-12(a)]. (a) Find the forces exerted on the ladder by the floor and the wall. (b) What is the coefficient of static friction between the ladder and the floor?

We begin by isolating the ladder and labeling the forces acting on it, as shown in Figure 4-12(b). The unknown force exerted by the floor at the point a is called **A**, with components $\mathbf{A}_x$ and $\mathbf{A}_y$. The entire weight of the ladder **W** of 30 lb acts vertically downward through its center of gravity located at the middle of the ladder. The weight of the man **M** of 200 lb acts vertically downward through a point 20 ft up the ladder. The smooth wall exerts a force **B** which must be perpendicular to the wall. Once again, italic symbols represent the magnitudes

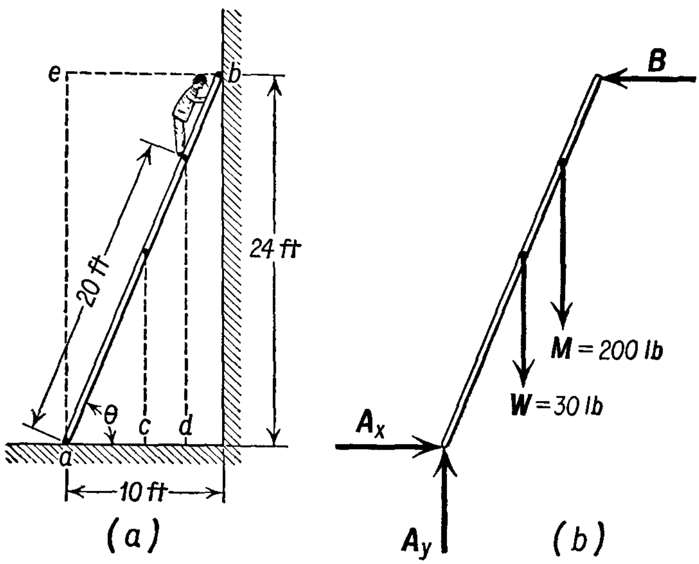


Fig. 4-12

of the forces, with directions given by the directions of the arrows. From the condition for equilibrium for the x components of the forces acting on the ladder, we have

$$\sum F_x = A_x - B = 0,$$

while for the y components we have

$$\sum F_y = A_y - 30\text{ lb} - 200\text{ lb} = 0,$$

so that

$$A_y = 230\text{ lb}.$$

Applying the conditions that the sum of the torques on the ladder must be zero, we choose an axis perpendicular to the plane of the paper through any convenient point such as a and get

$$\sum G_a = 0 = A_x \times 0 + A_y \times 0 - W \times \overline{ac} - M \times \overline{ad} + B \times \overline{ea}.$$

Substituting numerical values, we obtain

$$-30\text{ lb} \times 5\text{ ft} - 200\text{ lb} \times \tfrac{100}{13}\text{ ft} + B \times 24\text{ ft} = 0,$$

from which

$$B = \frac{150\text{ lb ft} + 1{,}540\text{ lb ft}}{24\text{ ft}} = 70.4\text{ lb};$$

and since $A_x = B$, from a preceding equation,

$$A_x = 70.4\text{ lb}.$$

The coefficient of static friction has been defined from the equation

$$F_r = fN.$$

In this example the force A_x is the frictional force, and A_y is the normal force, so that the coefficient of friction is equal to

$$f = \frac{A_x}{A_y} = \frac{70.4 \text{ lb}}{230 \text{ lb}} = 0.31.$$

Note that the coefficient of static friction was obtained from an analysis of the forces on the ladder when the ladder was on the point of slipping, when the force of static friction was at its maximum value.

Problems

4-1. Determine the torque produced by a force of 6 lb acting horizontally on the top of a bicycle wheel 24 in. in diameter with respect to an axis through its axle.

4-2. A torque of 5 ft lb is required to swing open a door which is 30 in. wide. What is the least force that must be exerted to open the door if it is applied (a) at a distance of 30 in. from the line of hinges and (b) at a distance of 24 in. from this line?

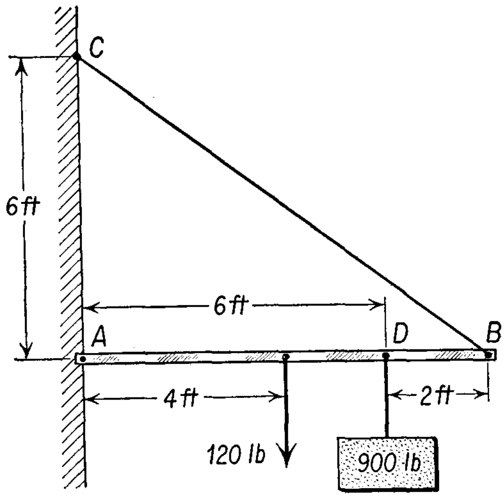


Fig. 4-13

4-3. A uniform horizontal bar AB, 8 ft long and weighing 120 lb, is pinned to the wall at A, while a steel cable 10 ft long extends out from a point C on the wall and is fastened to the bar at the point B, as shown in Figure 4-13. This bar supports a weight of 900 lb at a point D, 6 ft from the wall. Determine (a) the tension in the cable, (b) the vertical component and (c) the horizontal component of the force at A.

4-4. A man carries a bar 6 ft long which has two loads, one of 40 lb and the other of 60 lb, hung from its ends. At which point should the man hold the bar to keep it horizontal? Neglect the weight of the bar.

4-5. If the bar in Problem 4-4 is uniform and weighs 20 lb, determine the point at which the man should hold the bar to keep it horizontal.

4-6. A load of 180 lb is hung from a bar 10 ft long at a point 6 ft from one end. Two men carry this bar in a horizontal position. How big a force does each man exert, assuming that the bar is supported at its ends?

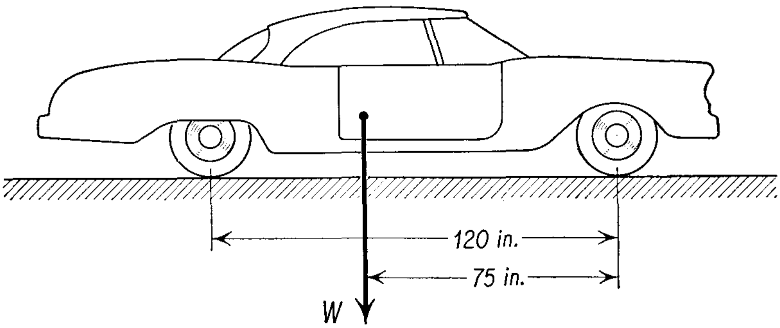


Fig. 4-14

4-7. A car weighing 3,200 lb has a wheel base of 120 in., and its center of gravity is 75 in. from the front wheels (see Figure 4-14). Determine the force (a) that the two front wheels exert on the ground and (b) that the two rear wheels exert on the ground.

4-8. A car weighing 3,600 lb has a wheel base of 125 in., and its center of gravity is 80 in. from the front wheels. Two passengers sit in the front seat. If their combined weight is 400 lb and if their center of gravity is at a point 60 in. from the front wheels, determine the shift in the center of gravity produced by the passengers.

4-9. A boom in the form of a uniform pole weighing 400 lb is hinged at the lower end. The boom is held at an angle of 60° with the ground by means of a horizontal cable attached to its upper end. (a) Determine the tension in the cable when there is no load on the boom. (b) Determine the tension in the cable when a load of 1,000 lb is attached to the upper end of the boom.

4-10. A door 8 ft high and 3 ft wide weighs 80 lb, and its center of gravity is at its geometrical center. The door is supported by hinges 1 ft from top and bottom, each hinge carrying half the weight. Determine the horizontal component of the force exerted by each hinge on the door.

4-11. A uniform ladder 25 ft long rests against a smooth vertical wall. The ladder weighs 30 lb. The lower end of the ladder is 15 ft from the wall. A man weighing 150 lb climbs up the ladder until he is 20 ft from the base of the ladder, at which point the ladder starts to slip. What is the coefficient of friction between the ladder and the floor?

4-12. Two rods, each of length 10 ft and weight 5 lb, are joined to make a 30° V. Find the center of gravity of the V.

4-13. Find the center of gravity of a collection of weights located at the

corners of an equilateral triangle, each side of length a. The three weights are 1, 2, and 3 lb, respectively. Place the x-axis along the line joining the 1 and 3 lb weights with the origin at the 1 lb weight.

4-14. A card table is made of 4 straight legs of dimensions 1 in. × 1 in. × 24 in., each weighing 1 lb, which are fastened to the corners of a square table top 30 in. on an edge by 1 in. thick. The table top weighs 5 lb. Find the center of gravity of the table.

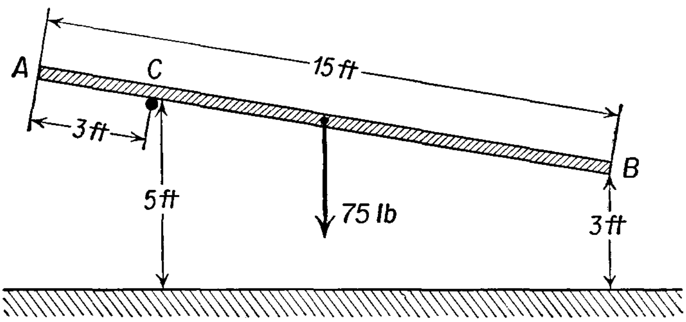


Fig. 4-15

4-15. A uniform beam 15 ft long weighing 75 lb is supported 3 ft from its upper end A by a smooth cylindrical rail which is 5 ft from the ground, as shown in Figure 4-15. What force must be exerted at the lower end B of the beam, located 3 ft from the ground, in order to support the beam?

4-16. A chain 5 ft long is placed on a horizontal table so that part of it hangs over the edge. If it starts to slip when 2 ft of chain hang over the side, find the coefficient of starting friction between the chain and the table.

4-17. Find the location of the center of gravity of a square sheet of metal of edge 4 in. which has had a smaller square of edge 1 in. cut out of one corner.

4-18. A uniform ladder 20 ft long and weighing 35 lb rests against a smooth wall at an angle of 30° to the wall. A 200-lb man stands 15 ft up the ladder. If the coefficient of friction between the floor and the ladder is 0.1, what additional horizontal force must be exerted at the base of the ladder to keep it from slipping?

4-19. Show that the center of gravity of a thin uniform board cut in the form of an isosceles triangle of altitude h is at a point $\frac{2}{3}h$ from the vertex on the perpendicular bisector of the base. [HINT: Choose a set of x-y coordinate axes with the origin at the vertex and the x axis along the perpendicular bisector, as shown in Figure 4-16. Take an element of the board formed by two lines a distance dx apart parallel to the base. The area of this element is $2y\,dx$, and its weight is $dw = \sigma \cdot 2y\,dx$, where σ is the weight per unit area. Then apply Equation (4-5a). Note that

$$y = \frac{b}{a}x$$

where $2b$ is the width of the base.]

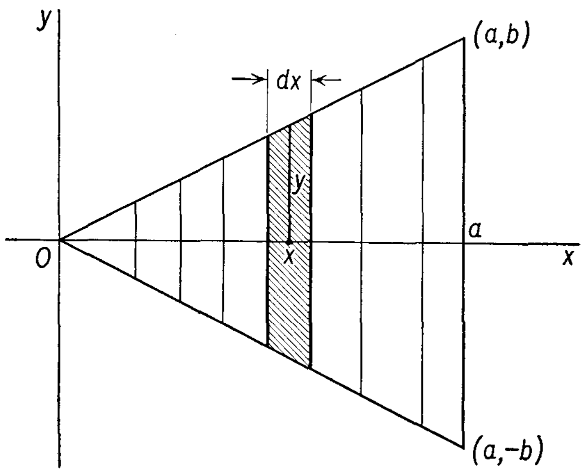


Fig. 4-16

4-20. Find the center of gravity of a thin board cut in the form of a 3–4–5 right triangle. [HINT: Apply the result of Problem 4-19.]

4-21. An irregular slab of material is pivoted at one corner by a horizontal pin, and is supported by a vertical force of 80 lb located 10 ft to the right of the pin. The slab weighs 200 lb. (a) How far to the right of the pivot is the center of gravity located? (b) What is the force on the object due to the pivot?

4-22. Show that if the resultant of a set of concurrent forces is zero, the sum of the moments of these forces about any axis is zero.

4-23. Using the second condition for the equilibrium of a body, show that when a body is in equilibrium under the action of three nonparallel forces, these forces must pass through a single point; that is, the forces are concurrent.

5

Force and Motion

5-1 Starting and Stopping Motion

All of us have many times had the experience of setting a body in motion. If we analyze any of these experiences, we readily recall that in each case some *force* was required to start the object moving. In throwing a ball, moving a piece of furniture, or pulling a sled, the force needed to start the object moving is supplied by one's muscular effort as a push or a pull. In more complex cases, such as setting a car or an airplane in motion, the analysis, although more complicated, will also show that a force is required to start the body moving.

There are many cases in which the force that acts on the body to produce the motion is not directly discernible. It was Newton who first showed that the acceleration of a freely falling body is produced by a force which acts between the earth and the body, called the *force of gravitation.* We shall encounter other such action-at-a-distance forces in electricity and magnetism, and in molecular and atomic physics.

Once a body has been set in motion by the action of a force, it will not necessarily stop moving when the force is removed. A sled in motion along a level road will continue to move in a straight line along the road, although with diminishing speed. The reduction in speed is due to the *force of friction* between the runners of the sled and the ground. If there is clean snow on the ground, the force of friction will be very small; if ashes or sand have been dumped on the snow, the force of friction will be greater, and the sled will come to rest much sooner.

The above examples illustrate the fact that a *force is required to change the state of motion of a body.* It was Sir Isaac Newton (1642–1727) who first recognized the relationship between force and the state of motion of the body on which it is acting. He epitomized the entire science of mechanics in the form of three statements which have become known as Newton's laws of motion. Although the first and third laws have been previously stated and discussed, they are sufficiently important to bear repetition.

5-2 Newton's Laws of Motion

Newton's three laws of motion can be stated as follows:

First law: *A body at rest will remain at rest, and a body in motion will continue in motion with constant speed in a straight line, as long as no net force acts on it.*

Second law: *If a net force acts on a body, the body will be accelerated; the magnitude of the acceleration is proportional to the magnitude of the force, and the direction of the acceleration is in the direction of the force.*

Third law: *Whenever one body exerts a force on another, the second body exerts a force equal in magnitude and opposite in direction on the first body.*

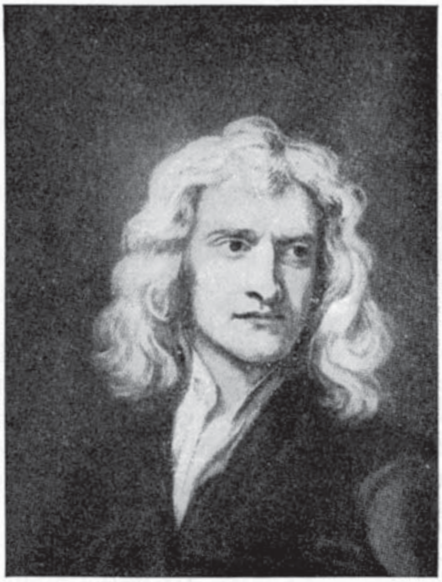


Fig. 5-1 Sir Isaac Newton (1642–1727). One of the greatest physicists of all time, he developed the law of universal gravitation; epitomized the subject of mechanics in the three laws of motion which bear his name; made important contributions to optics. The publication of his *Principia*, the Mathematical Principles of Natural Philosophy, in 1687, was an epoch-making event for science. (Courtesy of *Scripta Mathematica.*)

5-3 Newton's First Law

Newton's first law states that a body at rest will remain at rest, and a body in motion will continue in motion with constant speed in a straight line, as long as no net force acts on it.

An examination of this first law shows that a body at rest and a body moving with constant velocity have one characteristic in common: there is no net external force acting upon either one. This is the case when the resultant of all the external forces acting on the body is zero. As we have already seen, this is the condition for the equilibrium of a particle; this is also the condition for the translational equilibrium of a rigid body.

According to Newton's first law, a train moving at a constant velocity along a level track is in equilibrium. It is acted upon by several external

forces whose resultant is zero. Consider the forces acting on a train of cars being pulled by a locomotive (see Figure 5-2). The weights $\mathbf{W}_1$, $\mathbf{W}_2$, $\mathbf{W}_3$, of the cars act vertically downward through the respective centers of gravity. They are opposed by the forces $\mathbf{N}_1$, $\mathbf{N}_2$, $\mathbf{N}_3$, and so on, which the tracks exert upward on the wheels of the train to support the weight. The sum of these upward forces must equal the total weight of the train. There are also frictional forces which oppose the motion of the train. Some of these frictional forces occur between the wheels and the tracks and in the wheel bearings; there is also another type of frictional force owing to the

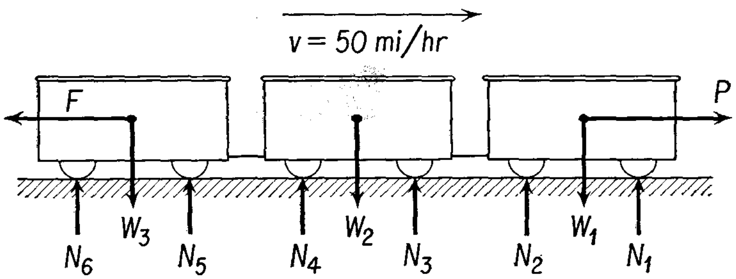


Fig. 5-2 A train moving with constant velocity has no net force acting on it. $W_1 + W_2 + W_3 = N_1 + N_2 + N_3 + N_4 + N_5 + N_6$. $P = F$.

resistance of the air to motion through it. All of these frictional forces are represented in the figure by the single force **F**. The effect of these frictional forces would be to reduce the speed of the train; to prevent this reduction in speed, the locomotive supplies a force **P** equal to **F** in magnitude but in the forward direction. There is no net force acting on the train when it is moving with constant velocity.

In this illustration we have taken a very liberal view of the meaning of *body* and of *net force*. We have considered the collection of all the cars of the train as a body, or as a *system* which could be surrounded by an imaginary box. Everything within the walls of the box was considered to be the body, and only the forces acting from outside the box upon objects inside the box were considered as forces acting on the body. In addition to the forces illustrated in the figure, each car exerts a force upon the two cars immediately adjacent to it. Nevertheless, these *internal* forces can be disregarded in our analysis of the over-all motion of the system, and our attention can be focused upon the *external* forces acting upon the system. From Newton's third law the sum of these internal forces must be zero. This procedure is analogous to what we have already done in the study of the equilibrium of a rigid body when we considered only the external forces acting on the rigid body and paid no attention to the internal forces which connected one particle to another particle of the body.

Implicit in the statement of Newton's first law is a property common to all objects—the property known as *inertia.* The inertia of a body is that property of a body associated with the first law, that a body at rest will remain at rest unless acted on by a net force, and that a body in motion will continue to move with uniform velocity unless acted on by a net force.

The magician who whisks a cloth from under the dinner dishes on a table, the prankster who places a brick under a hat on the sidewalk, have a qualitative understanding of the concept of inertia. We shall attempt to systematize and formalize this concept in the following sections of this chapter.

5-4 Newton's Second Law

Newton's second law states that if a net force acts on a body, the body will be accelerated; the magnitude of the acceleration is proportional to the magnitude of the force, and the direction of the acceleration is in the direction of the force.

Let us examine the meaning of the second law of motion. When a single force acts upon a body, an acceleration results. The body is accelerated in the direction in which the force acts. The magnitude of the acceleration is proportional to the magnitude of the force and to some quality of the body which has not yet been specified. Since both force and acceleration are vector quantities, we may write Newton's second law in the form of an equation:

$$\mathbf{F} = k\mathbf{a}. \tag{5-1}$$

The quantity k is used to represent a constant of proportionality, a scalar quantity, having whatever dimensions are necessary to give the equation formal meaning. The constant k must clearly depend on the properties of the body, for the other quantities in the equation do not. The value of k must also depend upon the choice of units for $\mathbf{F}$ and $\mathbf{a}$. It is desirable to break up k into two parts, one of which depends only upon the properties of the body, and another which depends only upon the choice of units in which to express the magnitudes of these quantities. Thus we rewrite Equation (5-1) as

$$\mathbf{F} = KM\mathbf{a}, \tag{5-2}$$

where both K and M must be scalar quantities. We use the constant K to represent that part of k which is associated with a choice of units, and let the symbol M represent the part of k which is associated with the body being accelerated. We have already referred to the resistance of a body to a change in its state of motion through the qualitative concept of inertia. The quantitative measure of the inertia of a body is its *mass,* represented in Equation (5-2) by the symbol M.

Given a choice of the constant K, to be associated with the choice of units used to represent $\mathbf{F}$ and $\mathbf{a}$, Equation (5-2) represents both a definition of mass and a recipe for its experimental determination. If we had chosen to define a unit of force in terms of the deflection of a spring, we could determine the mass of a body by using the spring to exert a force on the body resting on a frictionless table. The resulting acceleration might be determined through the measurement of the distance traversed in a known time. Such a determination of mass would be called a *dynamic* determination. We have already seen in Chapter 1 that the customary method for the determination of mass is based upon comparing the earth's gravitational force upon the unknown mass and a standard mass in a beam balance.

It is considerably simpler to embody a standard of mass as a preservable physical entity, say in the form of a piece of metal, than it is to embody a standard of force. While the standard of mass may be protected from alteration by wear and corrosion, the properties of a spring which determine the force it exerts vary with the age and condition of the spring. Consequently, mass is often taken to be a fundamental mechanical quantity, along with distance and time, and the force is considered to be a derived quantity, whose definition is based upon Newton's second law of motion.

5-5 Absolute Systems of Units

When numerical values are used with an equation involving physical quantities, such as Equation (5-2), these numerical values must be accompanied by appropriate units. There are many different sets of units in actual use today, each set consistent within itself, each chosen for some special merit which it is supposed to have for the particular group of experiments or investigations under consideration. An *absolute* system of units is one in which the unit of *force* is defined without reference to gravity, as in the two metric systems of units discussed in this section. A *gravitational* system of units is one in which gravity, or weight, is used as the basis of the definition of a unit of force, as in Section 5-6. Quantities expressed in one set of units can be converted more or less readily into any other set of units. Most physicists prefer to base the systems of units upon length, mass, and time as the fundamental concepts. Of these systems one of the most widely used is the cgs system in which the centimeter, gram, and second are the units for the respective fundamental quantities.

For convenience, the constant K of Equation (5-2) is set equal to 1, a pure number without physical dimensions. When $K = 1$ Equation (5-2) becomes

$$\boxed{\mathbf{F} = M\mathbf{a}.} \tag{5-3}$$

Equation (5-3) is the form most commonly used to represent Newton's second law of motion. It must be emphasized that the quantity **F** is the net force, or unbalanced force, or the resultant force acting on the body. If the force is entirely in the x direction, the acceleration must also be in the x direction. Resolving both the force and the acceleration into components parallel to each of the three coordinate axes, we obtain the component form of Equation (5-3).

$$\mathbf{F}_x = M\mathbf{a}_x, \tag{5-4a}$$

$$\mathbf{F}_y = M\mathbf{a}_y, \tag{5-4b}$$

$$\mathbf{F}_z = M\mathbf{a}_z, \tag{5-4c}$$

where the symbols $\mathbf{F}_x$, $\mathbf{F}_y$, and $\mathbf{F}_z$ represent the x, y, and z components of the net or resultant force acting upon the body, and $\mathbf{a}_x$, $\mathbf{a}_y$, and $\mathbf{a}_z$ represent the x, y, and z components of the acceleration.

In the cgs absolute system of units, the mass of a body is expressed in *grams*, and the acceleration is expressed in *centimeters per second per second.* A unit of force must be introduced that will be consistent with Equation (5-3). This unit of force is called a *dyne* and is defined as *that force, which, acting on a one-gram mass, produces an acceleration of one centimeter per second per second.* Thus Equation (5-3), together with legally defined units of mass, length, and time, has been used to generate a unit of force.

Suppose that a force F acts on a body whose mass is 1 gm and that it produces an acceleration of 1 cm/sec^2. Then Equation (5-3) would read

$$F = 1 \text{ gm} \times 1 \frac{\text{cm}}{\text{sec}^2} = 1 \frac{\text{gm cm}}{\text{sec}^2} = 1 \text{ dyne.}$$

Illustrative Example. A loaded car has a mass of 2,800 gm. (a) What horizontal force is required to give this car an acceleration of 80 cm/sec^2? (b) What velocity will this car acquire if it starts from rest and the force acts on it for 8 sec?

(a) Using Equation (5-3), and noting that **F** and **a** have the same direction, we may write

$$F = Ma,$$

and substituting values for M and a, we get

$$F = 2{,}800 \text{ gm} \times 80 \frac{\text{cm}}{\text{sec}^2},$$

or

$$F = 224{,}000 \text{ dynes.}$$

(b) The velocity of the car can be determined with the aid of the equation $v = u + at$ with $u = 0$, $a = 80$ cm/sec^2, and $t = 8$ sec, yielding

$$v = 80 \frac{\text{cm}}{\text{sec}^2} \times 8 \text{ sec},$$

so that

$$v = 640 \frac{\text{cm}}{\text{sec}} \cdot$$

Another absolute system which is widely used is the mks system of units based upon the meter, kilogram, and second as the respective units of length, mass, and time. The unit of force in the mks system is the *newton*, which is defined as that *force which, acting on a one-kilogram mass, produces an acceleration of one meter per second per second.*

If a force F acts on a body whose mass is 1 kg and produces an acceleration of 1 m/sec^2, then, from Equation (5-3), we have

$$F = 1 \text{ kg} \times 1 \frac{\text{m}}{\text{sec}^2} = 1 \frac{\text{kg m}}{\text{sec}^2} = 1 \text{ nt}.$$

We can obtain the relationship between a newton and a dyne from the above equation thus:

$$1 \text{ nt} = 1 \text{ kg} \times 1 \frac{\text{m}}{\text{sec}^2}$$

$$= 1{,}000 \text{ gm} \times 100 \frac{\text{cm}}{\text{sec}^2} = 100{,}000 \frac{\text{gm cm}}{\text{sec}^2},$$

so that $$1 \text{ nt} = 100{,}000 \text{ dynes} = 10^5 \text{ dynes}.$$

Illustrative Example. A force of 800 nt is applied to a mass of 160 kg. (a) Determine the acceleration produced. (b) If the body starts from rest, determine the distance the body travels if the force acts on it for 12 sec.

(a) Solving Equation (5-3) for the acceleration, we get

$$a = \frac{F}{m},$$

and, substituting numerical values for F and M, we obtain

$$a = \frac{800 \text{ nt}}{160 \text{ kg}} = 5 \frac{\text{m}}{\text{sec}^2}.$$

(b) Since the initial velocity $u = 0$, we can use the equation

$$s = \tfrac{1}{2}at^2$$

for determining the distance s traveled at the constant acceleration of 5 m/sec^2 for 12 sec, obtaining

$$s = \tfrac{1}{2} \times 5 \frac{\text{m}}{\text{sec}^2} \times 144 \text{ sec}^2,$$

from which $$s = 360 \text{ m}.$$

The above two systems of absolute units are based on the metric system and will be used throughout this book.

5-6 British Gravitational System of Units

While the legally defined unit of mass is the pound mass, defined as 1/2.20462 kilogram, it is inconvenient to use the pound mass as the basis of a system of units, for in everyday terminology, and in many engineering applications, the word "pound" commonly refers to weight rather than to mass. To conform to this common usage, the British gravitational system of units has chosen to define the pound of force as the weight of the standard one-pound body at sea level and at 45° latitude.

Modern engineering practice tends to avoid the use of the pound mass by introducing a new unit of mass called a *slug*. The *slug* is defined as *that unit of mass, which, when acted on by a force of one pound, will acquire an acceleration of one foot per second per second.* In the British gravitational system of units, the unit of force is the *pound*, the unit of mass is the *slug*, the unit of distance is the *foot*, and the unit of time is the *second.*

In everyday language the word "pound" is commonly used as a unit of mass and as a unit of force. Usually one can infer from the context of a statement whether the pound is used as a unit of force or as a unit of mass. For example, in Chapter 3 the pound was consistently used as a unit of force. In statements where it is possible to interpret the word "pound" as either force or mass, the terms "pound force" or "pound mass" should be used to avoid ambiguity.

Let us examine the relationship between the pound mass and the slug. From Table 2-1 we see that at latitude 45° the acceleration of gravity is 32.17 ft/sec^2. This is the acceleration acquired by any freely falling body at sea level, and in particular it is the acceleration which would be acquired by a pound mass falling freely at sea level. The weight of a pound mass at this latitude and elevation has been defined as the pound of force. But the weight of a body is the force of the earth's gravitational attraction. Writing M as the mass in appropriate units, and substituting in Equation (5-3), with $F = 1$ lb and $a = 32.17$ ft/sec^2, we find

$$1 \text{ lb} = M \times 32.17 \text{ ft/sec}^2;$$

hence

$$M = \frac{1}{32.17} \text{ slug.}$$

A pound mass has a mass in slugs given by 1/32.17 slug; thus the mass of one slug is 32.17 lb mass.

In general, if the weight of a body W is the only force which acts upon it, it is a freely falling body and has an acceleration g. We can apply Newton's second law to a freely falling body by setting $F = W$ and $a = g$ in the equation $F = Ma$, to obtain

$$W = Mg. \tag{5-5}$$

The appropriate units for mass, length, time, and force for use in Equation (5-3) are shown in Table 5-1. Only when units appropriate to a

TABLE 5-1 SYSTEMS OF UNITS

System	Mass	Length	Time	Force
Metric absolute—cgs	Gram	Centimeter	Second	Dyne
Metric absolute—mks	Kilogram	Meter	Second	Newton
British gravitational	Slug	Foot	Second	Pound

particular system are used is the constant K of Equation (5-2) equal to 1, and, in fact, it is this consideration which converts a collection of apparently unrelated quantities into a system of units.

Illustrative Example. An automobile weighing 3,200 lb starts from rest and acquires a speed of 30 mi/hr in 5 sec. Determine the resultant force on the automobile.

The acceleration of the automobile is

$$a = \frac{30 \text{ mi/hr}}{5 \text{ sec}} = \frac{44}{5} \frac{\text{ft}}{\text{sec}^2}.$$

The mass of the automobile is found by substitution in Equation (5-5):

$$3{,}200 \text{ lb} = M \times 32 \frac{\text{ft}}{\text{sec}^2},$$

yielding

$$M = 100 \text{ slugs}.$$

To find the force we substitute the now known values of M and a in appropriate units in Equation (5-3):

$$F = 100 \text{ slugs} \times \frac{44}{5} \frac{\text{ft}}{\text{sec}^2} = 880 \text{ lb}.$$

A general class of problems of great value in developing understanding of Newton's second law deals with two or more bodies connected by a rope which is passed over a pulley. The bodies may be hanging freely or may be supported on inclined planes, and to add additional complications the planes may be made rough. In such problems we analyze the forces acting on each body separately, and then tie the system together through an algebraic statement about the way the various parts are connected. We call these connections the *constraints* under which the system is required to move. Thus a body sliding on a horizontal table is constrained to move in the horizontal plane of the table. A bead sliding along a wire is constrained to move along that wire. Two bodies connected by an inextensible rope must always be a fixed distance apart.

Illustrative Example. A useful experimental device, called *Atwood's machine,* consists of two bodies suspended by a rope which is passed over a frictionless fixed pulley, as illustrated in Figure 5-3. Let us call the masses of the suspended objects M_1 and M_2, the tension in the ropes S_1 and S_2, as illustrated in the figure, and their respective accelerations a_1 and a_2. Suppose that M_1 and M_2 are known and we wish to find the accelerations of the two bodies and the tension in the rope.

The force of gravity acting upon these bodies is equal to the weights of the bodies M_1g and M_2g, respectively. Each body experiences an upward force

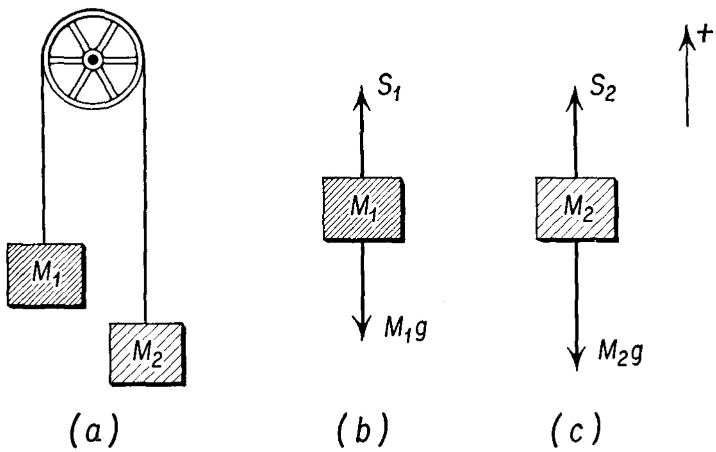


Fig. 5-3 Atwood's machine.

produced by the pull of the rope on it. Let us call the tension in the rope acting on the first body S_1, and the tension in the rope acting on the second body S_2. Following our customary sign conventions, we call the direction vertically upward positive.

First we imagine the bodies M_1 and M_2 to be completely isolated in space, as shown in Figures 5-3(b) and 5-3(c), and apply Equation (5-3) to determine their motion. Applying Newton's second law, we get

for M_1: $$S_1 - M_1g = M_1a_1; \tag{a}$$

for M_2: $$S_2 - M_2g = M_2a_2. \tag{b}$$

Now we examine the connection between the two bodies. First, because the rope which connects them passes over a frictionless pulley, the tension in the rope is everywhere the same. By the very nature of a tensile force, when the rope pulls M_1 upward, it must also pull M_2 upward. Thus the rope acts simultaneously on M_1 and M_2 in the directions indicated in Figures 5-3(b) and 5-3(c), and we may write

$$S_1 = S_2 = S. \tag{c}$$

Now we consider a second aspect of the connection between the two bodies. The distance between them, measured along the rope, is always the length of the rope. Thus if the body M_1 moves 1 ft upward, the body M_2 must move 1 ft downward. A positive displacement of M_1 generates an equal negative displacement of M_2. If the body M_1 is given a positive acceleration, the body M_2 must be given an equal negative acceleration. We may write

$$a_1 = -a_2 = a. \tag{d}$$

Substituting equations (c) and (d) into (a) and (b), we have

$$S - M_1 g = M_1 a,$$

$$S - M_2 g = -M_2 a.$$

Subtracting the second equation from the first, we obtain

$$M_2 g - M_1 g = M_2 a + M_1 a,$$

from which

$$a = \frac{M_2 - M_1}{M_2 + M_1} g.$$

Thus, if M_2 is greater than M_1, the body M_2 receives a positive acceleration of magnitude a, while the body M_2 experiences a negative acceleration of equal magnitude. If we multiply the first equation by M_2, the second by M_1, and add, we find

$$(M_1 + M_2)S = M_1 M_2 g,$$

from which

$$S = \frac{M_1 M_2}{M_1 + M_2} g.$$

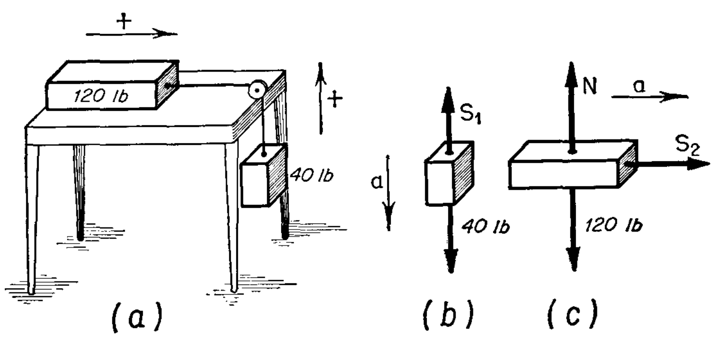


Fig. 5-4

Illustrative Example. A box weighing 120 lb is placed on a smooth table. A cord tied to this box passes over a smooth pulley fixed to the edge of the table. Another box weighing 40 lb is fastened to the other end of the cord, as shown in Figure 5-4(a). Determine the acceleration of the two bodies and the tension in the cord.

First we imagine the two bodies to be isolated in space and examine the forces

acting on them, as in Figures 5-4(b) and 5-4(c). The only forces acting on the 40-lb body are its weight and the tension in the rope supporting it, which we shall call S_1. Three forces act on the 120-lb body. These are the weight of the body of 120 lb, acting vertically downward, the force of the smooth table on the body acting vertically upward, which we call N, and the tension in the rope acting on it, which we call S_2. Since the 120-lb body is constrained to move in a horizontal plane, it can have no vertical acceleration, and

$$N = 120 \text{ lb}.$$

The resultant of the forces acting on the 120-lb body is S_2 acting to the right. Writing a_1 for the acceleration of the 40-lb body and a_2 for the acceleration of the 120-lb body, we have, from Equations (5-4a) and (5-4b),

$$S_1 - 40 \text{ lb} = \tfrac{40}{32} \text{ slug} \times a_1, \tag{a}$$

and

$$S_2 = \tfrac{120}{32} \text{ slug} \times a_2. \tag{b}$$

Examining the nature of the constraint imposed by the rope, we note first that the magnitude of the tension in the rope must be the same at both ends of the rope, and that the directions chosen for S_1 and S_2 are appropriate, for a positive value of S_1 implies a positive value of S_2. We write

$$S_1 = S_2 = S. \tag{c}$$

Next we find that a displacement of the 120-lb weight to the right implies an equal displacement of the 40-lb weight downward. Thus a positive displacement of the 120-lb weight implies an equal negative displacement of the 40-lb weight, and a positive acceleration of the 40-lb weight implies an equal negative acceleration of the 120-lb weight, and we write

$$+a_2 = -a_1 = a. \tag{d}$$

Substituting equations (c) and (d) into (a) and (b), we have

$$S - 40 \text{ lb} = \tfrac{40}{32} \text{ slug} \times (-a), \tag{e}$$

and

$$S = \tfrac{120}{32} \text{ slug} \times a. \tag{f}$$

Subtracting the second from the first of these equations, we have

$$-40 \text{ lb} = -\tfrac{160}{32} \text{ slug} \times a, \tag{g}$$

from which

$$a = 8 \frac{\text{ft}}{\text{sec}^2}; \tag{h}$$

substituting this value of a in equation (f) yields

$$S = 30 \text{ lb}.$$

We note that the 120-lb weight receives a positive acceleration of 8 ft/sec^2, while the 40-lb weight receives an equal negative, or downward, acceleration.

It is particularly important to emphasize the procedure used in solving these pulley problems. In each case we have systematically isolated the bodies involved in the problem and have examined the forces acting on each

of them. To each body we then applied Newton's second law. The constraints which related the several motions to give a sufficient number of relationships to solve the problem were then introduced. In more advanced courses in mechanics, more sophisticated methods are developed for the solutions of such problems, but in all cases the methods depend on Newton's equation, and if a problem is soluble by any method it is soluble by the persistent and systematic application of Newton's second law of motion.

5-7 Weight and Mass

Although considerable space has already been devoted to a discussion of the distinction between weight and mass, the subject is of sufficient importance to warrant further emphasis. If we consider two freely falling bodies at the same place on the earth's surface, one of which has a mass M and weight W, while the other has a mass m and weight w, we find, by application of Equation (5-5) to each of the bodies, that

$$W = Mg,$$

and

$$w = mg,$$

from which

$$\frac{W}{w} = \frac{M}{m}. \tag{5-6}$$

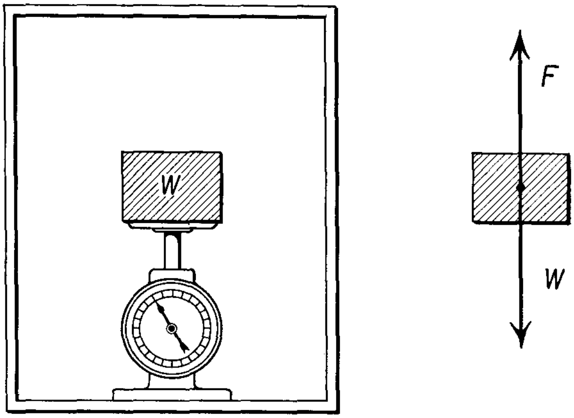
Thus the magnitudes of the weights of two bodies at the same place are in the same ratio as their masses. This is the reason the beam balance can be used to determine the mass of an unknown object in terms of a standard mass.

As we go from place to place, the value of g changes with latitude and with altitude. The mass of a body, however, remains constant unless the body is traveling with a speed comparable to the speed of light, which is about 186,000 mi/sec, in which case the mass of the body increases over its mass at rest. We shall restrict this discussion to bodies moving with speeds which are small in comparison with the speed of light. The weight W of a body of constant mass M depends upon the particular place where the weight is measured. In the systems of units used in this book, the weight of a body is properly referred to in units of force. Thus the weight of a body is properly expressed in dynes, in newtons, or in pounds. In these terms the operations customarily undertaken in a chemical laboratory, called "weighings," are more properly "massings," for the analytical chemist is not interested in the force of the earth's attraction but in the quantity of matter present in a sample.

The weight of an object is measured by the deflection of a calibrated spring. The reading of such a spring scale varies with the state of motion

of the scale. When the scale is at rest, or moving with uniform motion in a straight line, the reading of the scale is the same, but when the motion is accelerated, the reading of the scale depends on both the mass of the object and the amount of acceleration. This is the source of the sensation of heaviness which is experienced when standing in an elevator being accelerated upward, or of lightness in the same elevator when its acceleration is downward.

Fig. 5-5



Illustrative Example. A weight W rests on a spring scale which is placed on the floor of an elevator. The scale reads 50 lb when the elevator is at rest. The elevator is started, accelerating upward at a rate of 16 ft/sec^2 for 1 sec, then continues at constant speed for 5 sec, and finally is decelerated at the rate of 16 ft/sec^2 for 1 sec. What is the reading of the scale during the first second? During the next 5 sec? During the last second?

The forces acting on the body are its weight W, acting downward, and the force of the scale F, acting upward, as shown in Figure 5-5.

During the first second the acceleration is 16 ft/sec^2 in the upward direction. Substituting in Newton's equation, we find that

$$F - 50\text{ lb} = \tfrac{50}{32}\text{ slug} \times 16\,\frac{\text{ft}}{\text{sec}^2},$$

so that

$$F = 75\text{ lb}.$$

During the next 5 sec the acceleration is zero. Hence the scale reads 50 lb.

For the last second the acceleration is 16 ft/sec^2 in the downward direction. Thus $a = -16$ ft/sec^2, and we find that

$$F - 50\text{ lb} = \tfrac{50}{32}\text{ slug} \times \left(-16\,\frac{\text{ft}}{\text{sec}^2}\right),$$

or

$$F = 25\text{ lb}.$$

5-8 Motion on an Inclined Plane

When a block of mass M is placed on an inclined plane, the forces acting on the block are due to the force of gravity and to the forces exerted on the block by the plane itself.

When there is no friction, the force exerted by the plane must be perpendicular to the surface of the plane, as illustrated in Figure 5-6(a). The

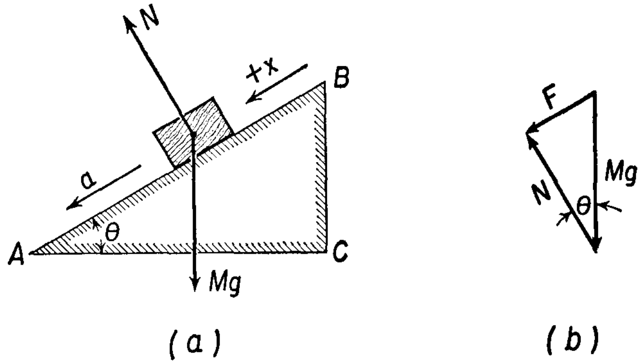


Fig. 5-6 Forces acting on a block placed on a frictionless inclined plane.

resultant of the force of gravity Mg and the normal force N of the plane must be a force F parallel to the plane, as shown in Figure 5-6(b) whose magnitude is given by

$$F = Mg \sin \theta,$$

where the angle θ is the angle the plane makes with the horizontal direction, and the direction of F is down the plane, as shown in the figure. Choosing the direction of the x axis as parallel to the plane, positive downward, we may find the acceleration of the block, from Equation (5-4a), as

$$Mg \sin \theta = Ma,$$

yielding

$$a = g \sin \theta. \tag{5-7}$$

The acceleration of a body on a frictionless inclined plane is down the plane and depends on the angle of inclination but is independent of the mass of the body. Looking at the same problem another way, we see that the acceleration g due to gravity is a vector quantity, directed vertically downward. The constraint of the plane prohibits such motion and only permits the body to move along the plane itself. The component of the acceleration of gravity along the plane is of magnitude $a = g \sin \theta$ directed down the plane.

When Galileo was studying the laws of motion, clocks of sufficient accuracy to time the motion of freely falling bodies were not yet available.

To slow down the motion so that it could be studied with available timing devices, he made use of the properties of a smooth inclined plane.

If a body is in motion on a rough inclined plane, the force of friction between the body and the plane affects the motion. The frictional force F_r acts so as to oppose the motion of the body along the plane. To determine the magnitude of the frictional force, we resolve the force of gravity W into components parallel and perpendicular to the plane, as shown in Figure 5-7, and find the parallel component to be of magnitude $W \sin \theta$ and the perpendicular component to be $W \cos \theta$.

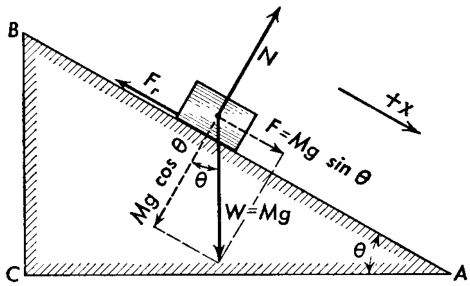


Fig. 5-7 Analysis of forces which act on a body that is sliding down a rough inclined plane.

The relationship between the frictional force and the normal force has been given as

$$F_r = fN,$$

so that

$$F_r = fMg \cos \theta,$$

where f is the coefficient of sliding friction between the body and the plane. If no other forces act on the body, its motion will be down the plane, and the frictional force will be directed up the plane, as shown in the figure. If we consider the x direction as parallel to the plane, positive downward, we find, on substituting in Newton's equation,

$$Mg \sin \theta - fMg \cos \theta = Ma,$$

$$a = g \sin \theta - fg \cos \theta. \tag{5-8}$$

Equation (5-8) becomes equivalent to Equation (5-7) when there is no friction.

If the angle of the plane is reduced to some critical value θ_c, the object will just slide down the plane with no acceleration. Equation (5-8) then gives

$$0 = g \sin \theta_c - fg \cos \theta_c,$$

from which

$$f = \tan \theta_c, \tag{5-9}$$

which is identical with a result obtained in Section 3-6.

Illustrative Example. A heavy wooden crate weighing 200 lb is pulled up a wooden plane, inclined at an angle of 20°, by a force of 150 lb. The coefficient of kinetic friction between the two surfaces is 0.30. (a) Determine the acceleration of the crate. (b) If the rope hauling it breaks, discuss the subsequent motion of the crate.

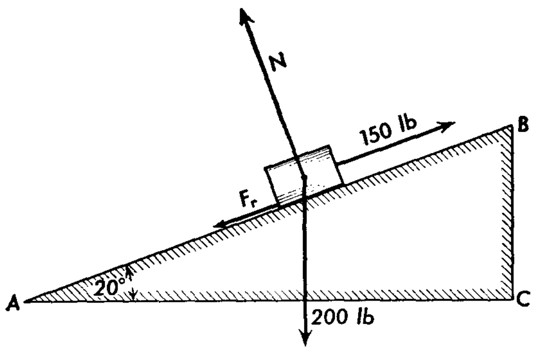


Fig. 5-8

(a) Figure 5-8 shows the forces which act on the crate as it moves up the plane. The force N which the plane exerts on the crate is perpendicular to its surface and is equal to the normal component of the weight, since there is no acceleration normal to the plane. Hence

$$N = W \cos\theta = 200 \text{ lb} \times 0.9397 = 187.9 \text{ lb}.$$

The force of friction is

$$F_r = fN = 0.3 \times 187.9 \text{ lb} = 56.4 \text{ lb}.$$

The component of the weight parallel to the plane is

$$F = W \sin\theta = 200 \text{ lb} \times 0.3420 = 68.4 \text{ lb}.$$

Choosing the direction of the x axis as parallel to the plane, positive downward, we find, from Equation (5-4a),

$$-150 \text{ lb} + 56.4 \text{ lb} + 68.4 \text{ lb} = \tfrac{200}{32} \text{ slug} \times a,$$

so that

$$a = -4.03 \frac{\text{ft}}{\text{sec}^2}\cdot$$

(b) When the rope breaks, the forces parallel to the plane which act on the crate are the component of the weight parallel to the plane and the frictional force. The direction of the frictional force is now reversed, since it always acts in a direction to oppose the motion. We find, substituting in Equation (5-4a),

$$-56.4 \text{ lb} + 68.4 \text{ lb} = \tfrac{200}{32} \text{ slug} \times a,$$

so that

$$a = +1.92 \frac{\text{ft}}{\text{sec}^2}\cdot$$

5-9 Motion through the Air

Objects falling through the air are acted upon by the resistance of the air as well as by the force of gravity. Some of this resistance may be due to the viscosity of the air and the rest to turbulence. This resistive force is often called *drag*, in connection with the flight of aircraft, and is the only reason an airplane does not cut off its engines after once assuming flying speed at a desired altitude. The effects of drag are also evident in the dust raised by the wind and in the transport of gravel and sand by flowing water. Drag is often put to use in engineering in the pneumatic conveying of grain and similar materials.

Experience shows that the resistance of air to motion through it increases as the velocity of the body increases. A body falling through the air for a sufficient time will ultimately reach a *terminal velocity*, at which time the force due to the resistance of the air is equal to the weight of the body. The body then continues to move downward with this limiting velocity.

For simplicity, let us assume that, for the case of a spherical body moving slowly through the air the resistance varies directly with the velocity. We may write

$$\mathbf{R} = K\mathbf{v}, \tag{5-10}$$

where K is a constant of proportionality depending on the cross-sectional area of the body and the viscosity of the air. As the velocity of fall increases, the magnitude of the force $\mathbf{R}$ increases until it becomes equal to the weight of the body. Thus

$$R = Kv_l = W,$$

from which

$$v_l = \frac{W}{K}, \tag{5-11}$$

where v_l is the limiting or the terminal velocity of the body.

Thus the terminal velocity of fall v_l of raindrops depends upon their weight. When raindrops reach the surface of the earth, the larger and heavier drops are moving faster than the light ones. The effective cross-sectional area of a man wearing an opened parachute is considerably greater than that of a man wearing a closed parachute, and the associated increase in K makes a very important difference to the man who is forced to leave an airplane in flight. The variation in K with cross-sectional area has been applied for centuries in the winnowing of grain to remove chaff and is today widely used in the cleaning of seed.

While the path of a freely falling projectile is parabolic, the path of a projectile in air is not. The speed of the projectile is steadily diminished by the resistance of the air. A baseball caught in the outfield is much

easier to catch than the same ball would have been if it could have been caught in the infield.

5-10 Pairs of Forces. Newton's Third Law

In our discussion of Newton's first and second laws, our attention was focused on one body on which a set of external forces acted. If we now analyze the origin of each of these forces, we find that each force is produced by the action of some other body on the one under discussion. If we push a trunk along the floor, the trunk exerts a force against our hands. If a ball is hit with a bat, not only is there a force exerted by the bat on the ball, but the ball also exerts a force on the bat. An automobile which is standing still pushes down on the ground at each of the surfaces of contact between its tires and the ground. At each region of contact, the ground exerts a force upward equal to that exerted by the car. *Newton's third law states that whenever one body exerts a force on another, the second body exerts a force equal in magnitude and opposite in direction on the first body.* This law is sometimes called the law of *action* and *reaction.*

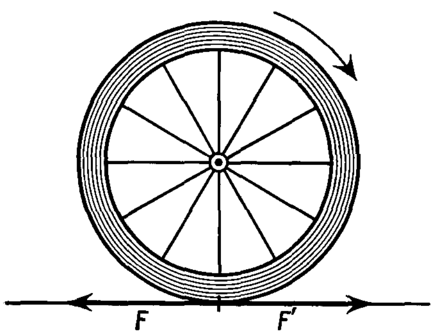


Fig. 5-9 Force F' exerted by the ground on the wheel is equal in magnitude but opposite in direction to the force F exerted by the wheel on the ground.

As an illustration of Newton's third law, consider the manner in which a car is set in motion. To start the car moving forward, there must be a net or unbalanced horizontal force acting on the car. To produce this horizontal force, the engine is started and then connected by means of gears and shafts to the rear wheels, causing them to turn in a clockwise direction, as shown in Figure 5-9. Because of the friction between the tires and the ground, the wheels exert a force F to the left (backward) on the ground; the ground exerts an equal and opposite force F' forward on the rear wheels. It is this horizontal force F' which makes the car go forward. To understand that it is the push of the ground on the driving wheels which makes the car go forward, just think of driving experiences on a winter day with ice on the ground, when the friction between the tires and the ground is very small. What usually happens is that the wheels

spin in a clockwise direction, but since there is little frictional force available, the wheels merely spin around, and the car does not move.

Problems

5-1. How big a force is required to give a 40-gm mass an acceleration of 150 cm/sec^2?

5-2. What constant force is required to give a body weighing 120 lb an acceleration of 4 ft/sec^2?

5-3. An automobile weighing 2,800 lb starting from rest acquires a speed of 40 mi/hr in 12 sec. Assuming that the acceleration is uniform, determine the unbalanced force which is acting on the automobile during this time.

5-4. A box whose mass is 350 gm rests on a table. A steady horizontal force is applied to this box. After 5 sec the box has acquired a speed of 40 cm/sec. Determine the force acting on the box.

5-5. A box of 800 gm mass is projected across a horizontal table with an initial speed of 150 cm/sec. It comes to rest on the table after having traversed a distance of 180 cm. Determine the frictional force opposing the motion.

5-6. A box whose mass is 12 kg is given an acceleration of 25 m/sec^2 on a horizontal surface. (a) Determine the resultant force acting on the box. (b) If the box starts from rest, determine the speed it will acquire in 8 sec.

5-7. A train weighing 450 tons has its speed increased from 20 mi/hr to 50 mi/hr in 15 sec. What force is supplied by the locomotive to produce this acceleration?

5-8. A steel cable supports an elevator weighing 2,500 lb. What is the tension in the cable when the elevator is moving (a) upward with a uniform velocity of 600 ft/min and (b) downward with a uniform velocity of 500 ft/min?

5-9. A steel cable supports an elevator weighing 1,800 lb. Starting from rest, the elevator acquires a velocity upward of 600 ft/min in 2 sec. (a) What is the resultant force acting on the elevator? (b) What is the tension in the cable?

5-10. The elevator of Problem 5-9, when going down, acquires a velocity of 500 ft/min in 2 sec. (a) What is the resultant force acting on the elevator? (b) What is the tension in the cable?

5-11. A steel ball whose mass is 250 gm is attached to the end of a cord. The ball is pulled upward with an acceleration of 120 cm/sec^2. Determine (a) the unbalanced force acting on the ball and (b) the tension in the cord.

5-12. A cube whose mass is 1,600 gm rests on a smooth table. A cord which is attached to the center of one face of the cube passes over a frictionless pulley at the edge of the table. A steel ball whose mass is 800 gm is fastened to the free end of the cord. Determine (a) the acceleration of each body and (b) the tension in the cord.

5-13. A box weighing 72 lb is placed on a smooth horizontal table. A cord which is connected to the center of one face of the box passes over a smooth pulley at the edge of the table. A steel ball weighing 24 lb is then fastened to the other end of the cord. Determine (a) the acceleration of each body and (b) the tension in the cord.

5-14. Two boxes, one weighing 16 lb and the other weighing 4 lb, are attached to the ends of a cord. The cord is placed over a frictionless pulley which is free to rotate about a horizontal axis. Determine the acceleration of each box.

5-15. A cord passes over a fixed frictionless pulley. A cylinder whose mass is 3 kg is suspended from one end of the cord, and another cylinder whose mass is 2 kg is suspended from the other end. Determine (a) the acceleration of the system and (b) the tension in the cord.

5-16. A series of frictionless inclined planes all have the same heights but have different lengths. Show that the time required for an object to slide down any of these planes is directly proportional to the length of the plane.

5-17. A car weighing 3,000 lb and moving with a speed of 20 mi/hr reaches a hill having a 5 per cent grade and starts coasting downhill. Determine (a) the component of the weight acting downhill and (b) the speed the car will acquire if it coasts for 400 ft, assuming friction is negligible. [NOTE: A hill having a 5 per cent grade is one which rises 5 ft for every 100 ft of length.]

5-18. A body whose mass is 3 kg is projected up an inclined plane with an initial velocity of 5 m/sec. The plane is inclined at an angle of 30° to the horizontal, and the coefficient of kinetic friction between the plane and the body is 0.2. Determine (a) how far up the plane the body will go before coming to rest, (b) its acceleration down the plane, and (c) the speed it will have when it reaches its starting point.

5-19. A box slides down a 30° inclined plane with an acceleration of 4 ft/sec^2. Determine the coefficient of friction between the box and the plane.

5-20. A box whose mass is 18 kg rests on a table. A cord tied to this box passes over a frictionless pulley at the edge of the table. A cylinder whose mass is 6 kg is hung from the free end of the cord. The coefficient of friction between the box and the table is 0.25. Determine (a) the acceleration of the box, (b) the tension in the cord, and (c) the distance the cylinder will move in 3 sec.

5-21. A boy takes a running start with a sled and acquires a speed of 8 ft/sec. If the coefficient of friction between sled and snow is 0.10, how far will the sled move on a level road before coming to rest?

5-22. A boy coasts down a hill on a sled, reaching level ground with a speed of 30 ft/sec. If the coefficient of friction between the steel runners and the snow is 0.05 and the boy and sled weigh 150 lb, find how far the sled will travel before coming to rest.

5-23. Show that if the force due to the resistance of the air varies with the square of the velocity of a falling body, the limiting velocity of fall is proportional to the square root of the weight of the body.

5-24. Two men, one weighing 180 lb and the other weighing 120 lb, are on ice skates. Each holds one end of a taut rope. The heavier man exerts a force of 20 lb on the rope. (a) How big a force does the lighter man exert? (b) What is the acceleration of each man? Neglect friction.

5-25. A 5-gm bullet is fired from a gun whose barrel is 60 cm long. The bullet leaves the gun with a muzzle velocity of 2,500 cm/sec. What was the average force acting on the bullet?

5-26. A man weighing 150 lb stands on a platform weighing 42 lb. The platform is suspended by a rope which passes over a frictionless pulley. The man

pulls down on the free end of the rope to lift himself and the platform. (a) With what force must he pull on the rope if the system consisting of the man and the platform is to receive an upward acceleration of 3 ft/sec^2? (b) What is the maximum acceleration with which the man can raise the platform and still stay on the platform?

5-27. A rope inclined at an angle of 37° with the horizontal is used to drag a 50-kg block along a level floor with an acceleration of 1 m/sec^2. The coefficient of friction between the block and the floor is 0.2. What is the tension in the rope?

5-28. A projectile of mass 5 gm is fired from a gun with a muzzle velocity of 2,500 cm/sec directed due east at an angle of 45° with the horizontal. A wind is blowing from the north, exerting a steady force of 1,000 dynes against the projectile. Find the position of the projectile when it strikes the ground.

5-29. A body of mass 100 kg is hung from a rope which is passed over a frictionless pulley to a man on the ground who is interested in raising the body a distance of 25 m in the shortest possible time. The pulley is hung from the ceiling by a chain whose breaking strength is 2,000 nt. What is the shortest time in which the body can be raised?

5-30. Two bodies, each weighing 10 lb, are connected by a cord which passes over a light frictionless pulley. What vertical force must be applied to the pulley to raise the system with an acceleration of 5 ft/sec^2?

5-31. Two bodies, weighing 10 lb and 20 lb, rest upon a table. The two bodies are connected by a cord which passes over a light frictionless pulley. What is the least vertical force which can be applied to the pulley (a) to raise the 10-lb weight? (b) To raise the 20-lb weight?

5-32. If in Problem 5-31 a force of 50 lb is applied to raise the pulley, what will be the acceleration of (a) the 10-lb weight and (b) the 20-lb weight?

5-33. A block weighing 5 lb rests on a horizontal surface. The coefficient of friction between the block and the surface is 0.2. A horizontal force of 2 lb is applied to the block. (a) What is the acceleration of the block? (b) The system consisting of block, table, and applied force is placed on an elevator which rises at a constant speed of 5 ft/sec. What is now the acceleration of the block? (c) The elevator is then brought to a stop with a uniform acceleration of 4 ft/sec^2. During this period of vertical acceleration what is the horizontal acceleration of the block?

5-34. A pendulum bob weighing 1 lb is hung from the roof of a railroad car. The train is started with a constant acceleration of 3.2 ft/sec^2. (a) At what angle with the vertical does the pendulum bob hang? (b) What is the tension in the string?

5-35. In a train moving with constant acceleration it is observed that the chandeliers hang at an angle of 0.57° with the vertical. The train starts from rest. With what velocity is the train moving at the end of 2 minutes?

5-36. A mass of 10 kg and a second mass of 5 kg are connected by a string and rest on a horizontal frictionless table. A constant pull of 60 nt is applied to the 5 kg mass. Find (a) the acceleration of the system and (b) the tension in the string.

6

Circular Motion and Gravitation

6-1 Circular Motion

Our earlier discussion of the kinematics of a particle was developed principally from the point of view of being able to describe that motion easily within a rectangular coordinate system. Thus the most complex case with which we dealt was that of a projectile motion, in which the acceleration was constant and was directed along one of the coordinate axes. A more convenient framework within which to discuss rotational and circular motions is provided by a set of polar coordinates. In the present discussion we will restrict ourselves to motion in which the polar coordinate r is constant, or fixed; that is, the particle is constrained to move in a circular path.

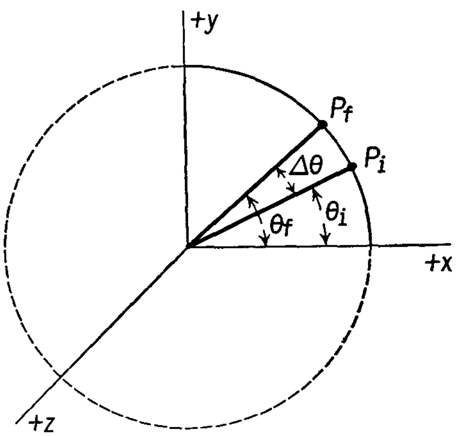


Fig. 6-1 Angular displacement.

6-2 Angular Displacement

When a particle is constrained to move in a circular path, it is convenient to superimpose a coordinate system on the motion so that the x-y plane is in the plane of the circle and the origin of coordinates lies at the center of the circle. If the particle is initially at the point P_i at angle θ_i and is finally at position P_f at angle θ_f, we say that its *angular displacement* $\Delta\,\theta$ is given by its final angular coordinate minus its initial angular coordinate. In the form of an equation we have

$$\Delta\theta = \theta_f - \theta_i, \tag{6-1}$$

as shown in Figure 6-1. In polar coordinate notation, angles are measured counterclockwise from the x axis. A positive angular displacement implies that the particle has been displaced in a counterclockwise direction around the circle, while a negative displacement implies a clockwise displacement.

An angular displacement may be expressed in degrees or in radians or in revolutions. As we have previously seen, 1 revolution corresponds to 360° or to 2π radians, so that

$$1 \text{ radian} = \frac{360°}{2\pi} = 57.3°.$$

6-3 Angular Speed and Angular Velocity

The *average angular speed* $\bar{\omega}$ (omega bar) is defined as the quotient of the angular displacement $\Delta\theta$ divided by the time interval Δt in which that displacement took place. In the form of an equation, we have

$$\bar{\omega} = \frac{\Delta\theta}{\Delta t}. \qquad \text{(6-2)}$$

When $\Delta\theta$ is expressed in radians and Δt is expressed in seconds, $\bar{\omega}$ is given in units of radians per second. Other appropriate units may be used.

The *instantaneous angular speed* at a point ω is obtained by a limiting process analogous to the process used in the definition of the instantaneous linear speed. Without going through the details of that process, we may write

$$\boxed{\omega = \lim_{\Delta t \to 0} \frac{\Delta\theta}{\Delta t} = \frac{d\theta}{dt}.} \qquad \text{(6-3)}$$

If the angular speed is constant, the average angular speed $\bar{\omega}$ is equal to the instantaneous angular speed ω. A particle whose angular speed is changing is said to have angular acceleration.

A particle on the rim of a spinning wheel is moving in circular motion with the angular speed of the wheel, but the angular speed alone is not a sufficient description of the motion, for, in order to know where the particle will be at some subsequent time, it is necessary to know the axis about which the wheel is rotating. To describe more completely a rotational motion with constant angular speed, it is necessary to specify both the angular speed and the direction of the axis of rotation.

We may describe the *angular velocity* of a particle in circular motion as a *vector quantity* whose magnitude is its angular speed and whose direction is along the axis about which the particle is rotating. The sense of the angular-velocity vector is given by the *right-hand rule* used to describe the

torque vector. If the bent fingers of the right hand are pointed in the direction of rotation of the particle, the outstretched thumb indicates the direction of the angular-velocity vector along the axis of rotation. Thus in Figure 6-1 the angular-velocity vector of the particle rotating about the z axis from P_i to P_f would be in the positive z direction.

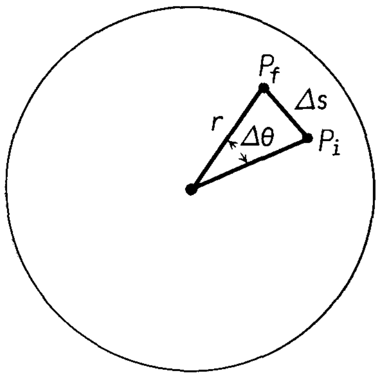


Fig. 6-2

6-4 Angular Speed and Linear Speed

The angular speed of every particle of a rotating disk is equal to the angular speed ω of the disk, but the instantaneous linear speed of each particle depends upon its position. The linear speed of a particle is given by the distance traversed divided by the elapsed time. In a short time interval Δt, a particle of the disk initially at a point P_i has been displaced to the position P_f, through a small arc of length Δs, while the angular displacement of every particle of the disk has been $\Delta\theta$, as shown in Figure 6-2. We have

$$\Delta\theta = \frac{\Delta s}{r},$$

where r is the radial distance of the particle from the axis of rotation. Thus

$$\Delta s = r\,\Delta\theta,$$

and, dividing both sides of the equation by the time interval Δt, we obtain

$$\frac{\Delta s}{\Delta t} = r\frac{\Delta\theta}{\Delta t},$$

from which we see that the instantaneous linear speed v with which the particle moves is the product of r by its angular speed ω. In the form of an equation we have

$$\boxed{v = \omega r.} \qquad \textbf{(6-4)}$$

From Equation (6-4) we see that the particles near the rim of a disk which rotates about a fixed axis move with greater linear speed than do particles which are nearer the axis of rotation. The particles on the axis of rotation have zero linear speed; they are at rest.

Illustrative Example. An airplane propeller is rotating with uniform angular

speed of 1,800 rpm. The blades of the propeller are 6 ft long. Determine the linear speed of a point (a) 2 ft from the axis and (b) 6 ft from the axis.

Expressing the angular speed in radians per second, we get

$$\omega = \frac{2\pi \times 1{,}800}{60}\,\frac{\text{radians}}{\text{sec}} = 188.4\,\frac{\text{radians}}{\text{sec}}.$$

(a) Using Equation (6-4) with $r = 2$ ft, we get

$$v = 188.4\,\frac{\text{radians}}{\text{sec}} \times 2\ \text{ft} = 376.8\,\frac{\text{ft}}{\text{sec}};$$

and when $r = 6$ ft, we get

(b) $$v = 188.4\,\frac{\text{radians}}{\text{sec}} \times 6\ \text{ft} = 1{,}130\,\frac{\text{ft}}{\text{sec}}.$$

As discussed in Chapter 1, the term "radian" has no physical dimensions since it is the ratio of two lengths; hence such a unit as $\frac{\text{radians}}{\text{sec}} \times$ ft is equivalent to the unit $\frac{\text{ft}}{\text{sec}}$.

6-5 Angular Acceleration

Since angular velocity is a vector quantity, it can change in either direction or magnitude or in both. An airplane generally cruises with its engines turning at a steady rate. Its propellers therefore rotate at a fixed angular speed. When the airplane makes a turn, the angular speed of the propellers remains fixed, but the direction of the axis of rotation is changing and therefore the angular velocity is changing, with important consequences which will be discussed in a later section. In this section we shall consider changes in the magnitude of the angular velocity only. Just as for linear acceleration, we shall define the angular acceleration as the change in angular velocity in a time interval Δt.

We define the *average angular acceleration* $\bar{\boldsymbol{\alpha}}$ (alpha bar) as the change in angular velocity $\Delta\boldsymbol{\omega}$ divided by the time interval Δt in which that change takes place. In the form of an equation we have

$$\bar{\boldsymbol{\alpha}} = \frac{\Delta\boldsymbol{\omega}}{\Delta t}. \tag{6-5}$$

The instantaneous angular acceleration $\boldsymbol{\alpha}$ is again obtained by a limiting process, and we may write

$$\boxed{\boldsymbol{\alpha} = \lim_{\Delta t \to 0} \frac{\Delta\boldsymbol{\omega}}{\Delta t} = \frac{d\boldsymbol{\omega}}{dt}.} \tag{6-6}$$

Since the angular acceleration is given by the result of dividing $\Delta\boldsymbol{\omega}$, a vector, by Δt, a scalar, the angular acceleration $\boldsymbol{\alpha}$ is a vector quantity. In the present chapter we shall deal only with the case in which the motion takes place about a fixed axis. The angular-velocity vector will always be directed parallel to that axis. Changes in the angular velocity will therefore also be directed parallel to the axis of rotation. Thus in this case the angular-acceleration vector must be parallel to the axis of rotation.

6-6 Equations of Motion for Constant Angular Acceleration

The defining equations for linear motion and for angular motion are identical except for the symbols used. In the equations representing linear motion, the symbols x, v, and a need only be replaced by θ, ω, and α to obtain the corresponding equations for angular motion. From the definitions of velocity and acceleration, a few mathematical manipulations enabled us to obtain equations relating such quantities as displacement, velocity, acceleration, and time, which were of considerable usefulness in describing the motion of a particle. For rotational motion it will not be necessary to repeat the development of these equations, for we have deciphered the code which enables us to translate freely from angular motion to linear motion. To translate equations developed for linear motion to equations appropriate for use in angular motion, we simply replace the symbol x by θ, v by ω, and a by α. The symbol u which was used to represent the initial linear velocity will be replaced by the symbol ω_i to represent initial angular velocity.

We display the correspondence between the equations representing linear motion and those representing angular motion as follows:

Linear Motion	Angular Motion	
$x = \bar{v}t$	$\theta = \bar{\omega}t$	(6-7a)
$v = u + at$	$\omega = \omega_i + \alpha t$	(6-7b)
$x = ut + \frac{1}{2}at^2$	$\theta = \omega_i t + \frac{1}{2}\alpha t^2$	(6-7c)
$v^2 = u^2 + 2ax$	$\omega^2 = \omega_i^2 + 2\alpha\theta$	(6-7d)
$\bar{v} = \dfrac{u+v}{2}$	$\bar{\omega} = \dfrac{\omega_i + \omega}{2}$	(6-7e)

Both sets of equations are for uniformly accelerated motion. The equations for linear motion represent motion in a straight line, here chosen as the x axis, with constant acceleration a, while the equations for angular motion represent rotation about a fixed axis with constant angular acceleration α. Although, to avoid complication, we have not discussed the com-

ponent form of the equations for rotational motion, it is evident that both sets of equations may also be thought of as the component form of the general vector equations for the motion of a particle, with the linear equations representing the x component of the motion, while the angular equations represent the components along an unspecified axis of rotation.

Illustrative Example. The angular velocity of an airplane propeller is increased from 1,800 rpm to 2,200 rpm in 10 sec. Determine (a) its angular acceleration and (b) the angle traversed during this period.

(a) We may apply the definition of angular acceleration directly with the units given in the problem and get

$$\alpha = \frac{2{,}200 - 1{,}800}{10}\,\frac{\text{rpm}}{\text{sec}} = 40\,\frac{\text{rev}}{\text{min sec}},$$

and we may find the total rotation by the application of Equations (6-7) with these units as

$$\theta = \frac{2{,}200 + 1{,}800}{2}\,\frac{\text{rev}}{\text{min}} \times \frac{1}{6}\text{ min}$$
$$= 333 \text{ rev},$$

or (b) we may express the angular velocities in radians per second and apply Equations (6-7). The final angular velocity is

$$\omega = \frac{2\pi}{60} \times 2{,}200\,\frac{\text{radians}}{\text{sec}},$$

and the initial angular velocity is

$$\omega_i = \frac{2\pi}{60} \times 1{,}800\,\frac{\text{radians}}{\text{sec}},$$

so that

$$\alpha = \frac{\dfrac{2\pi}{60}(2{,}200 - 1{,}800)\,\dfrac{\text{radians}}{\text{sec}}}{10\text{ sec}},$$

$$\alpha = 4.19\,\frac{\text{radians}}{\text{sec}^2};$$

and
$$\theta = \frac{\dfrac{2\pi}{60}(2{,}200 + 1{,}800)\,\dfrac{\text{radians}}{\text{sec}}}{2} \times 10\text{ sec},$$

$$\theta = 2{,}090 \text{ radians}.$$

6-7 Uniform Circular Motion

In our previous study of accelerated motion, we restricted ourselves to the motion produced by the action of a force which remains constant in mag-

nitude and direction throughout the motion. The particular path followed by the body depended upon the relationship of the direction of the initial velocity and the direction of the force. When the initial velocity was parallel to the force, the motion was linear; when the initial velocity was at any other angle to the direction of the force, the motion was parabolic. We shall now consider another important type of motion, that of *uniform motion in a circle;* that is, motion in a circular path in which the speed of the body remains constant in magnitude, but where the direction in which the body is moving changes constantly. We shall show that this type of motion is produced by a force which is always constant in magnitude but continually changing in direction in such a way that it is always at right angles to the velocity of the body.

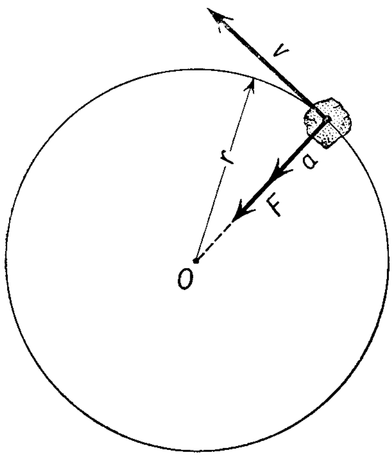


Fig. 6-3 Force **F** acting on a particle which is moving with uniform motion in a circular path is directed toward the center of the circle.

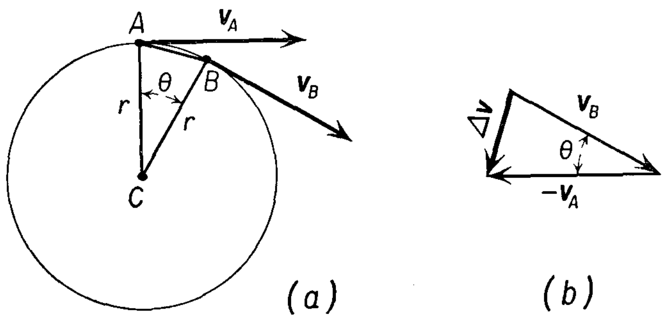
Consider a particle moving with velocity **v** in a horizontal circular path of radius r, as in Figure 6-3. The direction of the velocity is always tangent to the path. If the force **F** acts at right angles to the velocity, it can have no component in the direction of the velocity; that is, it cannot change the speed of the body. The only effect of the force is to change the direction of the motion. Since the radius of a circle is always at right angles to the tangent, the force **F** must act along a radius. Furthermore, the force must be directed *toward the center of the circle.* According to Newton's first law, a particle moves with uniform velocity **v** in a straight line when the resultant force on it is zero, and such a motion would take it away from the center of the circle. To bring it back toward the center requires that there be an acceleration toward the center, and hence there must be an unbalanced force acting toward the center of the circle. One way of supplying such a force is to tie one end of a flexible string to the particle and to tie the other end to a pin at the center of the circle. A flexible string is used because such

a string can support only a tension, hence it must exert a pull toward the center on the particle.

6-8 Centripetal Acceleration

We can derive the expression for the acceleration of a particle moving in uniform circular motion in terms of the speed of the particle and the radius of the circle by considering two neighboring positions A and B of a particle moving with uniform speed in a circle of radius r, as shown in Figure 6-4(a).

Fig. 6-4



The velocity of the particle at A is $\mathbf{v}_A$ and its velocity at B is $\mathbf{v}_B$. The change in velocity $\Delta\mathbf{v}$ of the particle is the vector difference of the two velocities $\mathbf{v}_B - \mathbf{v}_A$. To find the difference between these two velocities, we add $-\mathbf{v}_A$ to $\mathbf{v}_B$, as shown in Figure 6-4(b). If Δt is the time to go from A to B, the average acceleration $\bar{\mathbf{a}}$ of the particle is

$$\bar{\mathbf{a}} = \frac{\mathbf{v}_B - \mathbf{v}_A}{\Delta t} = \frac{\Delta\mathbf{v}}{\Delta t}\cdot$$

Now the triangle formed by these vectors and the triangle ABC are similar, since they are both isosceles triangles and have equal angles θ. We may therefore write

$$\frac{\Delta v}{v_A} = \frac{\overline{AB}}{\overline{AC}} = \frac{\overline{AB}}{r}\cdot$$

Dividing both sides of the equation by Δt and multiplying both sides of the equation by v_A, we obtain

$$\frac{\Delta v}{\Delta t} = \frac{\overline{AB}}{\Delta t \times r}\, v_A,$$

so that

$$\bar{a} = \frac{\overline{AB}}{\Delta t \times r}\, v_A.$$

To find the instantaneous acceleration of the particle, we take the limit of $\bar{a}$ as Δt approaches zero; the points A and B approach each other, and the chord $\overline{AB}$ approaches the length of the arc $\widehat{AB}$; the velocity v_A becomes equal to the velocity v at the limiting point, and the distance $\overline{AB}$ may be replaced by $v\,\Delta t$. Substituting these values into the previous equation, we find

$$\boxed{a = \frac{v^2}{r}\cdot} \tag{6-8}$$

A glance at Figure 6-4 shows that when the angle θ is very small, the vector $\Delta\mathbf{v}$ is at right angles to the velocity $\mathbf{v}_A$ and is directed toward the center of the circle.

When an object moves in uniform circular motion with linear speed v, *it is accelerated toward the center of the circle with an acceleration* a *given by* a = v²/r.

We call the acceleration of a particle moving in uniform circular motion *centripetal acceleration;* the word "centripetal" means "directed toward the center."

A particle moving in uniform circular motion with constant linear speed v may also be described as moving with uniform angular speed ω with the linear and angular speed related by Equation (6-4) as $v = \omega r$, where r is the radius of the circle in which the particle is moving. We may therefore rewrite Equation (6-8) in terms of ω as

$$a = \frac{v^2}{r} = \omega^2 r. \tag{6-9}$$

6-9 Centripetal and Centrifugal Forces

Knowing the magnitude and the direction of the acceleration of a particle moving in uniform circular motion, it is a simple matter to compute the force which acts on the particle. If the mass of the particle is m, we find, upon application of Newton's second law,

$$F = \frac{mv^2}{r} = m\omega^2 r \tag{6-10}$$

for the magnitude of the force acting toward the center to keep the body moving with uniform circular motion. The force acting toward the center is called the *centripetal* force. It must be remembered that the centripetal force acts *on* a body moving in a circular path.

Referring again to the particle that is being whirled around at the end

of a string, as in Figure 6-3, we note that the centripetal force is the pull of the string on the particle. From Newton's third law, the particle exerts an equal and opposite force on the string. This reaction is sometimes referred to as *centrifugal* force, meaning directed away from the center. If the string is cut during the circular motion, there will no longer be a force acting on the particle. The instant the string is cut, the particle will continue to move in the direction it was moving at the time; that is, the particle will go off at a tangent to the circle.

In ordinary conversation there is a great deal of confusion about the terms "centripetal force" and "centrifugal force." If one asks why the particle flies off when the string is cut, the answer seems always to be centrifugal force. If the question is raised as to why the passenger in a car seems to be thrown against the door when the car goes around a curve, the same answer invariably is given—centrifugal force. In terms of Newton's laws of motion, this is obviously incorrect. An observer, viewing the car turning the corner from some perspective above the car, would observe that the passenger, who has little or no radial force exerted on him by the slippery seat covers, moves in a straight line with uniform speed, while the car, which is acted on by the frictional force of the road, moves in a circular path. When the trajectories of the passenger and the side of the car intersect, the passenger says that he has been thrown against the door. The passenger views the world as though he were at its center, and if he finds himself in contact with the door, he assumes that he must have been forced toward it.

Illustrative Example. A stone weighing 0.5 lb tied to a string 2 ft long is placed on a smooth horizontal table. The other end of the string is tied to a pin at the center of the table. The stone is given a push for a short time, and acquires a speed of 6 ft/sec. (a) Determine the tension in the string. (b) If the breaking strength of the string is 15 lb, determine the maximum speed with which the stone can be whirled.

(a) The mass of this stone is

$$m = \frac{0.5 \text{ lb}}{32 \text{ lb/slug}} = \frac{1}{64} \text{ slug}.$$

The centripetal force F required to keep it moving in a horizontal circle of radius $r = 2$ ft with a speed of $v = 6$ ft/sec is, from Equation (6-10),

$$F = \frac{1}{64} \text{ slug} \times \frac{36 \text{ ft}^2/\text{sec}^2}{2 \text{ ft}},$$

so that

$$F = 0.28 \text{ lb}.$$

(b) If the breaking strength of the string is 15 lb, this represents the maximum centripetal force that the string can apply to the stone. Using this value for

F in Equation (6-10) and letting v be the unknown maximum speed of the stone, we get

$$15 \text{ lb} = \frac{1}{64} \text{ slug} \times \frac{v^2}{2 \text{ ft}},$$

from which

$$v^2 = 1{,}920 \frac{\text{ft}^2}{\text{sec}^2},$$

and

$$v = 43.9 \frac{\text{ft}}{\text{sec}}.$$

6-10 Banking of a Curved Road

A car rounding a curve can be considered as moving in an arc of a circle or, in some cases, in a series of such arcs of slightly different radii. In order to move the car in a circular path of radius r, an outside force must act on the car, and this force must be directed toward the center of the circle. In the

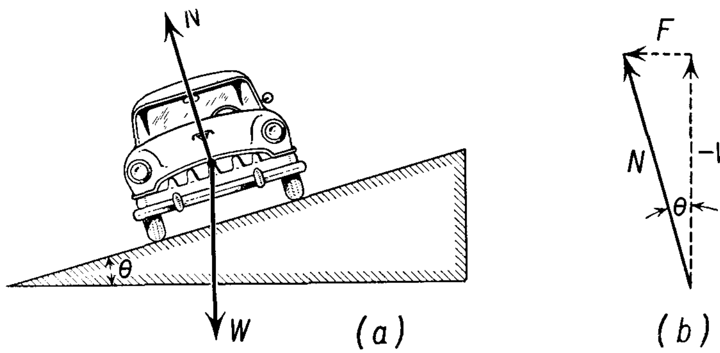


Fig. 6-5 (a) Car moving on a curve in a road banked at an angle θ to the horizontal. (b) The components of the normal force N are $-W$ and F.

case of an automobile rounding a curve, this force is supplied by the friction between the road and the tires. The frictional force which can be supplied by the contact between the road and the wheels of a car is variable, depending on the conditions of the road and the tires of the vehicle. It is much more desirable for design purposes to pretend that the road is smooth and to bank the road so that the normal force exerted by a smooth road has a horizontal component of magnitude and direction equal to the required centripetal force. The angle at which the road is banked will depend on the speed of the vehicle. Thus when a railroad curve is marked with a speed of, say, 40 mi/hr, this does not mean the maximum speed at which the turn should be traversed, but rather the speed for which the turn was

designed. In general, small variations from the design speed are compensated for by frictional forces.

An airplane moving through the air is affected by the lift of the air on the wings. In level flight this force is directed vertically upward. An airplane may execute a turn by banking in such a manner that the force on the wings has a horizontal component of the desired magnitude and direction to supply the centripetal force necessary for the circular motion.

In Figure 6-5(a), if a car is to move in a circular path, the normal force N exerted on the car by the road must both support the weight of the car W and supply a horizontal component F equal to the required centripetal force. From Figure 6-5(b) we see that the horizontal component is

$$F = N \sin \theta = \frac{mv^2}{r},$$

and the vertical component is

$$-W = N \cos \theta = mg.$$

From these equations we find that

$$\tan \theta = \frac{v^2}{rg}. \tag{6-11}$$

Essentially the same analysis can be applied to the banked turn of an airplane, as illustrated in Figure 6-6. We note from Equation (6-11) that the angle of inclination of the road, or the angle at which an airplane should be banked upon making a turn, depends only upon the speed and the radius of the turn and does not depend on the mass of the vehicle. It is this latter consideration which makes it possible to bank highways.

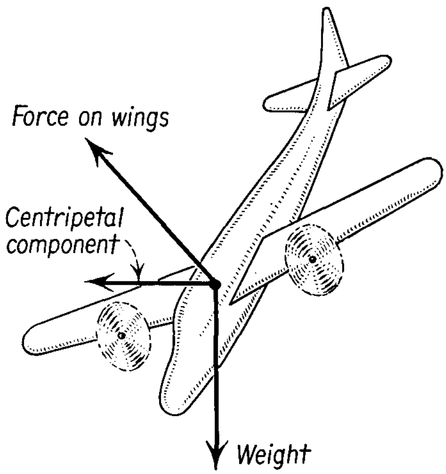


Fig. 6-6 The banked turn of an airplane.

Illustrative Example. A truck weighing 2.5 tons rounds a curve in a level road at 30 mi/hr. The curve is in the form of a circular arc of 1,200 ft radius. (a) Determine the lateral force exerted by the road on the tires to keep the truck moving in its circular path. (b) Find the angle of banking if the turn is to be executed without friction.

(a) The road must exert a force toward the center of the circle in which the car is moving of magnitude $F = \dfrac{mv^2}{r}$. Now

$$m = \frac{W}{g} = \frac{2.5 \times 2{,}000}{32}\ \text{slugs}; \qquad v = 30\,\frac{\text{mi}}{\text{hr}} = 44\,\frac{\text{ft}}{\text{sec}}; \qquad r = 1{,}200\ \text{ft};$$

substituting these values, we get

$$F = \frac{2.5 \times 2{,}000}{32}\text{ slugs} \times \frac{\left(44\,\dfrac{\text{ft}}{\text{sec}}\right)^2}{1{,}200\text{ ft}},$$

from which $$F = 252\text{ lb} = 0.126\text{ ton}.$$

(b) The angle of banking may be determined from Equation (6-11). Substituting numerical values, we have

$$\tan\theta = \frac{v^2}{rg} = \frac{(44\text{ ft/sec})^2}{1{,}200\text{ ft} \times 32\text{ ft/sec}^2}$$

$$= 0.0504,$$

$$\theta = 2.9°.$$

6-11 Motion in a Vertical Circle

When circular motion takes place in a vertical plane, as, for example, when an airplane loops the loop, the motion is not uniform, and the speed varies from point to point on the circle. Consider a particle which acquires speed

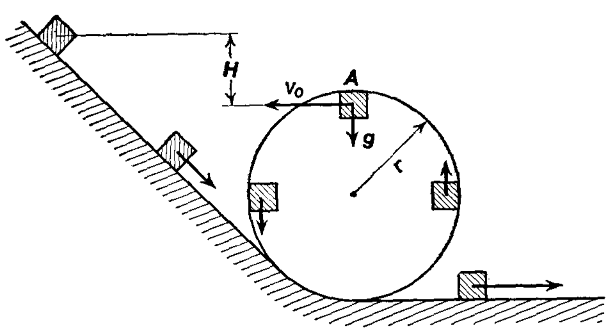


Fig. 6-7

by sliding down a frictionless inclined plane, as shown in Figure 6-7, and then starts up the inside of a vertical circular track. It is obvious that the danger point is the highest point A of the track. The particle must negotiate this point with the proper speed if it is to travel safely around the track.

Suppose the particle is at point A under the track. The particle will stick to the track as long as it moves fast enough so that the track is required to exert some force against it to provide the necessary centripetal acceleration. The minimum speed the particle may have is one in which the force exerted by the track is zero. At this critical speed v_0, the only force acting on the particle is the force of gravity. Consequently, the

acceleration of gravity g must be equal to the required centripetal acceleration; that is,

$$g = \frac{v_0^2}{r},$$

from which

$$v_0 = \sqrt{rg}.$$

If the speed of the particle is greater than this minimum speed, its acceleration toward the center will be greater than g; this means that the

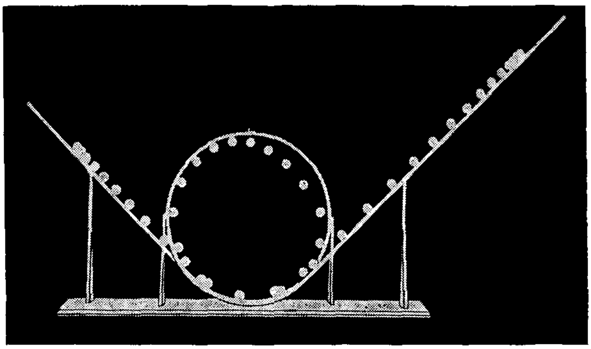


Fig. 6-8 Multiflash photograph of a ball which starts on an inclined plane but does not acquire sufficient speed to loop the loop in a vertical circle. (Reproduced by permission from *College Physics*, 2nd ed., by Sears and Zemansky, 1952; Addison-Wesley Publishing Company, Inc., Reading, Mass.)

track will have to exert a force toward the center to keep it moving in the circular path; the particle will stick to the track. If its speed is less than the minimum safe value of v_0, the particle will leave the track and follow the parabolic path of a projectile (see Figure 6-8).

6-12 Angular Acceleration and Linear Acceleration

When a rigid body rotates about a fixed axis with constant angular acceleration α, each particle of the body has the same angular acceleration, but the linear acceleration $\mathbf{a}$ of each particle will be different. The linear acceleration $\mathbf{a}$ of a particle may be resolved into two components, one component $\mathbf{a}_t$ tangent to the path and given by the time rate of change of the *speed* of the particle, and the other component $\mathbf{a}_c$, the centripetal acceleration, directed perpendicular to the velocity and parallel to the radius of the circle in which the particle is moving.

The relationship between the length of arc s subtended by an angular displacement θ on a circle of radius r is well known as the basis of definition

of the radian as

$$s = \theta r.$$

Differentiating this equation once with respect to time for circular motion of constant r, we obtain

$$\frac{ds}{dt} = \frac{d\theta}{dt} r.$$

By definition, the instantaneous linear speed v is equal to $\frac{ds}{dt}$, while the instantaneous angular speed ω is equal to $\frac{d\theta}{dt}$, and we may rewrite this equation as

$$v = \omega r,$$

which we have seen before as Equation (6-4). If we differentiate once again with respect to time, neglecting the change in direction of **v** and considering only the change in magnitude, that is, the change in the angular speed of the particle, we obtain

$$\frac{dv}{dt} = \frac{d\omega}{dt} r.$$

By definition, the rate of change of the linear speed is the component of the acceleration measured along the path, hence the *tangential acceleration* a_t, while the rate of change of the angular velocity is the angular acceleration α. Thus we find

$$a_t = \alpha r. \tag{6-12}$$

We have already seen that the central, or radial, component of the acceleration which is due to the change in direction of **v** may be expressed in terms of the angular velocity from Equation (6-9) as

$$a_c = \frac{v^2}{r} = \omega^2 r. \tag{6-13}$$

The tangential component of the acceleration depends upon the angular acceleration α, while the central component of the acceleration depends upon the angular velocity ω. Both the tangential and the central, or radial, components of acceleration depend upon the distance of the particle from the axis of rotation. If the particle is moving with uniform speed in a circle, the radial acceleration is the total acceleration. If the speed is increasing or decreasing while the particle is moving in a circle, then its total acceleration **a** is made up of two components $\mathbf{a}_t$ and $\mathbf{a}_c$. Since these two components are at right angles to each other, we have

$$a = \sqrt{a_t^2 + a_c^2}. \tag{6-14}$$

Illustrative Example. A circular pulley 4 ft in diameter is mounted so that it can rotate about an axis passing through its center. One end of a cord which is

wound around the pulley is being pulled in a horizontal direction, as shown in Figure 6-9, with an acceleration of 6 ft/sec^2. (a) Determine the angular acceleration of the pulley. (b) Assuming the pulley to have been at rest initially, determine the acceleration of the lowest point B on the rim of the pulley at the end of 10 sec.

(a) The point A at the top of the pulley is the point at which the rope just leaves the pulley. The tangential component of the acceleration a_t of this point is the same as the acceleration of a point on the rope. We can, therefore, use the acceleration of this point for determining the angular acceleration of the pulley. From Equation (6-12) we get

$$\alpha = \frac{a_t}{r} = \frac{6\ \text{ft/sec}^2}{2\ \text{ft}} = \frac{3\ \text{radians}}{\text{sec}^2}\,.$$

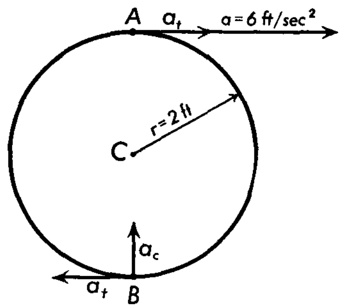


Fig. 6-9

(b) The angular speed of the pulley at the end of 10 sec is, from Equations (6-7),

$$\omega = \omega_i + \alpha t.$$

Since the pulley starts from rest, $\omega_i = 0$, and we have

$$\omega = 3\,\frac{\text{radians}}{\text{sec}^2} \times 10\ \text{sec} = 30\,\frac{\text{radians}}{\text{sec}}\,.$$

Since the angular acceleration is clockwise, the lowest point on the rim of the wheel is moving toward the left. The two components of the acceleration at this instant are

$$a_t = \alpha r = 6\,\frac{\text{ft}}{\text{sec}^2}\,,\ \text{to the left,}$$

and

$$a_c = \omega^2 r = \left(30\,\frac{\text{radians}}{\text{sec}}\right)^2 \times 2\ \text{ft} = 1{,}800\,\frac{\text{ft}}{\text{sec}^2}\,.$$

6-13 Periodicity of Uniform Circular Motion

One of the interesting properties of uniform circular motion is that it is *periodic;* that is, a particle in uniform circular motion traverses a full circumference in a time T. This time T is called the *period* of the motion. The period remains the same no matter how often the motion is repeated.

In one complete revolution the angular displacement is 2π radians, and the distance traversed is $2\pi r$ where r represents the radius of the circle. The period, or time, for one complete revolution is therefore equal to

$$T = \frac{2\pi r}{v}\,, \tag{6-15}$$

or

$$T = \frac{2\pi}{\omega}\,.$$

We have already shown that the acceleration of a particle in uniform circular motion is

$$a = \omega^2 r.$$

Substituting the value of ω, from Equation (6-15), yields

$$a = \frac{4\pi^2 r}{T^2} \tag{6-16}$$

for the relationship between acceleration and period for uniform circular motion.

6-14 Planetary Motion

One important type of periodic motion which has been studied and recorded for centuries is that of the bodies constituting the solar system. Theories concerning these bodies have changed with the centuries, and to a certain extent these changes mirror man's intellectual progress. Of course, it has not always been known that these bodies were part of a system which was tied to the sun, but knowledge of the planets predates recorded history, for after the sun and the moon these bodies are often the most prominent objects in the sky that are visible to the naked eye, and they have the very special character that they wander in the heavens among the fixed stars.

Among the early theories which held sway for centuries was that associated with the name of Claudius Ptolemy (c. A.D. 150) and known as the *geocentric* theory of the universe. In the geocentric theory the earth was assumed to be at the center of the universe, and the sun, the moon, the planets, and the stars were thought to move around it in complicated paths. In a sense, this was a theory which most nearly corresponded to direct observation and to man's rather egocentric view of his own place in the universe.

Several centuries before Ptolemy, Aristarchus of Samos (c. 310–230 B.C.) proposed a theory in which the sun was fixed at the center of the universe and the earth revolved around the sun in a circular orbit, called the *heliocentric* theory of the universe. He also recognized that the stars appeared fixed in position because their distances from the earth were tremendous in comparison with the distance of the earth from the sun. Very few of the early astronomers accepted the heliocentric conception of the universe; from the second to the sixteenth century, only the geocentric theory of Ptolemy was taught and used. In the sixteenth century Nikolaus Copernicus (1473–1543) revived and extended the heliocentric theory of Aristarchus and thus started a revolution in scientific thought which was carried forward by Kepler, Galileo, and Newton. In the heliocentric theory

of Copernicus (see Figure 6-10), the sun was considered at the center of the universe, the planets revolved around the sun in circular orbits, and the fixed stars were assumed to lie in a sphere surrounding the solar system.

Copernicus' theory was not readily accepted by scientists of that period. Tycho Brahe (1546–1601), a famous Danish astronomer, made very careful and accurate measurements of the motions of the planets and the

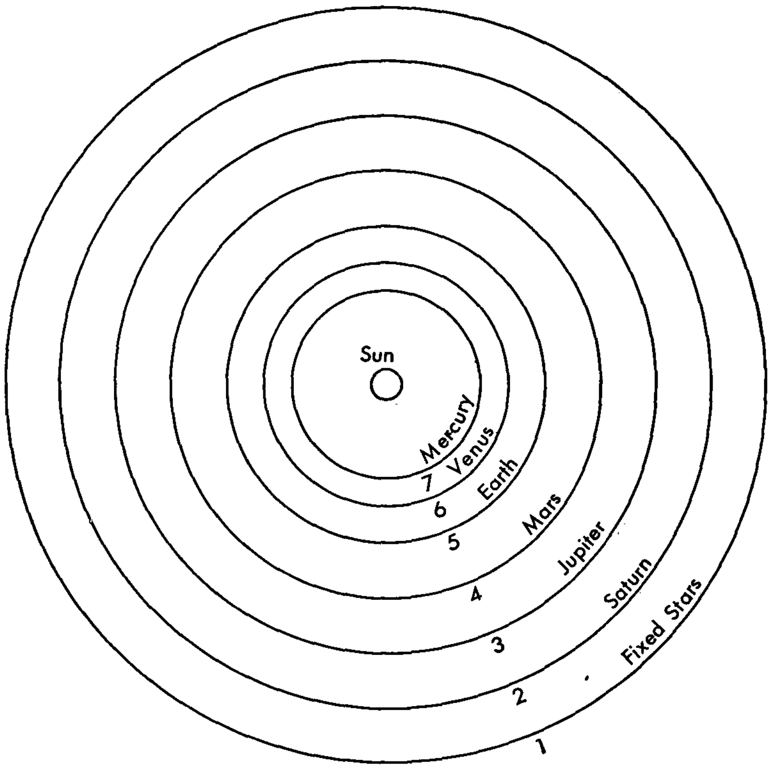


Fig. 6-10 Orbits of the planets and the fixed stars in the heliocentric theory of the universe according to Copernicus.

sun. He had never become convinced of the correctness of the Copernican hypothesis, but his extensive and careful measurements which he bequeathed to another astronomer, Johannes Kepler (1571–1630), laid the foundations of modern astronomy. It may be noted here that Brahe's observations were made without telescopic instruments. The telescope had not yet been invented.

From his study of the data accumulated by Tycho Brahe, Kepler deduced three laws which accurately described the motions of the planets about the sun. Kepler's three laws follow.

First law: *Each planet moves around the sun in an elliptic path (or orbit) with the sun at one focus of the ellipse.*

Second law: *As the planet moves in its orbit, a line drawn from the sun to the planet sweeps out equal areas in equal intervals of time* (see Figure 6-11).

Third law: *The squares of the periods of the planets are proportional to the cubes of their mean distances from the sun.*

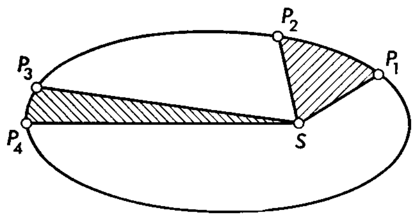


Fig. 6-11 The path of a planet about the sun S is an ellipse. $P_1P_2P_3P_4$ represent positions of a planet in its orbit at different times. The speed of a planet is such that an imaginary line joining the sun and the planet would sweep out equal areas in equal intervals of time. For example, area SP_1P_2 is equal to area SP_3P_4.

It can be seen that the simplified picture of the planetary system proposed by Copernicus is not sufficiently accurate; however, the elliptical orbits of the planets are not far removed from circles.

In large measure it was to explain Kepler's laws that Newton invented the laws of motion and the law of universal gravitation that we shall discuss in the next section. From astronomy, from pure speculation about the motion of the planets, has come the stimulus for much of today's engineering and for the foundations of modern science. Centuries of careful observation and profound thought have been distilled into a few carefully worded statements about the behavior of material bodies, and today every schoolboy knows more about the universe than did Kepler or even Newton.

6-15 Newton's Law of Universal Gravitation

Although Kepler's laws give an adequate description of the motions of the planets, they do not give a physical explanation of the cause of the motion. Newton, having introduced the concept of force into mechanics, now applied this concept to help explain the cause of the motions of the planets around the sun. He developed the *law of universal gravitation* which states that *any two bodies in the universe attract each other with a force which is directly proportional to the product of the masses of the two bodies and inversely proportional to the square of the distance between them.*

Stated in mathematical form, this law becomes

$$F \propto \frac{Mm}{r^2},$$

where M is the mass of one body, m is the mass of the other body, r is the distance between them, and F is the force that one body exerts on the

other. This proportionality may be replaced by an equation by replacing the proportional sign by an equal sign and a constant of proportionality, thus

$$\boxed{F = G_0 \frac{Mm}{r^2},} \tag{6-17}$$

where G_0 is the constant of proportionality known as the *universal constant of gravitation, or the gravitational constant.* The force of gravitation is directed along a line joining the two bodies.

Newton's law of universal gravitation, when combined with his laws of motion, predicts with great accuracy the motions of the planets. In fact, astronomy is one of the more important observational means of verification of Newtonian mechanics. In the last 40 or 50 years, it has been found that Newton's laws of motion have had to be extended to include some additional concepts in order to deal properly with atoms and molecules by a theory called *quantum mechanics*, and have had to be extended for even larger aggregates of matter when these are moving with extremely high speeds. It is extremely impressive, however, that Newton's laws have been shown to be a correct formulation of the rules of order which nature imposes upon material objects in the range of size from microscopic to astronomical, a range of from 10^{-6} m to 10^9 m, or more. The enormous success of the Newtonian view of the rational character of the world has had repercussions in other areas of human experience, and the philosophy of rationalism owes much of its stimulus to the great success of physics and astronomy in developing a rational view of nature.

Newton's law of universal gravitation is applicable to all particles in the universe. To use it most fully, it is necessary to evaluate the constant G_0. Of course the numerical value of G_0 will depend upon the units used for mass, force, and distance. Since the units appropriate to these quantities have already been defined through the equation $F = ma$, we cannot set G_0 equal to 1 but must determine its value experimentally. We can give G_0 a simple physical interpretation, no matter what system of units is used, by imagining two unit masses placed a unit distance apart; that is, $M = 1$, $m = 1$, and $r = 1$. The force F_1 with which two such masses attract each other is, from Equation (6-17),

$$F_1 = G_0;$$

that is, G_0 can be interpreted as the force with which two unit masses will attract each other when placed a unit distance apart.

The first experimenter to evaluate G_0 was Henry Cavendish (1731–1810). One method of determining the gravitational constant G_0 makes use

of a very delicate torsion balance such as that sketched in Figure (6-12). This consists of a fine elastic fiber AB suspended from some support at A; a small stiff metal rod CD is fastened to B. Two identical metal spheres, usually silver or gold, each of mass m, are mounted on the ends CD. Two much more massive spheres made of lead are placed near the small spheres, one in front of the sphere at D, the other behind the sphere at C. Each of the lead spheres has a mass M. The force of attraction between each lead sphere and the small metal sphere near it produces a torque about AB as an axis; the two torques are in the same direction and cause the fiber to twist through a small angle. The angle of twist can be measured by reflecting a beam of light from a small mirror attached to the fiber onto a scale. By shifting the positions of the large lead spheres so that one is now behind D and the other in front of C, the fiber is made to twist in the opposite direction. From a calibration of the fiber, the force F that each lead sphere of mass M exerts on the small metal sphere of mass m can be computed. The distance r between the centers of the spheres is also measured. Putting these data into Equation (6-17) will give the value of F.

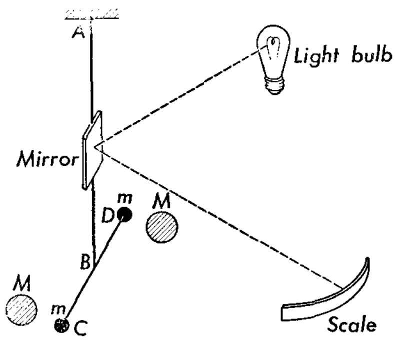


Fig. 6-12 Method of determining the gravitational constant G_0 using a delicate torsion balance.

The value of G_0 determined experimentally is

$$G_0 = 6.670 \times 10^{-8} \frac{\text{dyne cm}^2}{\text{gm}^2} = 6.670 \times 10^{-11} \frac{\text{nt m}^2}{\text{kg}^2}.$$

Because the force of gravitational attraction between ordinary objects is extremely small, we are not normally conscious of the gravitational attraction between adjacent objects, and we neglect it entirely in engineering calculations in comparison to the force of the earth's attraction and to the forces exerted by structural members upon each other.

6-16 The Gravitational Field

The force that exists between two particles because of their masses acts no matter how far apart these masses may be. This is one example of a type of force which is called an *action-at-a-distance* force, for it requires no material medium to transmit the force. We know, for example, that the space between the planets of the solar system is essentially empty. There is another way of thinking about gravitational forces and that is to imagine

that in the space all around a particle, and extending to infinity in all directions, there exists a *gravitational field.* Whenever any other particle finds itself in this gravitational field, it will experience a force F, given by Equation (6-17). We can define a new term called the *intensity of the gravitational field at any point in space as the force* F *which acts on a particle at this point divided by the mass* m *of the particle situated there.* Let us denote the gravitational field intensity by the letter **I**; then

$$\mathbf{I} = \frac{\mathbf{F}}{m}. \tag{6-18}$$

Note that the gravitational field intensity **I** is a vector quantity. Substituting the value of **F** from Equation (6-17) to obtain the magnitude of I yields

$$I = G_0 \frac{M}{r^2}. \tag{6-19}$$

Equation (6-19) shows that the gravitational field intensity varies inversely with the square of the distance from the particle of mass M which is the source of the field. The intensity is a vector quantity; its direction is that of the force **F** which acts on a particle placed anywhere in the field, and since the force is always one of attraction, its direction is always toward the mass M.

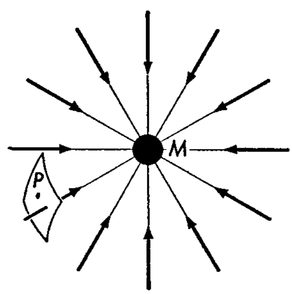


Fig. 6-13 Radial gravitational field around a small concentrated mass M. The number of lines of force through a unit area at P is proportional to the intensity I of the gravitational field at P.

We can develop a graphical method for representing the gravitational field so that it will show at a glance both the magnitude and the direction of the field intensity. This is illustrated in Figure 6-13 in which radial lines are drawn converging upon the mass M; a scale can be chosen so that the number of lines passing perpendicularly through a unit area at any point such as P will be proportional to the intensity I at that point. The direction of the field is given by the direction of the lines. Such lines are called *lines of force.*

An interesting case is that of the earth's gravitational field. Newton was the first to prove that the field outside a spherical mass is identical with

that of a mass concentrated at the center of the sphere. Hence, at points outside the earth's surface, the gravitational field intensity is given by Equation (6-19). However, we have been using the term "weight" to describe the force which the earth exerts on a mass m placed anywhere in its field. The intensity of the earth's gravitational field at the surface of the earth I_e is therefore

$$I_e = \frac{W}{m} = g. \tag{6-20}$$

The term which we have been calling the acceleration of a freely falling body is identical with the gravitational field intensity. From Equation (6-20) we note that the units for gravitational field intensity can be either those of acceleration or the ratio of force to mass, such as lb/slug, dynes/gm, nt/m.

Comparing Equations (6-19) and (6-20), we find that

$$g = G_0 \frac{M}{r^2} \quad \text{for } r \geqq R_e, \tag{6-21}$$

in which M is now the mass of the earth, and r is the distance of a point from the center of the earth. Equation (6-21) holds only for points outside the surface of the earth, that is, for distances r greater than the radius of the earth R_e. We have treated g as though it were independent of altitude in the solution of projectile problems. We see that if we set $r = R_e + h$, where h is the height above the surface of the earth, we may take h as small compared to R_e and to a good approximation we may set $r = R_e$.

Illustrative Example. Equation (6-21) may be used to determine the mass of the earth once G_0 has been measured. For this reason the experiment measuring the value of G_0 is popularly called "weighing the earth." Taking known values of $g = 9.80$ m/sec², $G_0 = 6.67 \times 10^{-11}$ nt m²/kg², $r = R_e = 6{,}380$ km $= 6.380 \times 10^6$ m, and solving for M, we get

$$M = \frac{gR^2}{G_0}$$

$$= \frac{9.8 \dfrac{\text{m}}{\text{sec}^2} \times (6.38 \times 10^6 \text{ m})^2}{6.67 \times 10^{-11} \dfrac{\text{nt m}^2}{\text{kg}^2}},$$

from which $$M = 5.98 \times 10^{24} \text{ kg}.$$

Illustrative Example. With the development of modern high-speed rockets, it is interesting to inquire what speed a particle should have in order to become a satellite of the earth, that is, to travel in an approximately circular path around the earth.

Any particle traveling with a speed v on a circular path of radius r has an acceleration toward the center of the circle of

$$a = \frac{v^2}{r}.$$

If it is traveling in the earth's gravitational field, then approximately $a = g$, so that

$$g = \frac{v^2}{r},$$

from which
$$v = \sqrt{gr}.$$

Near the surface of the earth $r = 4{,}000$ mi, and

$$g = 32\,\frac{\text{ft}}{\text{sec}^2} = \frac{32}{5{,}280}\,\frac{\text{mi}}{\text{sec}^2}.$$

Substituting these values into the above equation yields

$$v = \sqrt{\frac{32\text{ mi}}{5{,}280\text{ sec}^2} \times 4{,}000\text{ mi}} = 5\,\frac{\text{mi}}{\text{sec}}.$$

6-17 The Curvature of Space

From Newton's first law, we might infer that one of the consequences of this law is that an observer cannot measure his own velocity except with respect to an outside reference frame. Velocity is a relative quantity rather than an absolute quantity. This concept was one of the building blocks of the theory of special relativity by Albert Einstein (1879–1955) and is now part of the basic structure of physics. Another concept due to Einstein is the impossibility of distinguishing, when inside a closed system, between a gravitational field and accelerated motion. Let us consider an observer moving through space in a rocket ship which is completely enclosed. He has on board a massive object which is hung from a spring balance. A reading of the balance will tell him either that he is being accelerated or that he has entered a gravitational field, but no measurement he can make within the rocket ship can distinguish between the two possibilities.

With this as a starting point, we see that we may consider any gravitational field as though it were an accelerated enclosure. Even a ray of light traversing an intense gravitational field must behave as though it were traversing an accelerated chamber. A ray of light passing horizontally through an elevator accelerated in the upward direction would appear to be deflected toward the floor; that is, it would enter the elevator through an aperture in the side wall at some distance s above the floor and would

leave the elevator through an aperture in the opposite wall at some lesser distance s' above the floor. To an observer in the elevator, the light beam would appear to have been deflected toward the floor. A massive object located in an elevator which is being accelerated in the upward direction behaves as though there were a gravitational field directed toward the floor. By analogy, we would expect a ray of light passing through an intense gravitational field to be deflected toward the source of the field. Measurements made, during a solar eclipse, of starlight passing by the edge of the sun indicate that the light is bent by the sun's gravitational field, in quantitative agreement with these ideas.

In this connection it is of interest to reconsider just what we mean by a straight line in space. In practice, when we wish to determine the straightness of a straightedge or when we wish to determine the straight line connecting two points, we sight along the straightedge or we look through a surveyor's transit. In practice, a straight line is defined as the path of a ray of light. Since the path of a ray of light is curved in the vicinity of a gravitational field, we must infer that space itself is curved in the vicinity of a gravitational field.

Problems

6-1. A flywheel of a steam engine is rotating with a uniform angular speed of 180 rpm. (a) Express this angular speed in radians per second. (b) Determine the linear speed of a portion of this wheel which is at a distance of 2 ft from the center. (c) Through what angle, in degrees, will the wheel have rotated in 10 sec?

6-2. The fuel supply is shut off from an engine when its angular speed is 1,800 rpm. It stops rotating 15 sec later. (a) Determine its angular acceleration, assuming it to be constant. (b) Through what angular displacement, in radians, will the wheel have rotated before it comes to rest?

6-3. The angular speed of an automobile engine is increased from 3,000 rpm to 3,600 rpm in 20 sec. (a) Determine its acceleration, assuming it to be uniform. (b) Determine the number of revolutions made by the engine in this time.

6-4. A uniformly accelerated wheel initially is rotating with an angular velocity of 10 radians/sec, and after 150 radians is rotating with an angular velocity of 25 radians/sec. (a) Determine the angular acceleration of the wheel. (b) For how long a time was the acceleration applied?

6-5. Derive Equation (6-7b) from the definitions of angular velocity and angular acceleration.

6-6. Derive Equation (6-7c) from the definitions of angular velocity and angular acceleration.

6-7. Derive Equations (6-7d) and (6-7e).

6-8. A drum 2 ft in diameter is free to turn on a fixed horizontal axis, as shown in Figure 6-14. A rope is wrapped around the drum. The free end of the rope is tied to a hanging weight which is observed to fall with an acceleration of

16 ft/sec^2 as the rope unwinds from the drum. If the weight starts from rest, (a) determine the angular velocity of the drum at the end of 5 sec. (b) If the weight has a mass of 2 slugs, determine the tension in the rope. (c) Determine the angular displacement of the drum at the end of 10 sec.

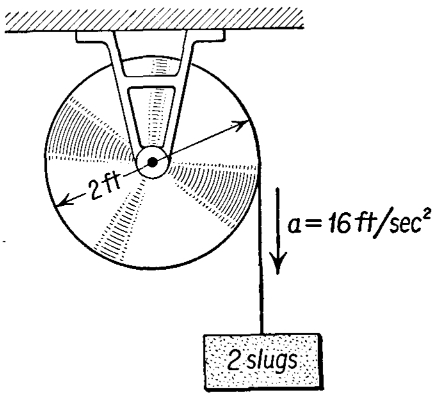


Fig. 6-14

6-9. A stone whose mass is 150 gm is attached to a cord 30 cm long and placed on a smooth horizontal table. The stone is then whirled in a circular path with a speed of 25 cm/sec. Determine (a) the tension in the cord and (b) the acceleration of the stone.

6-10. A car weighing 3,000 lb rounds a curve of 600 ft radius at a speed of 40 mi/hr. What lateral force must the ground exert on the tires to keep this car moving in this circular path? In what direction is this force?

6-11. Assume that the moon is moving in a circular path of 380,000 km radius about the earth. The period of revolution is 27.3 days. (a) Calculate the acceleration of the moon. (b) Calculate the mass of the earth.

6-12. An automobile rounds a curve of 800 ft radius at a speed of 50 mi/hr. What is the minimum value of the coefficient of friction between the tires and the road in order that there be no skidding?

6-13. (a) Determine the angle at which a road should be banked if the radius of the curve is 1,600 ft and if it is to supply the necessary centripetal force to a car traveling at 50 mi/hr. (b) If the same car is driven at 75 mi/hr over the banked road, what must be the coefficient of friction between the tires and the road in order that there be no skidding?

6-14. Uniform circular motion may be described in polar coordinates in terms of the equations $r = R$, a constant, and $\theta = \omega t$. The rectangular coordinates of a point in uniform circular motion are given by the equations

$$x = R \cos \omega t \qquad \text{and} \qquad y = R \sin \omega t.$$

From these relationships, show by the use of the calculus that $a = \omega^2 R$.

6-15. A stone of mass 100 gm is tied to the end of a string 50 cm long. The stone is twirled as a conical pendulum, so that it rotates in a horizontal circle 30 cm in radius, as shown in Figure 6-15. Determine (a) the angular speed of rotation and (b) the tension in the string.

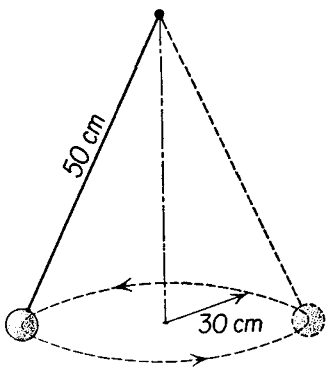


Fig. 6-15 Conical pendulum.

6-16. A small car whose mass is 25 gm moves on the inside of a vertical circular track of radius 40 cm. (a) Determine the minimum speed that the car must have at the top of the circular track in order to move in this circular path. (b) Assuming that it has the minimum safe speed at the top of the track, determine its speed at the bottom of the track. What is the force that the track exerts on the car (c) at the top of the track and (d) at the bottom of the track?

6-17. A pilot pulls his plane out of a dive by moving in the arc of a vertical circle with a speed of 600 mi/hr. (a) Determine the minimum radius of this circle if his acceleration is not to exceed $7g$ at the lowest point. (b) Determine the force which acts on a pilot weighing 180 lb.

6-18. The period of Jupiter is 11.86 years. With the aid of Kepler's third law, determine its distance from the sun. The distance from the earth to the sun is 1.49×10^8 km.

6-19. Assuming the earth to move around the sun in a circular orbit of radius 1.49×10^8 km with a period of 365.3 days, compute (a) the speed of the earth in its orbit and (b) the acceleration of the earth relative to the sun. The mass of the earth is 5.98×10^{24} kg. (c) Determine the gravitational force between the sun and the earth.

6-20. Two blocks, each of mass 1 kg, rest on a horizontal table a distance of 1 m apart. (a) What is the force with which they attract each other? (b) What is the minimum value of the coefficient of static friction if the blocks are to remain at rest on the table?

6-21. In Equation (6-21) set $r = R + h$, where R is the radius of the earth and h is the height above the surface of the earth. Expand $r^{-2} = (R + h)^{-2}$ by the binomial expansion. What is the greatest value h may have if g is to remain a constant (a) within 1 per cent? (b) Within 0.1 per cent?

6-22. The equation $x^2 + y^2 = r^2$ defines a circle centered at the origin. If this equation is differentiated twice with respect to time, an equation relating the x and y components of velocity and acceleration results. For the special case of $x = r$, $y = 0$ show that

$$a_x = -\frac{v^2}{r}.$$

6-23. The distance from the earth to the sun is 149×10^6 km. Assuming that the earth's orbit around the sun is circular, calculate the mass of the sun.

6-24. In an experiment on the determination of the universal constant of gravitation G_0, a lead sphere whose mass is 2,000 gm was placed near a gold sphere of mass 4 gm with their centers 6 cm apart. The force of attraction between them was found to be 1.45×10^{-5} dyne. Determine the value of G_0 from these data.

6-25. Assuming that a planet moves in a circular orbit of radius r around the sun of mass M with a period T, show that

$$T^2 = \frac{4\pi^2}{G_0 M} r^3.$$

Compare this with Kepler's third law.

6-26. A satellite, launched by Russia and called "Sputnik," traverses an orbit around the earth at an altitude of about 500 miles. Determine (a) the value of g at this altitude; (b) the linear velocity of the satellite; and (c) its period.

6-27. Two equal masses of 2 slugs each are joined by a rod of negligible mass which is pivoted at its center. The system rotates in a horizontal plane at an angular velocity of 20 rad/sec. Each mass is 1 ft from the axis of rotation. What is the tension in the rod?

7

Work and Energy

7-1 Work Done by Forces

An extremely important concept that has been developed in physics is that of the *work done* on a body by the action of some external agent which exerts a force on this body and produces motion. For example, whenever someone lifts a body, he does work by exerting a force upward on it and moving it upward. Whenever a steam locomotive pulls a train, a series of processes takes place in the steam engine of the locomotive which enables it to exert a force on the train and move it in the direction of the force. The term *work*, as used in physics, is a technical term. Whenever work is done by an external agent on a body, *the work done is the product of the force which acts on the body and the distance through which the body moves while the force is acting on it, provided that the force and the distance through which the body moves are parallel to each other.*

In the initial development of the concept of work, we shall restrict our discussion to work done by a constant force. We shall later (Section 7-8) remove this restriction and treat the more general case of work done by a variable force.

If a constant force F acts on a body for a distance s in the direction of the force, then the work done $\mathscr{W}$ is, from the definition,

$$\boxed{\mathscr{W} = Fs.} \qquad \textbf{(7-1a)}$$

If the force F and the distance s are not parallel, then only that component of the force which is in the direction of the motion does the work. For example, if a heavy block is to be moved, it may be more convenient to pull with a force F at some angle θ with respect to the ground, as shown in Figure 7-1. The component of F in the direction of the motion is $F \cos \theta$; if the block is moved through a distance s while this force is acting on it,

the work done $\mathscr{W}$ is

$$\boxed{\mathscr{W} = Fs \cos \theta.} \tag{7-1b}$$

Work is a scalar quantity. The concept of physical work is often a confusing one, in part because of the way the word "work" is used in everyday language, and in part because there is no direct physiological analogue

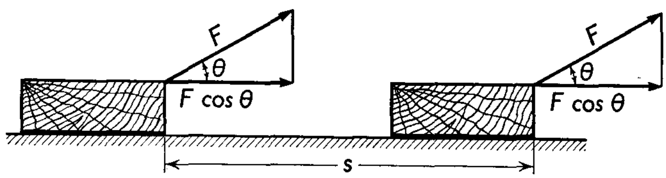


Fig. 7-1 Work done by a constant force F acting at an angle θ to the direction of its motion is $Fs \cos \theta$.

to the physical concept of work. In physics no force exerted, no matter how great nor for how long a time, generates any work unless there is a displacement. No force generates work unless the force has a component in the direction of the displacement. The centripetal force which a string

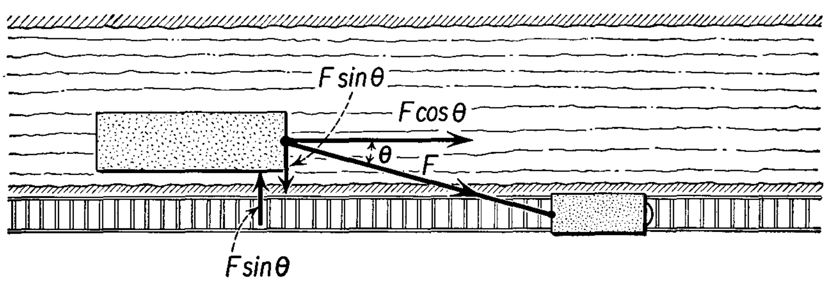


Fig. 7-2

exerts on a stone in uniform circular motion does no work upon the stone. *The sensation of tiredness has no direct relationship to physical work.* Suppose, for example, that you are called upon to support one end of a car while the driver changes a tire. When the operation is concluded you will have done no work, for, although you were called upon to exert a large force, there was no displacement of the car. Consider an even more unlikely situation. Suppose a barge is being towed through a canal by an engine located alongside the barge canal, as shown in Figure 7-2. If the force exerted by the engine is F, the component of the force parallel to the

canal is $F \cos \theta$, while the component perpendicular to the canal is $F \sin \theta$. The perpendicular component tends to urge the barge against the side of the canal. You are called upon to exert a force against the side of the barge of magnitude $F \sin \theta$ to keep the barge from scraping along the walls of the canal, and you accompany the barge on its trip through the canal, continually exerting the required force. When the trip is completed you will have done no work on the barge, for the force exerted had no component in the direction of the displacement.

7-2 Units for Expressing Work

There are several different units that are used for expressing the work done. In every case the unit used must be equivalent to the product of a force by a distance. In the British gravitational system the unit used is the foot pound (ft lb), the product of the unit of distance by the appropriate unit of force. In the cgs system the analogous unit would be the dyne centimeter, but this unit has been given a special name, the *erg; one erg is the work done by a force of one dyne acting through a distance of one centimeter*, or

$$1 \text{ erg} = 1 \text{ dyne} \times 1 \text{ cm} = 1 \text{ dyne cm}.$$

In the mks system the unit of work is called the *joule. A joule is defined as the work done by a force of one newton acting through a distance of one meter;* that is,

$$1 \text{ joule} = 1 \text{ nt} \times 1 \text{ m} = 1 \text{ nt m}.$$

The relationship between the joule and the erg can be found readily from the facts that 1 nt = 100,000 dynes and 1 m = 100 cm. By the usual conversion procedure, we write

$$1 \text{ joule} = 1 \text{ nt m} \times \frac{100{,}000 \text{ dynes}}{1 \text{ nt}} \times \frac{100 \text{ cm}}{1 \text{ m}},$$

$$1 \text{ joule} = 10{,}000{,}000 \text{ dyne cm} = 10{,}000{,}000 \text{ ergs},$$

$$1 \text{ joule} = 10^7 \text{ ergs},$$

or

$$10^{-7} \text{ joule} = 1 \text{ erg}.$$

While both work and torque are compounded of the product of a force by a displacement and therefore have the same units, these are quite different things. Torque is a vector quantity, while work is a scalar quantity. Torque is the product of a force by a distance which is always measured in a direction perpendicular to the force, while work is the product of a force by a distance parallel to the direction of the force. While there

need be no displacement to generate a torque, on the other hand, no work can be done without a displacement of the force.

Illustrative Example. Referring to Figure 7-1, suppose that the body is pulled along a level floor by a rope making an angle of 30° with the floor. If the body is moved a distance of 15 m, and if the force F is 40 nt, the work done is

$$\mathscr{W} = Fs \cos\theta = 40 \text{ nt} \times 15 \text{ m} \times \cos 30°$$

$$= 40 \times 15 \times 0.866 = 520 \text{ nt m},$$

or

$$\mathscr{W} = 520 \text{ joules}.$$

7-3 The Scalar Product

We have seen in the preceding paragraphs that work is a scalar quantity, yet it is composed of the product of two vector quantities—the force and the displacement. Such operations occur quite often in physics and have

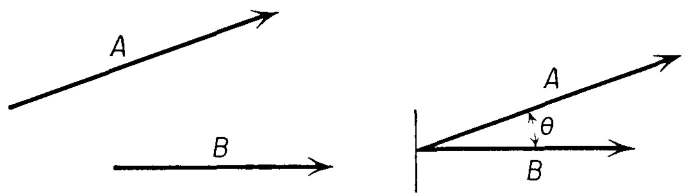


Fig. 7-3 The scalar or dot product of two vectors **A** and **B** is given by $\mathbf{A}\cdot\mathbf{B} = AB\cos\theta$, where θ is the angle between the two vectors when they are drawn from a common origin.

been given a special name, the *scalar product* of two vectors. If we have two vector quantities, such as the vectors **A** and **B** in Figure 7-3, *we define the scalar product of these two vectors as the product of the magnitude of* **A** *by the magnitude of* **B** *by the cosine of the angle between them.* The scalar product of the two vectors **A** and **B** is sometimes called the *dot product* because it is represented by writing **A·B**. Thus we have

$$\mathbf{A}\cdot\mathbf{B} = AB\cos\theta, \tag{7-2}$$

where θ is the angle between the vectors **A** and **B** when they are drawn from a common origin. The result of multiplying two vectors by the operation called the scalar product is a scalar quantity; that is, it has magnitude only; there is no direction to be associated with the scalar product.

We see that the definition of the scalar product is perfectly adapted for the representation of work, for work has been defined as a scalar quantity, the result of the multiplication of a vector, force, by another vector, displacement. If a force **F** applied to a body produces a displacement s,

the work done is the scalar product of the force and the displacement for

$$\mathscr{W} = \mathbf{F}\cdot\mathbf{s} = Fs\cos\theta. \tag{7-3a}$$

In an incremental displacement $\Delta\mathbf{s}$, the incremental work is

$$\Delta\mathscr{W} = \mathbf{F}\cdot\Delta\mathbf{s} = F\,\Delta s\cos\theta, \tag{7-3b}$$

and in the limit of small displacement,

$$d\mathscr{W} = \mathbf{F}\cdot d\mathbf{s} = F\,ds\cos\theta. \tag{7-3c}$$

In these formulas boldface type has been used to represent vector quantities, while italics have been used to represent the magnitudes of these quantities, or to represent scalar quantities. The great advantage of the representation of work as the dot product is that there is never any ambiguity as to which angle is referred to as θ in the formula $\mathscr{W} = Fs\cos\theta$. From the definition of the dot product, the angle θ is always the angle between the vectors $\mathbf{F}$ and $\mathbf{s}$ when these two vectors are drawn from a common origin. There is the further advantage in the representation $\mathscr{W} = \mathbf{F}\cdot\mathbf{s}$ in that there will be no confusion about the fact that work is a scalar quantity, for the result of the operation called the dot product is always a scalar quantity.

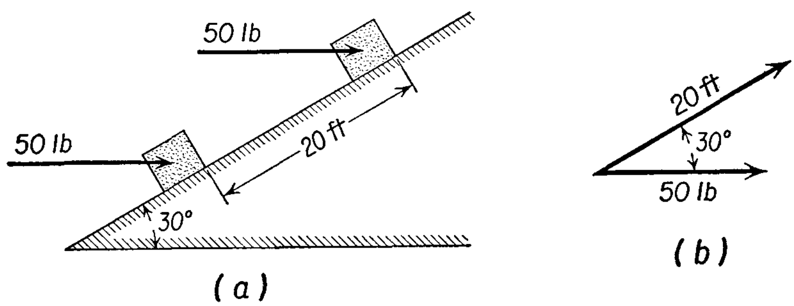


Fig. 7-4

Illustrative Example. A body is pushed up a 30° inclined plane a distance of 20 ft by a horizontal force of 50 lb, as shown in Figure 7-4(a). Find the work done on the body. In Figure 7-4(b) the force vector and the displacement vector are shown drawn from a common origin. The angle θ appropriate for use in Equation (7-1b) or (7-3a) is 30°. Substituting into the equation, we find

$$\begin{aligned}\mathscr{W} &= \mathbf{F}\cdot\mathbf{s} = Fs\cos\theta \\ &= 50\text{ lb} \times 20\text{ ft} \times \cos 30° = 50 \times 20 \times 0.866\text{ ft lb} \\ &= 866\text{ ft lb}.\end{aligned}$$

When the displacement $\mathbf{s}$ is in the same direction as the applied

constant force **F**, the angle θ is zero. The work $\mathscr{W}$ done is

$$\mathscr{W} = \mathbf{F}\cdot\mathbf{s} = Fs\cos 0^\circ = Fs.$$

Since F and s represent the magnitudes of the two vectors, they are both positive quantities. Thus $\mathscr{W}$ is positive and represents the work done *by* the agency applying the force on the body which has been displaced.

In many cases the displacement is opposite in direction to the applied force, as, for example, when a moving body is slowed down by the action of an external force. When the displacement vector **s** is opposite to the applied force **F**, the angle θ between the two vectors is 180°. Since the cos 180° $= -1$, we find that the work done *by* the applied force is negative, or

$$\mathscr{W} = -Fs.$$

We interpret this result by saying that negative work has been done by the agency applying the force to the body which has been displaced. Alternatively, we may say that positive work has been done by the body which has been displaced upon the agency exerting the force. Thus, when a baseball is caught by a fielder, the ball does work upon the fielder.

The discussion of Figure 7-2 illustrates a case in which the applied force is perpendicular to the displacement. Here the angle θ is 90°, and the work done is zero.

7-4 The Vector Product

The extension of the concept of multiplication to vector quantities requires some additional consideration. By analogy with arithmetic, it is clear that the resulting product should involve the product of the magnitudes of the two vectors, but the question remains as to what to do about the directions; and should the resulting quantity be a vector or a scalar? In the preceding section we have seen the virtue of one type of product, the scalar product, in which the product of two vector quantities is a scalar quantity which has been so defined as to be ideally suited to represent work.

A second product, called the *vector product*, has been defined so as to be ideally suited to represent torque. Given two vectors **A** and **B**, as in Figure 7-5, we define their vector product as a vector **C** which is perpendicular to the plane formed by **A** and **B** whose magnitude is given by

$$C = AB\sin\theta, \tag{7-4}$$

and whose direction is given by the right-hand rule, or by the direction of advance of a right-handed screw which is made to rotate from the direction of **A** into the direction of **B**. The vector product is usually represented by the symbol **×** and is therefore called the *cross product*, to distinguish it from the scalar product or dot product.

Suppose a force **F** acts on a body at a point P, as illustrated in Figure 7-6. The vector directed from the axis of rotation to the point P is **r**. Then

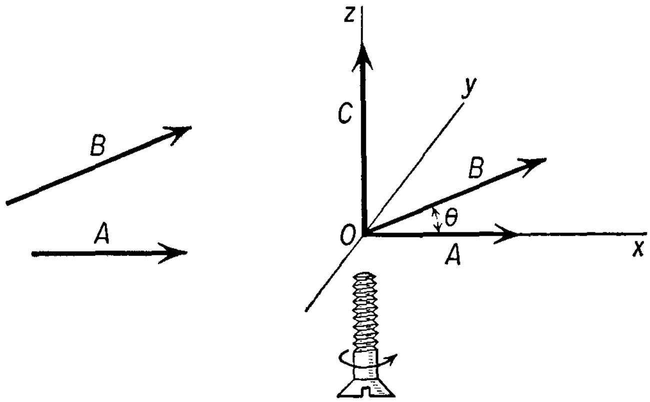


Fig. 7-5 The vector product, or cross product, of two vectors **A** and **B** is the vector **C**, written as $\mathbf{A} \times \mathbf{B} = \mathbf{C}$. The magnitude of **C** is given by $C = AB \sin \theta$, and the direction of **C** is perpendicular to the plane formed by **A** and **B** pointing in the direction of advance of a right-hand screw turned so as to advance from **A** to **B**. As in the scalar product, the angle θ is the angle between **A** and **B** when these are drawn from a common origin.

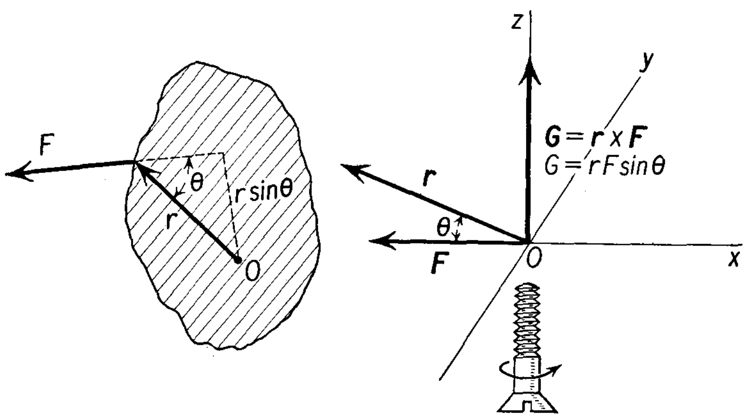


Fig. 7-6 The torque of the force **F** about an axis through O is given by $\mathbf{G} = r \times \mathbf{F}$.

the torque **G** developed by **F** about the axis of rotation is given by

$$\mathbf{G} = \mathbf{r} \times \mathbf{F}, \tag{7-5}$$

and the magnitude of **G** is G, given by

$$G = rF \sin \theta, \tag{7-6}$$

which is precisely the product of the force by the perpendicular distance between the line of action of the force and the axis of rotation, the moment arm. The direction of the torque vector is given by the same right-hand rule in the cross product as the direction previously given for the determination of the torque.

Again the great virtue of the cross-product notation for the representation of torque lies in the fact that the one rule for the determination of the cross product is adequate for many laws of physics. The same cross product will reappear in the discussion of angular momentum, in the relationship between electricity and magnetism, and in other places in physics. The cross product serves to simplify the notation and the formulas which must be learned.

We must note that the order of the factors which appear in the cross product is of some importance. While the magnitude of $\mathbf{A}\times\mathbf{B}$ is the same as the magnitude of $\mathbf{B}\times\mathbf{A}$, the two vectors are opposite to each other in direction, for a right-handed screw which rotates from $\mathbf{A}$ to $\mathbf{B}$ advances in the opposite direction from one which advances from $\mathbf{B}$ into $\mathbf{A}$, and we write

$$\mathbf{A}\times\mathbf{B} = -\mathbf{B}\times\mathbf{A}. \tag{7-7}$$

7-5 Work and Energy

An important question which arises from this discussion concerns the result of the work done by the various forces which act on different bodies. In some cases the results are immediately obvious. For example, the work done by a force which accelerates a body produces a change in its speed; the work done in lifting a body produces an increase in the height of the body with respect to its former position; the work done against frictional forces produces an increase in the temperature of one or more of the bodies involved. In other cases the results may not be so obvious. Some bodies may become charged electrically; others may become magnetized. These changes will be discussed at the appropriate places in the text. One general conclusion can be drawn here; that is, that whenever work is done, some change is produced in the body or system of bodies on which the forces acted. To describe these changes, another technical term is used. We say that the work done produces *a change in the energy of the body or system of bodies.* In the first case above, the energy of motion or the *kinetic energy* of the body is changed; in the second case the positional energy or *potential energy* of the body is increased. In each case *the change in energy is defined as equal to the work done on the body or system of bodies.* From this, it follows that the units used in expressing the energy of a system are the same as the units of work. We shall see that it is possible for a body to gain energy as a

result of work done upon it, and, conversely, a body may lose energy by doing work upon a second object.

7-6 Kinetic Energy

Suppose that a constant force F acts on a body of mass m for a distance s in the direction of F, as shown in Figure 7-7. The work done on the body by the force F is

$$\mathcal{W} = Fs.$$

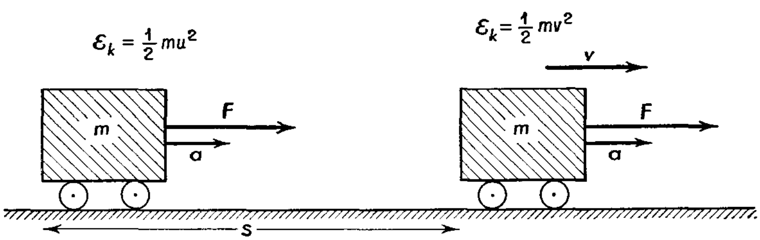


Fig. 7-7 The work done by a force in accelerating a body increases its kinetic energy.

Under the action of a constant force, the body will receive an acceleration a given by

$$a = \frac{F}{m},$$

and the velocity of the body will be increased from its initial value u to some final value v, given by Equation (2-27c) as

$$v^2 = u^2 + 2as,$$

and, substituting F/m for a, we find

$$v^2 = u^2 + 2\frac{Fs}{m}.$$

We now multiply the equation above by $\frac{m}{2}$, substitute the value $\mathcal{W}$ for the product Fs, and transpose to find

$$\boxed{\mathcal{W} = \tfrac{1}{2}mv^2 - \tfrac{1}{2}mu^2.} \qquad \textbf{(7-8a)}$$

We call the quantity $\frac{1}{2}mv^2$ the final *kinetic energy* of the body and $\frac{1}{2}mu^2$ its initial kinetic energy, and we say that the work done on the body by the force F acting through a distance s has produced a change in the kinetic energy of the body.

In general, the kinetic energy $\mathcal{E}_k$ of a body of mass m moving with speed v is given by

$$\boxed{\mathcal{E}_k = \tfrac{1}{2}mv^2.} \tag{7-8b}$$

Just as work is a scalar quantity, so is kinetic energy a scalar quantity.

7-7 Potential Energy in a Uniform Gravitational Field

When a mass m is placed in a uniform gravitational field in which the intensity of the gravitational field is g, the gravitational force is given by mg, the weight of the body. If such a body is lifted through a height h, the work done by the agency which lifts it is mgh (see Figure 7-8). We say that the work done has increased the *potential energy* $\mathcal{E}_p$ of the body, and we write as a defining equation

$$\boxed{\mathcal{E}_p = mgh.} \tag{7-9}$$

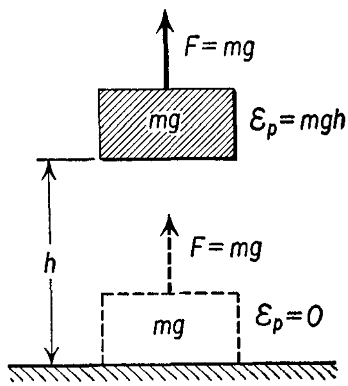


Fig. 7-8 The work done in lifting a body increases its potential energy.

The position at which the potential energy is zero is quite arbitrary and, in fact, makes no difference in the consideration of a particular problem. We shall always be interested in the change in potential energy associated with a change in position, and shall not attempt to ascribe a meaning to the value of the potential energy itself. For this reason it is often convenient to take the initial position of the body as the position of zero potential energy.

The potential energy in the earth's gravitational field is sometimes called the *gravitational potential energy*, to distinguish it from other forms of potential energy such as the energy of a stretched spring or the energy of an electric charge in an electric field, which will be discussed in the later chapters of this book.

7-8 Potential Energy of a Spring

Bodies which are acted on by external forces generally undergo changes in size or shape. A helical spring is an important example of such a body.

If a spring is fixed at one end and a force F is applied to the other end, an extension s is produced which is proportional to F and is given by the equation

$$F = ks, \tag{7-10}$$

where k is called the *constant of the spring;* if the force is measured in pounds and the extension is measured in feet, the constant k is given in pounds per foot. Each helical spring has a constant k associated with it, whose value depends upon the material from which it is made, the wire diameter, the number of turns per unit length, and the diameter of the spring itself. The linear relationship between applied force and displacement of a helical spring is the basis of the uniform scale of a spring balance.

To stretch a spring by an incremental amount Δs when it has already been stretched by an amount s requires that an amount of work $\Delta \mathscr{W}$ be done by the agency doing the stretching, which is given by

$$\Delta \mathscr{W} = F\,\Delta s = ks\,\Delta s.$$

The total work done in stretching the spring from the position of zero extension to the maximum extension s_m is given as the sum of the work done in the incremental extensions from the initial to the final positions. Letting the incremental displacement become infinitesimal, as the number of displacements increases appropriately, we write, in the language of the calculus,

$$\mathscr{W} = \int_0^{\mathscr{W}} d\mathscr{W} = \int_0^{s_m} ks\,ds = \tfrac{1}{2}ks_m^2.$$

The same result may be attained by observing that since the force on the spring varies at a constant rate, the average value of the force exerted by the spring, $\bar{F}$, is half the sum of the initial and final forces. In the form of an equation,

$$\bar{F} = \frac{ks_m}{2},$$

and the work done in stretching the spring is the product of the average force by the total displacement, or

$$\mathscr{W} = \bar{F}s_m = \tfrac{1}{2}ks_m^2\,.$$

Calling the elastic energy in a stretched spring $\mathscr{E}_e$, and dropping the subscript m, we get

$$\mathscr{E}_e = \tfrac{1}{2}ks^2. \tag{7-11}$$

The energy of a stretched spring is associated with position or deformation rather than with motion and is therefore a form of potential energy.

7-9 Conservation of Energy

In the absence of dissipative forces, such as friction, the total mechanical energy of a system is constant. This principle, known as the *principle of conservation of mechanical energy,* is of great usefulness in the solution of many problems in mechanics. Initially discovered in mechanics, the principle of energy conservation has since been extended to include heat energy, radiant energy, electrical energy, chemical energy, and the energy associated with transformations of mass, so that it has become one of the most fundamental and unifying principles of physics, and indeed, of all science. It was Albert Einstein who first showed, from his work in the theory of relativity, that mass and energy were equivalent and could be interchanged in accordance with the formula

$$\mathcal{E} \equiv m. \tag{7-12a}$$

If the equivalence between mass and energy is to be expressed in systems of units in which mass and energy units are defined separately, it is necessary to apply a conversion factor to convert units of mass to units of energy. In the cgs and mks systems we write

$$\mathcal{E} = mc^2, \tag{7-12b}$$

where c is the velocity of light, 3×10^{10} cm/sec or 3×10^8 m/sec. In the cgs system of units, $\mathcal{E}$ is expressed in ergs, m in grams, and c, the velocity of light, in centimeters per second. In the mks system $\mathcal{E}$ is expressed in joules, m in kilograms, and c in meters per second.

Einstein's mass-energy relationship has been widely verified experimentally. The nuclear reactor is an example of the practical application of this relationship, in which energy in the form of mass is converted to energy in the form of heat.

In more general and extended form, the principle of conservation of energy states that *energy may be transformed from one type to another without loss,* and that *in a closed system the total amount of energy remains constant.*

One of the chief contributions of physics in the service of society has been in identifying and defining the concept of energy. This remarkable idea, and its subsequent theoretical and experimental development, has resulted in mechanical devices for the utilization of fuels such as coal and petroleum, of the mechanical energy of rain, and of the conversion of mass to other forms of energy by nuclear processes. As a result, each generation in this last century has lived in a substantially different way from that of its parents.

7-10 Kinetic and Potential Energy Transformations

Suppose that a ball of mass m is thrown vertically upward with an initial velocity u. Since the acceleration of the ball is equal to g, directed verti-

cally downward, we may calculate from Equation (2-27c) that the ball will rise to a height h given by the equation

$$u^2 = 2gh.$$

Consider the application of the principle of the conservation of mechanical energy to the same problem. Since no mechanical energy is transformed into heat by frictional processes, the total energy that the ball has initially will also be the total energy at every point in its path. The kinetic energy of the ball at the instant it is thrown is $\frac{1}{2}mu^2$, and its potential energy is zero. At the highest point h in its path, its kinetic energy is zero and its potential energy is mgh. Equating the initial total energy to the final total energy, we find

$$\tfrac{1}{2}mu^2 = mgh,$$

from which

$$u^2 = 2gh.$$

The speed of the ball at any known height above the initial point can be found in a similar manner.

The motion of a simple pendulum provides another interesting example of the transformation of energy and the usefulness of the energy-conservation principle in the solution of problems in mechanics. A simple pendulum consists of a small ball of mass m attached to one end of a string of negligible weight and of length L. The other end of the string is attached to some fixed point O, as shown in Figure 7-9. When at rest, the string hangs vertically with the ball at its lowest position C. Let us call the potential energy of the pendulum zero when it is in this position. When the ball is drawn aside from its lowest position and released, it moves in the arc of a circle of radius L. Suppose the ball is moved to position A, at a height h above point C. The only work which is done on the ball is that done against the force of gravity in lifting the ball; it increases the potential energy by the amount mgh. As the ball swings from point A to point C some of its energy is transformed from potential to kinetic energy until, at C, all of it is kinetic energy. As it moves from C to B, some of its kinetic energy is now transformed into potential energy, and, at B, all of it is potential energy again. The pendulum may be viewed as a device for the continuous exchange of energy between these two forms.

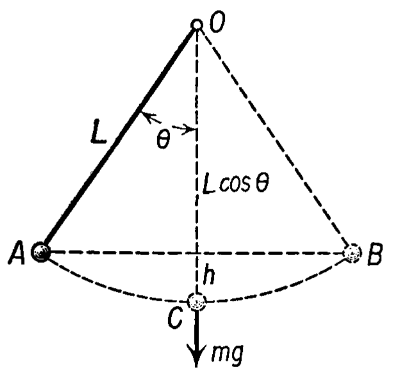


Fig. 7-9 Motion of a simple pendulum.

In the simple pendulum the ball moves with variable acceleration. The tension in the rope is not constant, and the methods of the preceding chapters are not adequate to analyze the motion of the pendulum. This analysis may be undertaken by the application of the energy principle.

Illustrative Example. A simple pendulum of length 10 cm with a pendulum bob of mass 5 gm is drawn aside so that the string makes an angle of 30° with the vertical. Find the linear speed of the pendulum bob when the pendulum has swung back so that its deflection is only 15°.

At any angle of deflection θ, the height h of the pendulum bob above its initial position C (Figure 7-9) is given by

$$h = L - L\cos\theta.$$

The potential energy of the pendulum, when $\theta = 30°$, is equal to $(\mathcal{E}_p)_i$ such that

$$(\mathcal{E}_p)_i = 5\text{ gm} \times g \times (10\text{ cm} - 10\text{ cm} \times \cos 30°).$$

The kinetic energy at this position is zero.
The potential energy of the pendulum, when $\theta = 15°$, is equal to $(\mathcal{E}_p)_f$ such that

$$(\mathcal{E}_p)_f = 5\text{ gm} \times g \times (10\text{ cm} - 10\text{ cm} \times \cos 15°).$$

The kinetic energy $\mathcal{E}_k$ at $\theta = 15°$, when the velocity of the bob is v, is

$$\mathcal{E}_k = \tfrac{1}{2}mv^2 = \tfrac{1}{2} \times 5\text{ gm} \times v^2.$$

Equating the total energy at $\theta = 30°$ to the total energy at $\theta = 15°$, we find

$$5 \times 10 \times 980 \times 0.134\text{ erg} = \tfrac{1}{2} \times 5\text{ gm} \times v^2 + 5 \times 10 \times 980 \times 0.034\text{ erg},$$

$$5 \times 10 \times 980 \times 0.10\text{ erg} = \tfrac{5}{2}\text{ gm} \times v^2,$$

so that

$$v = 44.3\,\frac{\text{cm}}{\text{sec}}\cdot$$

A mass which is oscillating at the end of a spring provides another illustration of an exchange between potential and kinetic energy. In this this case the energy is shared three ways. The mass may have kinetic energy and gravitational potential energy, while the spring itself has elastic energy associated with its stretch and compression. Here too, the force acting on the mass is not constant, and the methods developed in earlier chapters for treating motion resulting from the action of a constant force are inadequate to the problem. The principle of conservation of energy once again enables us to find a solution.

Illustrative Example. A 6-lb weight, when placed on a vertical spring, stretches it 2 in., at which point it is in equilibrium. The weight is then pulled down an additional 4 in. and released. (a) Determine the energy of the vibrating system. (b) What is the maximum upward displacement of the weight? (c) With what speed is the weight moving when it is 1 in. above the equilibrium position?

(a) The spring constant k may be determined from Equation (7-10). From the statement of the problem, we note that a force of 6 lb stretches the spring

2 in. or $\frac{1}{6}$ ft, and we find that

$$k = \frac{6\text{ lb}}{2\text{ in.}} = 3\,\frac{\text{lb}}{\text{in.}} = 36\,\frac{\text{lb}}{\text{ft}}\cdot$$

Let us take the zero level of energy as the neutral position of the spring, as shown in Figure 7-10. The vertical displacement of the weight will be called h, taken

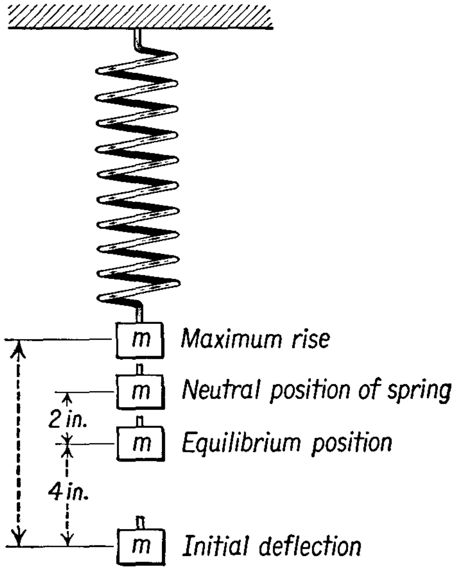


Fig. 7-10 The motion of an oscillating spring.

negative downward from the neutral position and positive upward from the neutral position. The energy $\mathcal{E}$ of the system at any position is given by

$$\mathcal{E} = \tfrac{1}{2}mv^2 + mgh + \tfrac{1}{2}kh^2. \tag{a}$$

We may evaluate $\mathcal{E}$ at the position of maximum displacement where $v = 0$. Noting that $mg = 6$ lb, $h = -\frac{1}{2}$ ft, and $k = 36\,\frac{\text{lb}}{\text{ft}}$, we find that

$$\begin{aligned}\mathcal{E} &= 0 + 6\text{ lb} \times (-\tfrac{1}{2}\text{ ft}) + \tfrac{1}{2} \times 36\,\frac{\text{lb}}{\text{ft}} \times (-\tfrac{1}{2}\text{ ft})^2\\ &= -3\text{ ft lb} + 4.5\text{ ft lb},\end{aligned}$$

so that $$\mathcal{E} = +1.5\text{ ft lb}.$$

(b) Since the system is acted on by no frictional forces, the total mechanical energy of the system must be constant. At the extremes of the motion, the weight is instantaneously at rest, and we may find these extremes by finding the values of h at which $v = 0$. Substituting these values in Equation (a), we get

$$1.5\text{ ft lb} = 0 + 6\text{ lb} \times h + \tfrac{1}{2} \times 36\,\frac{\text{lb}}{\text{ft}} \times h^2.$$

Omitting dimensions to simplify the expression, and transposing,

we get $$18h^2 + 6h - 1.5 = 0;$$

and, applying the quadratic formula, we find

$$h = \tfrac{1}{6}, \quad \text{and} \quad h = -\tfrac{1}{2}$$

for the two solutions; that is, the weight oscillates from a point 2 in. above the undisplaced spring position to a point 6 in. below the undisplaced spring position, or 4 in. above and below the equilibrium position of the hanging weight.

(c) When the weight is 1 in. above its equilibrium position, it is 1 in. below the neutral position of the spring. To find the speed v with which the weight is moving at this point, we again apply the condition for conservation of mechanical energy, as given by Equation (a). The total energy $\mathcal{E}$ of the system is still 1.5 ft lb, but now some of the energy is kinetic; the value of h is now $-\frac{1}{12}$ ft. The equation now becomes

$$1.5 \text{ ft lb} = \tfrac{1}{2} \times \tfrac{6}{32} \text{ slug} \times v^2 - 6 \text{ lb} \times (\tfrac{1}{12} \text{ ft}) + \tfrac{1}{2} \times 36 \frac{\text{lb}}{\text{ft}} \times (\tfrac{1}{12} \text{ ft})^2,$$

from which $$v^2 = 20 \frac{\text{ft}^2}{\text{sec}^2},$$

yielding $$v = \pm 4.47 \frac{\text{ft}}{\text{sec}};$$

that is, the weight may be rising or falling with a speed of 4.47 ft/sec when the weight is 1 in. above its equilibrium position.

7-11 Power

In many cases the time in which a given amount of work is done is of great importance. The term *average power* is defined as the average rate of doing work, or *the work done divided by the time during which the work is done.*

$$\overline{\mathcal{P}} = \frac{\mathcal{W}}{t}, \tag{7-13}$$

in which $\overline{\mathcal{P}}$ represents the average power and t the time in which the work $\mathcal{W}$ is done. We define the instantaneous power $\mathcal{P}$ as

$$\mathcal{P} = \lim_{\Delta t \to 0} \frac{\Delta \mathcal{W}}{\Delta t} = \frac{d\mathcal{W}}{dt}. \tag{7-14}$$

The units of power are the units of work divided by the time. Thus, in the British gravitational system, the unit of power is foot pounds per second; in the cgs system power is expressed in ergs per second; and in the mks system power is expressed in joules per second. Only in the mks system of units is the unit of power given a special name, the *watt.* By definition *one watt equals one joule per second.* Note that although watts and kilowatts

(abbreviated kw; equalling 1,000 watts) are widely used in connection with electrical apparatus, there is nothing particularly electrical about the definition of the watt. The watt is a unit of power which has been adopted by manufacturers of electrical equipment.

Another unit of power which is widely used in engineering but which is not an appropriate unit in any of our systems of units is the *horsepower* (hp) which owes its use to tradition rather than to any currently logical reason. When steam engines were first introduced, they were used to replace horses. Careful measurements of the rate at which a horse could do mechanical work were made by James Watt (1736–1819) who concluded that an average draft horse could exert a force of 150 lb while walking at a rate of 2.5 mi/hr. Thus 550 ft lb/sec or 33,000 ft lb/min were defined as the horsepower. These definitions are in current use today. The relationship between the foot pound per second, the watt, and the horsepower are

$$746 \text{ watts} = 1 \text{ hp} = 550 \frac{\text{ft lb}}{\text{sec}} \cdot$$

Such units as horsepower hour and kilowatt hour, often found in technical use, are units of power times units of time and therefore represent work.

We can derive an equation which is often quite useful, relating the power expended, the applied force, and the velocity. If the applied force and the displacement are in the same direction, we have seen that

$$d\mathscr{W} = F\,ds,$$

and, dividing by the time interval dt, we find

$$\mathscr{P} = \frac{d\mathscr{W}}{dt} = F\frac{ds}{dt},$$

so that

$$\mathscr{P} = Fv. \tag{7-15}$$

The power delivered to a body is the product of the force acting on it and the velocity of the body, when the force and the velocity are in the same direction. If we wish to take into account the possibility of different directions for the force and velocity, we perform the same operation of division by time on the vector form of the equation defining work. From Equations (7-3)

$$d\mathscr{W} = \mathbf{F}\cdot d\mathbf{s} = F\,ds\cos\theta,$$

and, dividing *by dt*,

$$\mathscr{P} = \mathbf{F}\cdot\mathbf{v} = Fv\cos\theta. \tag{7-16}$$

Illustrative Example. An engine is delivering 1,200 hp to an airplane in level flight at a uniform speed of 300 mi/hr. Determine the total of all the resistive forces (drag) acting on the airplane.

An airplane flying with uniform velocity is in equilibrium. Hence the force

supplied by the airplane engine and propeller must be equal and opposite to all the resisting forces acting on the airplane. This force can be found by means of Equation (7-15). The force supplied by the airplane is in the same direction as the velocity. The units supplied in the problem must first be converted to the system of units in which we will work, the British gravitational system.

$$\mathcal{P} = 1{,}200 \text{ hp} = 1{,}200 \text{ hp} \times \frac{550 \text{ ft lb/sec}}{1 \text{ hp}},$$

$$\mathcal{P} = 1{,}200 \times 550 \frac{\text{ft lb}}{\text{sec}} \cdot$$

$$v = 300 \frac{\text{mi}}{\text{hr}} = 300 \frac{\text{mi}}{\text{hr}} \times \frac{1 \text{ hr}}{3{,}600 \text{ sec}} \times \frac{5{,}280 \text{ ft}}{1 \text{ mi}},$$

$$v = 440 \frac{\text{ft}}{\text{sec}} \cdot$$

From Equation (7-15)

$$F = \frac{\mathcal{P}}{v} = \frac{1{,}200 \times 550 \text{ ft lb/sec}}{440 \text{ ft/sec}} = 1{,}500 \text{ lb}.$$

7-12 Simple Machines

In common terminology the words "engine" and "machine" are often used interchangeably. For present purposes we shall define an engine as a device which converts other forms of energy into mechanical energy, and a machine as a device which transmits mechanical energy. Neither an engine nor a machine is capable of delivering more energy in an interval of time than it consumes in that same interval of time, for this would be a violation of the principle of conservation of energy. In general, any device always delivers less energy than is supplied to it. We find it convenient to define the mechanical efficiency of a device as

$$e = \frac{\text{work delivered by the device}}{\text{work supplied to the device}} \cdot \qquad \textbf{(7-17)}$$

For an ideal frictionless machine the efficiency $e = 1$.

Most complicated machines can be considered to be made up of a combination of simple elements, called *simple machines*, such as the inclined plane, the lever, the pulley system, and the screw and nut. There are usually two reasons for using a simple machine: one is that the magnitude of the force which can be applied without the aid of a machine is insufficient to move the required load, and the other is that the direction of the applied force is not appropriate to the desired result. Input quantities will be distinguished by the subscript i, and output quantities by the subscript o. If a force F_i is applied to the machine through a distance s_i, the input

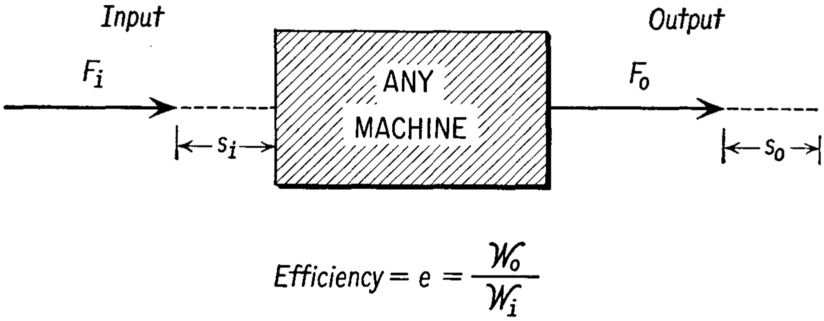


Fig. 7-11 Simple machines.

work will be $\mathcal{W}_i$; the machine will perform an amount of work $\mathcal{W}_o$ by applying a force F_o through a distance s_o, as indicated in Figure 7-11.

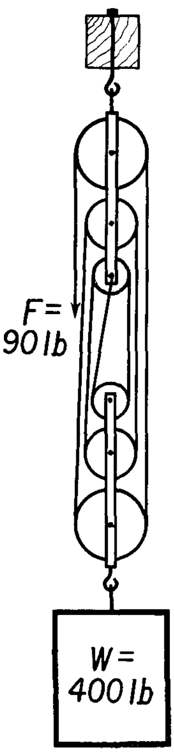


Fig. 7-12 Pulley system.

Illustrative Example. The pulley system of Figure 7-12 is used to hoist a safe weighing 400 lb by the application of a 90-lb force to the free end of the rope. Determine the efficiency of the pulley system.

To determine the efficiency of the system, we note that when the free end of the rope moves a distance $6h$, the weight moves a distance h. The work done on the machine is

$$\mathcal{W}_i = 90 \text{ lb} \times 6h = 540h \text{ lb}.$$

The work done by the machine is

$$\mathcal{W}_o = 400 \text{ lb} \times h = 400h \text{ lb}.$$

The efficiency of the machine is

$$e = \frac{\mathcal{W}_o}{\mathcal{W}_i} = \frac{400h}{540h} = 0.74.$$

Problems

7-1. A trunk weighing 150 lb is pulled across a floor for a distance of 12 ft by a horizontal force of 50 lb. (a) How much work is done? (b) If the trunk was pulled across the floor at uniform speed, what is the coefficient of kinetic friction between the trunk and the floor?

7-2. A man pulls a sled by means of a cord attached to it, exerting a force of 16 lb at an angle of 60° with the horizontal. How much work is done in pulling this sled for a distance of 250 ft?

7-3. (a) Calculate the work done in lifting a body whose weight is 140 lb through a height of 8 ft. (b) How much is the increase in its potential energy?

7-4. Determine the kinetic energy of an airplane whose weight is 30 tons if it is moving with a speed of 250 mi/hr.

7-5. A body weighing 100 lb is pushed up a rough inclined plane by a force of 75 lb acting parallel to the plane. The plane is inclined at an angle of 30° with the horizontal and is 24 ft long. (a) How much work is done in moving the body to the top of the inclined plane? (b) What is its potential energy when at the top of the plane? (c) How much work was done against friction?

7-6. A body weighing 75 lb slides down an inclined plane 16 ft high and 80 ft long. It reaches the bottom of the incline with a speed of 24 ft/sec. (a) What is its potential energy at the top of the inclined plane? (b) How much kinetic energy does it possess when it reaches the bottom of the plane? (c) Determine the force of friction between the body and the plane.

7-7. A box weighing 150 lb slides down an incline 20 ft long from the second floor of a building to the first floor 12 ft below. The frictional force exerted on the box by the incline is 48 lb. (a) How much potential energy does the box lose in sliding down? (b) How much energy is used up in moving the box against the frictional force? (c) How much kinetic energy does the box have when it gets to the bottom? (d) What is the coefficient of kinetic friction between the box and the plane?

7-8. A ball is thrown upward at an angle of 60° with the horizontal at a speed of 50 ft/sec. (a) From energy considerations, find the speed of the ball when it reaches the top of its path. (b) How high will the ball go?

7-9. A simple pendulum consists of a thin string of negligible mass with a steel ball of 450 gm mass attached to one end. The distance from the point of support to the center of the ball is 100 cm. The ball is pulled aside until the string makes an angle of 37° with the vertical. (a) How much potential energy does the pendulum have in this position? (b) With what velocity will the ball reach the lowest position after it is released? (c) How fast will the ball be moving when the string makes an angle of 10° with the vertical?

7-10. A helical spring hangs vertically with its lower end at $y = 0$. When a 45-lb weight is attached to it and lowered gently, the spring is stretched 1.5 in. The spring is pulled down an additional 4.0 in. and is released. (a) Determine the constant of the spring in lb/ft. (b) Determine the total energy of the system, taking the zero level of energy at $y = 0$. (c) Determine the y coordinate of the highest position the body will attain after being released. (d) Determine the speed of the body when $y = -2$ in.

7-11. What power must be delivered to a car which is moving at a speed of 45 mi/hr if the sum of all the resisting forces acting on the car is 180 lb?

7-12. How much power must a man weighing 160 lb develop if he runs up a flight of stairs 9 ft high in 5 sec?

7-13. The engines of a fighter plane deliver 2,000 hp to keep the plane in level flight with a constant velocity of 400 mi/hr. The plane weighs 5 tons. (a) Determine the sum of all the forces opposing the motion of the plane. (b) Assuming that these drag forces remain constant, what will be the speed of the plane

when it climbs at an angle of 5° with the horizontal? (c) What will be the speed of the plane when it descends at an angle of 5° with the horizontal?

7-14. A body weighing 96 lb drops from a height of 4 ft above the top of a spring and compresses it. If the constant of the spring is 12 lb/in., determine the decrease in length of the spring.

7-15. When a body is attached to a tension spring and gently lowered to its equilibrium position, the spring is stretched by an amount s. If the same body is attached to the spring and permitted to drop, show that the maximum deflection of the spring is $2s$.

7-16. The pulley system sketched in Figure 7-13 consists of an upper fixed block containing two pulley wheels and a lower movable block also containing two pulley wheels. (a) A body of weight W is attached to the movable block. How big a force F would have to be applied to lift the weight uniformly it there were no friction? (b) In one such pulley system a force of 80 lb was needed to lift a 240-lb weight. What was the efficiency of this simple machine?

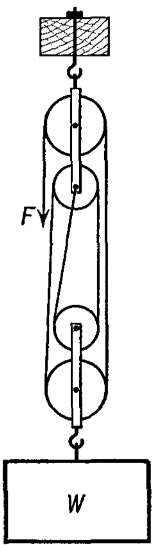


Fig. 7-13 Pulley system.

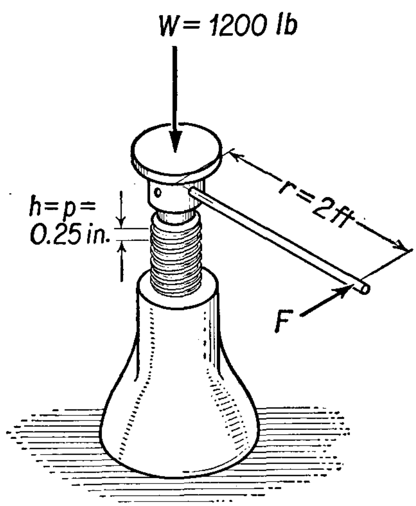


Fig. 7-14 Jack screw.

7-17. The screw thread of an automobile jack has a pitch of 0.25 in. (the pitch is the distance the screw moves forward in one complete rotation) and is operated by a lever 2 ft long, as shown in Figure 7-14. If the efficiency of the jack is 0.30, determine the force that must be applied to the end of the lever to lift a load of 1,200 lb.

7-18. The hammer of a pile driver weighs 1,200 lb and falls through a height of 6 ft to drive a pile into the ground. (a) How much energy does the hammer have when it strikes the pile? (b) If the pile is driven a distance of 6 in., determine the average resisting force acting on the pile.

7-19. A chemical balance customarily weighs to a sensitivity of about 0.1 mg. How much energy, in ergs, would be liberated in a chemical reaction if the change in mass were just detectable?

7-20. Two vectors **A** and **B** are drawn in the xy plane radiating out from the origin with the heads of the two vectors lying at the points whose coordinates are $(x, y) = (4, 0)$ and $(5, 12)$, respectively. Find (a) the scalar product $\mathbf{A} \cdot \mathbf{B}$ and (b) the vector product $\mathbf{A} \times \mathbf{B}$ between the two vectors.

7-21. Two vectors are drawn in the x-y plane. The tail of vector **A** lies at point (6, 0) while its head lies at point (9, 4). The head of vector **B** is at the origin, while its tail is at point (0, 6). Find (a) the dot product $\mathbf{A} \cdot \mathbf{B}$, (b) the cross product $\mathbf{A} \times \mathbf{B}$, and (c) the cross product $\mathbf{B} \times \mathbf{A}$ between the two vectors.

7-22. The output of an electric motor is 5 hp. Determine the velocity with which a load of 400 lb can be lifted?

7-23. An automobile weighing 2,500 lb is driven by an engine which develops 50 hp. On level ground the automobile has a maximum speed of 75 mi/hr. What is the greatest speed with which the automobile can climb a 10 per cent grade? (A 10 per cent grade is one which rises 1 ft in 10 ft along the incline.)

7-24. Starting with the definitions of the watt and the horsepower, show that 746 watts = 1 horsepower.

7-25. A ball weighing 4 ounces is at rest on the floor of a train which is moving at 60 mi/hr. What is the kinetic energy of the ball (a) as determined by an observer on the train? (b) By an observer on the ground? (c) Explain the discrepancy between the two cases.

7-26. The altitude at which the potential energy is zero may be set arbitrarily by choosing some level, such as sea level, as the level of zero altitude. In what way may we establish an arbitrary zero of kinetic energy? (HINT: Are the laws of mechanics equally valid in two coordinate frames moving at uniform speed with respect to each other?)

8

Hydrostatics (Fluids at Rest)

8-1 Three Phases of Matter

From our everyday experience, we have become familiar with the fact that matter occurs in three different forms—*solid*, *liquid*, and *gas*. Under ordinary conditions stone, iron, copper, and chalk, for example, are solids; water, oil, and mercury are liquids; air, hydrogen, and carbon dioxide are gases. Each one of these forms is called a *phase*. At times it is difficult to distinguish clearly between the solid and the liquid phases, as in a material such as tar which flows under the action of a force at ordinary temperatures. Metals at high temperatures flow or "creep" under the action of a force. Even where the different phases are clearly recognizable, materials undergo a phase change under different conditions of temperature and pressure. For the present we shall confine our discussion to the application of the principles of mechanics to bodies which remain in the same phase.

Liquids and *gases* are sometimes grouped together as *fluids* because they *flow* very readily under the application of an external force, while solids do not. A solid has a definite size and a definite shape, and these change only slightly under the application of external forces. For this reason it is possible to study the statics of *solids* by characterizing them as *rigid* bodies. *Liquids*, on the other hand, possess a definite size or volume but change their shape very readily. Liquids at rest generally take the shape of the containing vessel. If the containing vessel has a volume greater than that of the liquid put into it, there will be a *free surface* at the top of the liquid. A gas differs from a liquid in that a gas has neither size nor shape. A quantity of gas placed in a container will completely fill that container. There is no free surface. The volume of the gas is the volume of the container.

8-2 Pressure

There is a difference in the manner in which a force is applied to a fluid and the way it is applied to a solid. A force can be supported by a single point of a free solid, but it can only be supported by a surface of an enclosed fluid. In a discussion of the results of the application of forces to fluids, it is convenient to introduce a new term called *pressure.* If a force F is applied to the surface of a fluid and acts over an area A perpendicular to it, then the average pressure $\bar{P}$ is defined as

$$\boxed{\bar{P} = \frac{F}{A}.} \tag{8-1}$$

The pressure may be expressed in dynes per square centimeter, in pounds per square foot, in newtons per square meter, or in any other appropriate set of units. Pressure is a scalar quantity.

When a fluid is under pressure, it exerts a force on any surface which contains the fluid. Equation (8-1), which describes the average pressure, is not quite complete, for, while relating the magnitude of the force exerted to the pressure and the area, it must be accompanied by a statement about the direction of the force. We may make the equation more complete by considering *area as a vector quantity.* An element of area $\Delta\mathbf{A}$ may be described as a vector whose magnitude is the numerical value of the area and whose direction is perpendicular to the surface of the area element, as shown in Figure 8-1. If the element of area is part of a closed surface, it is conventional to choose the *outward normal* to the surface as the direction of the area vector. With this convention we may write

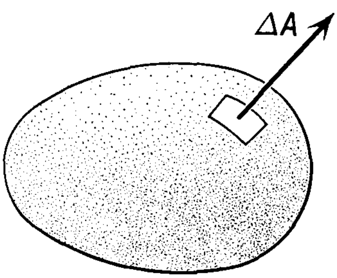


Fig. 8-1 An area vector $\Delta\mathbf{A}$ is directed along the outward normal to the surface.

$$\Delta\mathbf{F} = P\,\Delta\mathbf{A} \tag{8-2}$$

for the force $\Delta\mathbf{F}$ exerted by a fluid of pressure P against any surface element of area $\Delta\mathbf{A}$.

The pressure may vary from point to point within a fluid. We speak of the pressure at a point within a volume of fluid, meaning that we imagine the point to be surrounded by a small container and divide the total force exerted by the fluid against the walls of the container by the area of the container. Following the usual limiting processes of the calculus, we exam-

ine this quotient as the volume of the container gets smaller and smaller, and we call its limiting value the pressure at the point.

8-3 Density

In discussing distributions of matter such as solids or fluids, it is convenient to define a quantity called the density ρ (rho) as the mass per unit volume or,

$$\rho = \frac{m}{V}. \tag{8-3}$$

Thus the mass of a homogeneous body of material of volume V is given by $m = \rho V$. The units in which density is expressed are the ones appropriate to the system of units being used. In the cgs system of units the density is expressed in grams per cubic centimeter; in the mks system of units the density is expressed in kilograms per cubic meter; in the British gravitational system of units the density is expressed in slugs per cubic foot.

It is common engineering practice to use the word "density" to express the weight per unit volume in pounds per cubic foot. We distinguish this

TABLE 8-1 DENSITIES OF SOME COMMON SUBSTANCES

Solids	Density in gm/cm^3	Liquids	Density in gm/cm^3
Aluminum	2.70	Alcohol	0.79
Brass	8.44–8.70	Ether	0.74
Copper	8.93	Glycerin	1.26
Cork	0.22–0.26	Mercury	13.596
Glass, common	2.4–2.8	Oil, olive	0.92
Glass, flint	2.9–5.9	Oil, paraffin	0.8
Gold	19.3	Water	1.00
Ice	0.917		
Iron	7.03–7.9	**Gases**	**Density at 0°C, 760 mm Hg in gm/cm^3**
Lead	11.34		
Osmium	22.5		
Nickel	8.9		
Platinum	21.50	Air	0.001293
Silver	10.5	Ammonia	0.000771
Tungsten	18.6–19.1	Carbon dioxide	0.001977
Wood, cedar	0.49–0.57	Helium	0.000179
Wood, ebony	1.11–1.33	Hydrogen	0.000090
Wood, elm	0.54–0.60	Oxygen	0.001429
Wood, white pine	0.35–0.50		
Zinc	7.1		

quantity from the density defined above by referring to the weight per unit volume as the *weight density*.

The density of solids varies only slightly with changes in temperature and pressure and is customarily given in tables as the result of measurement at a standard pressure and temperature. The density of metals ranges over a factor of about 10 from the lightest, magnesium, which has a density of 1.80 gm/cm^3 to the most dense, osmium, which has a density of 22.5 gm/cm^3. The density of commercial metals and alloys depends upon their composition. Some values of the density of various substances are given in Table 8-1.

8-4 Pressure Due to Weight of a Liquid

Let us consider the equilibrium conditions of a small element of liquid volume submerged within the body of the liquid, as shown in Figure 8-2. Since the volume element is at rest, the horizontal forces exerted upon this volume element by the surrounding liquid must have a zero resultant. In the vertical direction we note that there must be a difference in the forces exerted on the top and bottom faces of the volume element sufficient to

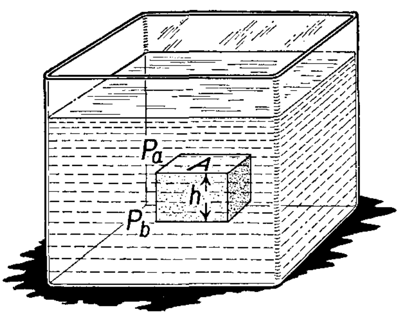


Fig. 8-2

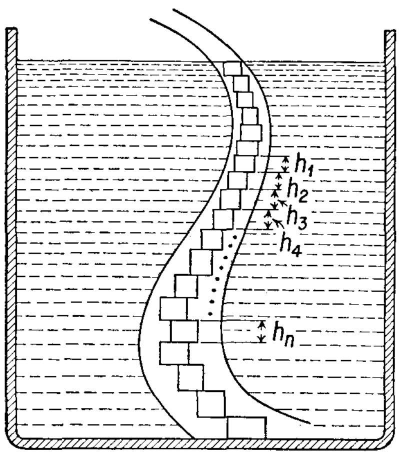


Fig. 8-3

support the weight of the liquid within that volume element. If the pressure at the level of the top face is P_a and the pressure at the level of the bottom face is P_b, the downward force exerted on the top face of the volume element by the surrounding liquid is P_aA, while the upward force exerted against the bottom face is P_bA. The difference between these forces must be equal to the weight of the liquid contained within the element of height h.

Thus we have

$$P_bA - P_aA = Ah\rho g,$$

and, dividing by A,

$$P_b - P_a = h\rho g. \tag{8-4}$$

Thus the pressure difference between two adjacent level surfaces in the liquid is given by $h\rho g$, where h represents the vertical distance between the two surfaces.

In the event that we wish to find the pressure within a liquid confined in a tube of irregular shape, such as the S-shaped tube of Figure 8-3, we may imagine the liquid to be subdivided into a succession of volume elements, one atop the other, each of which contributes a small increment of pressure, depending on its vertical height, to the pressure at the depth h.

Thus the pressure at all points at a given depth beneath the surface of a liquid depends upon the depth but not upon the shape of the container; *all points at the same horizontal level surface within a body of liquid are at the same pressure, as long as the liquid is at rest.* In fact, this statement is often taken as the definition of a level surface. If the pressure on the surface of a liquid is taken as zero, then the pressure P at any point a distance h below this level is given by

$$\boxed{P = h\rho g.} \tag{8-5}$$

Illustrative Example. Determine the pressure at the bottom of a column of mercury 70 cm high.

From Equation (8-5) we have

$$P = h\rho g.$$

We substitute the value of $h = 70$ cm, the density ρ of mercury $= 13.6$ gm/cm^3, and $g = 980$ cm/sec^2, and obtain

$$P = 70 \text{ cm} \times 13.6 \frac{\text{gm}}{\text{cm}^3} \times 980 \frac{\text{cm}}{\text{sec}^2},$$

$$P = 9.33 \times 10^5 \frac{\text{dynes}}{\text{cm}^2}.$$

8-5 Pressure in a Confined Liquid

In addition to the pressure due to its weight, a confined liquid may be subjected to an additional pressure by the application of an external force. Suppose the liquid is in a cylinder, as shown in Figure 8-4, and that a tight-fitting piston is placed on the surface of the liquid. If a force F is applied to the piston, it will remain in practically the same position, since the compressibility of liquids is very small. If A is the area of the piston, this

external force produces a pressure $P = F/A$ at the surface of the liquid. As we have seen from the preceding discussion, as we examine the pressure at increasing depth beneath the surface of the liquid, each element of volume adds a contribution to the pressure which is due to its own weight

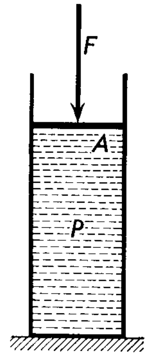


Fig. 8-4 A pressure P is produced in a confined liquid by the application of a force F on a piston of area A in contact with the liquid.

to the pressure at the top surface of that volume, so that the pressure P due to the external force is transmitted throughout every part of the liquid and acts on all surfaces in contact with the liquid. This is sometimes known as *Pascal's principle* and may be stated as follows:

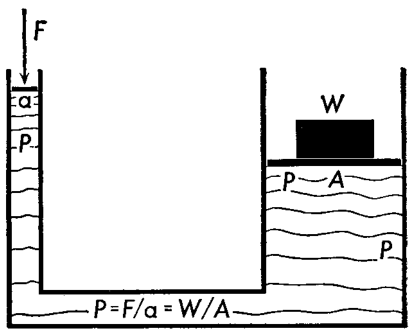


Fig. 8-5 Hydraulic press. The pressure P is the same everywhere throughout the liquid, if the pressure due to the weight of the liquid is negligible.

Whenever the pressure in a confined liquid is increased or diminished at any point, this change in pressure is transmitted equally throughout the entire liquid.

The operation of the hydraulic press, the hydraulic brakes of a car, and the hydraulic lift is based upon Pascal's principle. The hydraulic press, sketched in Figure 8-5, consists essentially of two connected cylinders, one of small cross-sectional area a, the other of large cross-sectional area A, each fitted with a piston. A liquid, usually oil or water, is supplied to it from a reservoir. By exerting a force F on the small piston, an additional

pressure $P = F/a$ is produced. This pressure is transmitted throughout the liquid and hence acts on the larger piston of area A. The force that can be exerted by the larger piston is then PA. If the hydraulic press is designed to lift a weight W, then

$$W = PA = F\frac{A}{a}.$$

The hydraulic press may be considered as a simple machine in which the force exerted by the machine divided by the force exerted on the machine, is equal to the ratio of the areas of its pistons; thus

$$\frac{W}{F} = \frac{A}{a}.$$

A force-distributing system such as that used to operate the brakes of a car offers the very great advantage that the force applied on each of the brakes is automatically equal. Furthermore, the force is transmitted with very little mechanical movement of the hydraulic link.

8-6 Atmospheric Pressure

The atmosphere is a layer of air surrounding the earth; its thickness has been estimated as about 500 to 600 mi. The density of the air decreases with increasing altitude. Since air has weight, this layer of air produces a pressure, called the atmospheric pressure, at the surface of the earth. The atmospheric pressure varies from day to day by about 5 per cent, the variations often accompanying changes in the weather. The pressure of the air is measured by a *barometer*, which often consists of an evacuated tube inverted in a dish of mercury, as shown in Figure 8-6.

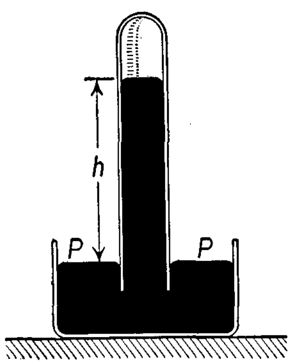


Fig. 8-6 Mercury barometer.

The atmosphere exerts a pressure P on the open surface of the mercury in the dish, and this is transmitted to the liquid in the tube. This pressure is balanced by the pressure due to the mercury in the tube at a height h above the open surface of the dish. To look at the barometer another way, we recall that the pressure is always the same at any level surface in a liquid. Consider the level surface defined by the surface of the mercury in the dish. Outside the barometer tube the pressure at this surface is entirely due to the atmosphere, so that the pressure of the mercury here is atmospheric pressure. Hence the pressure of the mercury within the tube at the level

of the surface of the mercury in the dish is also atmospheric pressure. Knowing the density of mercury and the height to which the column of mercury rises within the evacuated barometer tube, we can calculate the atmospheric pressure.

Since the atmospheric pressure varies from day to day and from place to place, scientific data are often corrected and reported for a standard atmospheric pressure, the pressure of the atmosphere when the mercury barometer stands 76.0 cm above the free surface of mercury in the dish. The density of mercury at 0°C is 13.60 gm/cm^3, so that the pressure of a standard atmosphere at 0°C is

$$P = 76 \text{ cm} \times 13.60 \frac{\text{gm}}{\text{cm}^3} \times 980 \frac{\text{cm}}{\text{sec}^2},$$

$$P = 1{,}013{,}000 \frac{\text{dynes}}{\text{cm}^2}.$$

This pressure can also be expressed as

$$P = 14.70 \frac{\text{lb}}{\text{in.}^2}.$$

A pressure of 1.013×10^6 dynes/cm^2, or 14.70 lb/in.^{2}, or the pressure at the base of a column of mercury 76 cm high at 0°C is called a pressure of one *atmosphere* (abbreviated atm). In meteorology other units of pressure are used; they are mentioned here only for the sake of completeness. *A pressure of one bar is defined as one million dynes per square centimeter. A millibar is one thousandth of a bar.* Standard atmospheric pressure can thus be expressed as 1,013 millibars.

8-7 Pressure Gauges: Manometers

We have already encountered one type of pressure gauge, the mercury barometer. A second type of pressure gauge is the *aneroid* barometer, which consists of a partially evacuated cylindrical box made of corrugated metal, as shown in Figure 8-7. The difference in pressure between the inside of the chamber and the atmosphere causes the faces of the box to deflect, and this is balanced by a steel spring. The motion of the upper surface of the box is coupled to a pointer whose indication is read on a scale. Such a gauge is often used as a pressure altimeter on board aircraft, for atmospheric pressure varies with altitude, and so can be used as a measure of altitude. Another type of gauge, a *Bourdon* gauge, consists essentially of a flattened brass tube, closed at one end and bent into circular form. When a fluid under pressure is admitted to the open end of the tube, the tube straightens

slightly, and this motion is coupled to a pointer whose deflection is calibrated.

The simplest type of pressure gauge is the *open-tube manometer*, illustrated in Figure 8-8, which consists essentially of a bent tube of transparent material with both arms vertical. One end A is open to the atmosphere, and the other end B is connected to the vessel in which the pressure

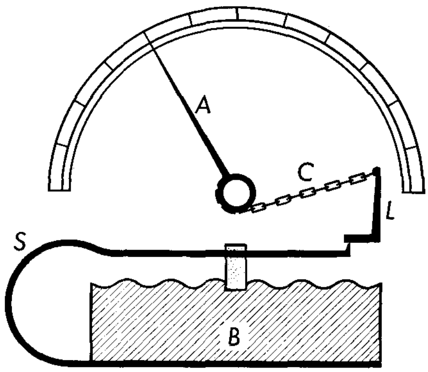


Fig. 8-7 Aneroid barometer. B is a box containing air at reduced pressure, S is a spring, L a lever, C a chain, and A a pointer.

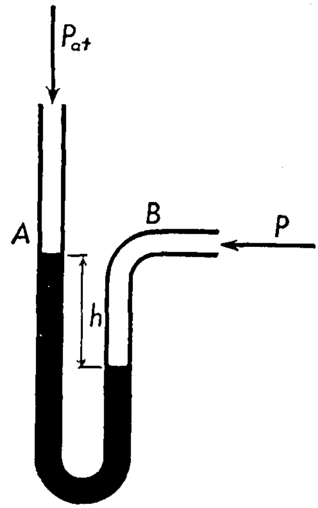


Fig. 8-8 Open-tube manometer.

is to be measured. The tube is partially filled with a liquid of density ρ. Suppose that the pressure within the vessel is greater than the atmospheric pressure, causing the liquid to rise in the column exposed to the atmosphere. Again we note that the pressure within a body of liquid is the same everywhere along a level surface. Choosing as the level surface of reference that of the lower liquid surface of the manometer, we observe that the pressure within the vessel P must be equal to the pressure of the atmosphere plus the pressure due to a column of liquid of height h, the difference in level between the liquid surfaces in the two arms of the manometer tube. Thus we have

$$P = P_{\text{atm}} + h\rho g. \tag{8-6}$$

In many technical applications it is the difference in pressure between the inside of a container and the atmosphere which is of importance, rather than the pressure itself. This is the case in a steam boiler, or in a

gas line, and in many other applications of the manometer. The difference $P - P_{\text{atm}}$ is called the *gauge pressure.* The pressure P is called the *absolute pressure.* A barometer thus reads the absolute pressure of the atmosphere, while the height h of an open-tube manometer is a measure of the gauge pressure. The gauge pressure of the atmosphere itself is, of course, zero.

8-8 Archimedes' Principle

The fact that some objects float in water while others sink to the bottom has been known for centuries; Archimedes (287–212 B.C.) was the first to discover the principle underlying these phenomena. To understand Archimedes' principle it is necessary to consider the forces acting on a body totally immersed in a liquid, as shown in Figure 8-9. There is a downward force on the body equal to its own weight W, and, in addition, there is a buoyant force B on the body, which acts in the upward direction. This can be understood by imagining the volume now occupied by the body to be occupied instead by an equal volume of liquid. This volume of liquid would have been in equilibrium, which means that its weight would have been supported by the action of the rest of the liquid. This support comes from the difference in pressure between the top and the bottom of this volume. Hence, no matter what material occupies this volume, there will be a force upward on it equal to the weight of the liquid displaced.

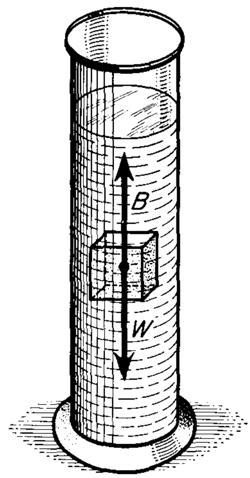


Fig. 8-9 Buoyant force on a block immersed in a liquid.

Archimedes' principle is a generalization of the result obtained above; it states that *any object partly or completely immersed in a fluid is buoyed up by a force equal to the weight of the fluid displaced.* The principle is applicable to both liquids and gases.

A body completely immersed in a fluid will sink if its weight is greater than the buoyant force and will rise if the buoyant force is greater than its weight. Equilibrium will be established when the weight of the liquid displaced is equal to the weight of the body. Thus a solid body which is less dense than water floats on the surface of the water, part below the water and part above the water. A ship afloat, for example, displaces its own weight of water. The weight of a ship is frequently expressed in terms of the weight of the water it displaces, and one speaks of ships with 10,000 tons displacement, and so on. There is usually a definite water line painted

on a ship, indicating the limit to which a ship may be submerged and still be safe.

A submarine is so designed that it can take water into specially built tanks so as to make its weight greater than the weight of the water it displaces when fully submerged. Because the density of water is essentially independent of pressure, water being an almost incompressible fluid, a submarine which has no forward motion when submerged will sink to the bottom of the sea. To enable the submarine to rise to the surface, water is forced out of the tanks by pumps.

Illustrative Example. A cylinder of brass 6 cm high and 4 cm² in cross-sectional area is suspended in water by means of a string so that its upper surface is 7 cm below the surface of the water, as shown in Figure 8-10. Determine (a) the force acting on the top of the cylinder, (b) the force acting on the bottom of the cylinder, and (c) the buoyant force acting on this cylinder.

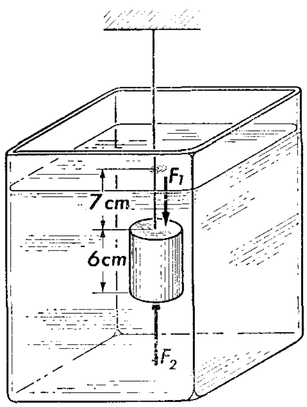


Fig. 8-10

(a) The force F_1 acting on the top of the cylinder is that due to the pressure of the water above it and is

$$F_1 = P_1A = h_1\rho gA$$

$$= 7\text{ cm} \times 1\,\frac{\text{gm}}{\text{cm}^3} \times 980\,\frac{\text{cm}}{\text{sec}^2} \times 4\text{ cm}^2,$$

$$F_1 = 27{,}440\text{ dynes.}$$

This force pushes down on the cylinder.

(b) The force F_2 acting on the bottom of the cylinder is due to the pressure of the water above it. Since the depth of the water at the bottom of the cylinder is 13 cm, this force is

$$F_2 = 13\text{ cm} \times 1\,\frac{\text{gm}}{\text{cm}^3} \times 980\,\frac{\text{cm}}{\text{sec}^2} \times 4\text{ cm}^2,$$

$$F_2 = 50{,}960\text{ dynes}$$

acting upward on the cylinder.

(c) The buoyant force B is the net force upward caused by the difference in pressures in the liquid. The forces which act on the walls of the cylinder are all directed horizontally, and their resultant is zero, as can be seen from the symmetry of the figure. Therefore, the buoyant force is simply the difference between the two vertical forces F_1 and F_2, thus

$$B = F_2 - F_1$$

$$= 50{,}960 \text{ dynes} - 27{,}440 \text{ dynes},$$

so that

$$B = 23{,}520 \text{ dynes}$$

and acts upward.

It is interesting to compare this buoyant force with the weight of water displaced. The volume of the cylinder is $h \times A = 24 \text{ cm}^3$. This is also the volume of water displaced. The weight of this displaced water is

$$W = mg = \rho V g$$

$$= 1 \frac{\text{gm}}{\text{cm}^3} \times 24 \text{ cm}^3 \times 980 \frac{\text{cm}}{\text{sec}^2},$$

$$W = 23{,}520 \text{ dynes},$$

which is in agreement with the earlier calculation.

Illustrative Example. A block of aluminum is attached to a balance. When suspended in air, the balance reads 250 gm. When the aluminum block is lowered so that it is completely immersed in water, the balance reads 160 gm. When the aluminum block is lowered so that it is completely immersed in alcohol, the balance reads 180 gm. The density of water is 1 gm/cm^3. Determine (a) the density of aluminum, and (b) the density of alcohol.

(a) The buoyant force of the water is the difference between the weight of the aluminum block in air and its weight when immersed in water; that is,

$$B = 250 \times 980 \text{ dynes} - 160 \times 980 \text{ dynes} = 90 \times 980 \text{ dynes}.$$

Thus the aluminum block displaces a mass of 90 gm of water. The volume of water displaced is

$$V = \frac{m}{\rho} = \frac{90 \text{ gm}}{1 \text{ gm/cm}^3}$$

$$= 90 \text{ cm}^3,$$

and this is equal to the volume of aluminum. Thus the density of aluminum is

$$\rho = \frac{m}{V} = \frac{250 \text{ gm}}{90 \text{ cm}^3},$$

$$\rho = 2.78 \frac{\text{gm}}{\text{cm}^3}.$$

(b) The amount of alcohol displaced by the aluminum block may be found

from the buoyant force exerted by the alcohol, which is given by

$$B = 250 \times 980 \text{ dynes} - 180 \times 980 \text{ dynes} = 70 \times 980 \text{ dynes},$$

so that the mass of alcohol displaced by the aluminum block is 70 gm. The volume of alcohol displaced is the volume of the aluminum block, which we have found to be equal to 90 cm^3. Thus the density of alcohol is

$$\rho = \frac{70}{90} \frac{\text{gm}}{\text{cm}^3},$$

$$\rho = 0.78 \frac{\text{gm}}{\text{cm}^3}.$$

For many technical purposes it is not necessary to make precise measurements of density; it is only important to know whether the density is greater than some minimum value or whether it lies between certain limits. For example, in testing die castings for internal porosity, one technique for rapid inspection of large quantities of castings is a flotation test where the die castings are immersed in a suitable liquid. If the castings are sound, their average density is greater than the density of the liquid, and the castings sink. If the castings contain excessive porosity, their average density is too low, and the defective castings may be skimmed off the surface of the flotation liquid.

8-9 Specific Gravity

It is sometimes convenient to refer the mass of a given object to the mass of an equal volume of water. The term *specific gravity* is used to denote the magnitude of this quotient, and since the specific gravity is a property of the material of which the object is made rather than of the size or shape of the object, we may conveniently define the specific gravity as the *ratio of the density of a body to the density of water.* Thus water has a specific gravity of 1. The specific gravity is a pure number and is independent of the system of units used to measure the density. Since the density of water in the cgs system of units is 1 gm/cm^3, the specific gravity of a substance has the same numerical value as its density in the cgs system.

From the definition of specific gravity, we see that a body of specific gravity less than 1 will float in water, while a body of specific gravity greater than 1 will sink.

The specific gravity of liquids is commonly determined by use of an instrument called the *hydrometer,* which is usually made in the form of a cylinder with a weighted bulb at one end, as shown in Figure 8-11. The depth to which the hydrometer will sink in a liquid depends on its specific gravity, so the hydrometer may be provided with a calibrated scale to read specific gravity directly, or to read some property associated with

specific gravity which is of immediate interest. Since the freezing point of a mixture of antifreeze and water is determined by the fraction of antifreeze in the solution, the hydrometer may be directly calibrated to read the freezing temperature of the mixture. Similarly, the specific gravity of a mixture of alcohol and water may be used to determine the percentage of alcohol in the mixture. A hydrometer is customarily used as a proof tester in alcoholic beverages. On such a scale, "100 proof" means 50 per cent alcohol content.

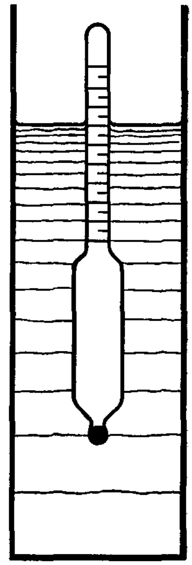


Fig. 8-11 Hydrometer.

8-10 The Centrifuge

In our discussion of rotational motion, we emphasized that a particle could move in uniform circular motion only if there was a centripetal force acting on the particle. Let us consider a particle of mass m placed on a horizontal turntable which is rotating with uniform angular speed ω about a vertical axis through its center. If the particle is at a distance r from the axis, a centripetal force equal to $m\omega^2 r$ must act on this particle to keep it in place on the turntable. This force may be supplied by friction between the particle and the table or, if there is no friction, by a string tied to a shaft at this axis and to the particle. If the motion of the particle is now considered from the turntable as the reference system, the particle seems to be experiencing a force $m\omega^2 r$ away from the center of the table, and it is kept in place by the opposing force produced by the pull of the string on it. The particle appears to experience a new kind of force field whose intensity is now given by $\omega^2 r$ instead of g. We may thus say that an observer located on a rotating coordinate system is led to believe that he is located in a force field directed radially outward from the center of the circle.

If a liquid contains particles of greater density than that of the liquid itself, these particles will separate out at the bottom of the liquid in the earth's gravitational field, given sufficient time. If the volume of the particle is V and its density is ρ, its weight will be given by $\rho V g$. The buoyant force on such a particle in a fluid of density ρ_0 is the weight of the displaced fluid $\rho_0 V g$, and the resultant downward force on the particle is $(\rho - \rho_0)Vg$. The particles of sediment suspended in the fluid are acted upon by a net force in the direction of the gravitational field and, in time, will settle out on the bottom of the container. Because of internal friction in the liquid, the settling-out process may be quite slow. The rate of separation of the solid particles from the liquid might be greatly expedited

by placing the suspension in an intense gravitational field, if such were available. A centrifuge, which is a machine designed to rotate a liquid at high speeds, simulates such a gravitational field, and the apparent force on the particle becomes $(\rho - \rho_0)V\omega^2 r$. Thus if a liquid is placed in a tube,

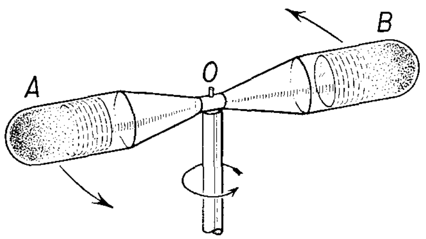


Fig. 8-12 The centrifuge. Tubes A and B are whirled around in a circular path about an axis through O. Particles of greatest density go to the ends of tubes farthest removed from O.

such as that shown in Figure 8-12, and the tube is rotated at high speed, the particles in the liquid will settle out very rapidly. Those particles which are denser than the liquid will be found at the bottom, that is, farthest from the axis of rotation, while those particles which are less dense than the liquid will be found near the top.

Illustrative Example. A liquid containing some solid particles is poured into the cup of a centrifuge which is then rotated at a speed of 6,000 rpm. Determine the apparent gravitational field intensity acting on a particle at a distance of 12 cm from the axis of rotation. Express this in terms of the earth's gravitational field intensity g.

The apparent gravitational field intensity acting on the particle is given by the formula $\omega^2 r$. We have

$$\omega = \frac{6{,}000 \times 2\pi}{60}\,\frac{\text{radians}}{\text{sec}} = 200\pi\,\frac{\text{radians}}{\text{sec}}.$$

Thus
$$\omega^2 r = (200\pi)^2 \times 12\,\frac{\text{cm}}{\text{sec}^2},$$

so that
$$\omega^2 r = 474 \times 10^4\,\frac{\text{cm}}{\text{sec}^2}.$$

Expressing this in terms of g, we have

$$\omega^2 r = \frac{474 \times 10^4}{980}\,g,$$

or
$$\omega^2 r = 4{,}840g.$$

Thus the particle in the centrifuge experiences an apparent gravitational field almost 5,000 times the magnitude of the earth's gravitational field.

Problems

8-1. A block of metal weighs 120 lb in air, 105 lb when immersed in water, and 108 lb when immersed in a certain liquid. Determine (a) the density of the metal, (b) the density of the liquid, and (c) the specific gravity of the liquid.

8-2. A metal sphere whose mass is 36 gm is attached by means of a string to one arm of an equal-arm balance. When the sphere is completely immersed in water, a mass of 23 gm is sufficient to balance it. Determine (a) the volume of the sphere and (b) the density of the metal.

8-3. A raft is made in the form of a rectangular box 8 ft by 10 ft by 4 ft deep. The raft weighs 2,500 lb. (a) How deep will this raft go when placed in fresh water which weighs 62.4 lb/ft^3? (b) What load can this raft carry without sinking? (c) What is the total force exerted by the water on the bottom of the raft when so loaded? (d) Assuming the load is uniformly distributed, will the bottom of the raft tend to bulge in at the center, bulge out, or remain flat?

8-4. A beaker partially filled with water is placed on a scale pan and found to have a mass of 500 gm. A string is attached to a stone and held so that the stone is completely submerged in the water but does not touch the beaker at any point. The scale now reads 550 gm. When the string is released and the stone rests on the bottom of the beaker, the scale reads 620 gm. Determine (a) the mass of the stone, (b) the density of the stone, and (c) the tension in the string.

8-5. A cube of iron 3 cm on an edge is placed in a dish of mercury. (a) How much of the cube is immersed in the mercury? (b) If water is poured over the mercury to a depth of 4 cm, what will be the depth of the iron in the mercury?

8-6. A piece of concrete whose mass is 150 kg has a density of 2,500 kg/m^3. A block of wood of density 500 kg/m^3 is to be fastened to the concrete block and placed in water so that they will both float almost completely submerged. What is the minimum mass of wood which can be used?

8-7. A **U** tube contains mercury at the bottom. Glycerin is poured into one arm so that the height of the glycerin column is 30 cm. How high a column of water must be poured into the other arm to bring the mercury to the same level in both arms?

8-8. What is the atmospheric pressure in dynes per square centimeter when the reading of a mercury barometer is 77 cm?

8-9. Hoover Dam is 1,180 ft long and 726 ft high. (a) What is the pressure at the bottom of the dam when the reservoir is full? (b) Assuming the face of the dam to be a plane rectangle, determine the total force pushing against the face of the dam.

8-10. From Problem 8-9, determine the torque exerted by the force of the water on the face of the dam (a) about a horizontal axis through the bottom of the dam and (b) through the top of the dam. [Hint: Use the methods of the calculus.]

8-11. An open-tube manometer containing water has one end connected to a city gas supply outlet. The difference in level between the two arms is 2.0 in. (a) Determine the gauge pressure of the gas. (b) If the height of the barometer is 76 cm of mercury, determine the absolute pressure of the gas.

8-12. Express the density of lead in slugs per cubic foot and in kilograms per cubic meter.

8-13. If every gram of air in the atmosphere were replaced by mercury, how deep would be the layer of mercury on the surface of the earth?

8-14. If it is desired to determine the mass of 1 cm^3 of aluminum to an accuracy of 1 per cent using a beam balance and calibrated brass weights, should the result of the measurement be corrected for the buoyant effect of the air? At what accuracy does the measurement need to be corrected for buoyancy?

8-15. A submarine having a volume of 165,000 cubic feet and weighing 4830 tons is floating at the surface of sea water of weight density 64 lb/ft^3. (a) What is the buoyant force on the submarine? (b) What volume of sea water must be admitted into the submarine so that it will just submerge?

8-16. A uniform wooden rod which weighs 5 lbs has a length of 8 ft and a volume of 0.1 ft^3. The rod is pinned at one end and is submerged in water, and is held in a horizontal position with the aid of a vertical force F at the opposite end. Assume the pinned end to be fixed. (a) What is the magnitude and direction of the total force on the rod due to the water? (b) What is the magnitude of the force F needed to keep the rod in equilibrium?

8-17. A 2.7 gm block of aluminum is suspended from a spring scale which hangs from the ceiling of an elevator. The block is then immersed in water. (a) What is the reading of the scale when the elevator is at rest? (b) The elevator is accelerated upwards at 490 cm/sec^2. What is the reading of the scale during this time? (c) What is the reading of the scale when the elevator is moving upward at a steady speed of 500 cm/sec? (d) Finally the elevator is brought to rest with an acceleration of 245 cm/sec^2. What is the reading of the scale during this time?

9

Hydrodynamics (Fluids in Motion)

9-1 Steady Flow of a Liquid

When a liquid flows through a pipe in such a way that it completely fills the pipe, and as much liquid enters one end of the pipe as leaves the other end of the pipe in the same time, then *the liquid is said to flow at a steady rate.* At any point of the pipe, the flow of the liquid does not change with time. The path of any particle of liquid as it moves through the pipe is

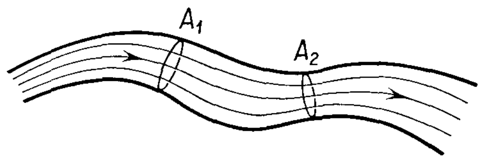


Fig. 9-1 Streamlines of a liquid flowing through a pipe at a steady rate.

called a *streamline.* We can map the flow of liquid through the pipe by drawing a series of streamlines following the paths of the particles of liquid, as shown in Figure 9-1. The instantaneous velocity of a liquid particle is always tangent to the streamline. The rate of flow may be represented by the density of streamlines, or the number of streamlines passing through a surface of unit area perpendicular to the direction of flow. Thus the streamlines will be close together where the liquid is moving rapidly and farther apart in regions of the pipe where the liquid is moving slowly. Similar conventions were used to represent the direction and magnitude of the gravitational field intensity.

Since the liquid is incompressible and there are no places in the pipe where the liquid can be stored, the volume of liquid which flows through any plane perpendicular to the streamlines in any interval of time must be the same everywhere in the pipe. Consider two typical planes whose intersections with the pipe have areas A_1 and A_2 perpendicular to the stream-

lines. The volume of liquid Q passing through area A_1 in unit time is

$$Q = A_1 v_1, \tag{9-1}$$

where v_1 is the velocity of the liquid at this point. Similarly, the volume of liquid passing through A_2 in unit time is

$$Q = A_2 v_2.$$

Since these two quantities must be equal for steady flow, we have

$$A_1 v_1 = A_2 v_2, \tag{9-2}$$

or

$$\frac{v_1}{v_2} = \frac{A_2}{A_1}.$$

Thus the velocity of the liquid at any point in the pipe is inversely proportional to the cross-sectional area of the pipe. The liquid will be moving slowly where the area is large and will be moving rapidly where the area is small.

Illustrative Example. Water flows out of a horizontal pipe at the steady rate of 2 ft^3/min. Determine the velocity of the water at a point where the diameter of the pipe is 1 in.

The area A of the 1-in. portion of the pipe is

$$A = \frac{\pi \times 1}{4 \times 144}\ \text{ft}^2 = 0.0055\ \text{ft}^2.$$

When we substitute the values for Q and A in Equation (9-1),

we obtain
$$\frac{2}{60}\ \frac{\text{ft}^3}{\text{sec}} = 0.0055\ \text{ft}^2 \times v,$$

from which
$$v = 6.10\ \frac{\text{ft}}{\text{sec}}\cdot$$

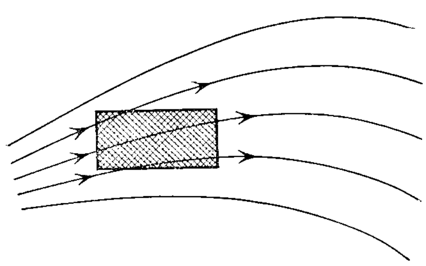


Fig. 9-2 The number of streamlines entering a volume element is equal to the number leaving the volume element when there is steady flow.

In the steady or streamline flow of a liquid, the total quantity of liquid flowing into any imaginary volume element of the pipe must be equal to the quantity of liquid leaving that volume element. If we represent the flow by streamlines, this implies that the streamlines are continuous and do not pile up anywhere within the liquid. The same number of streamlines enter

a volume element as the number which leave it, as shown in Figure 9-2. Another characteristic of streamline flow is that it is layerlike, or *lamellar*. There is no circulation of the liquid about any point in the pipe. We might imagine a small paddle wheel, as shown in Figure 9-3, placed in the pipe. When the flow is lamellar, no rotation of the paddle wheel is produced by the flow of liquid.

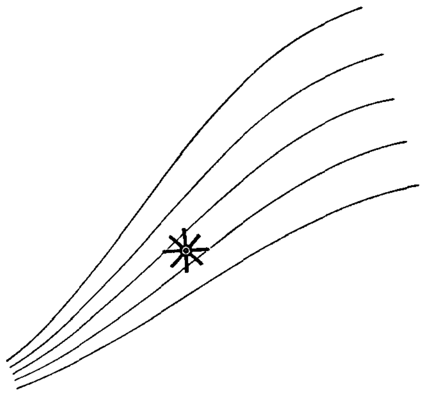


Fig. 9-3 A small paddle wheel placed in a flowing liquid will not rotate when the liquid is in steady or lamellar flow.

9-2 Bernoulli's Theorem

The fundamental theorem regarding the motion of fluids is due to Daniel Bernoulli (1700–1782), a Swiss physicist and mathematician. Bernoulli's theorem is essentially a formulation of the mechanical concept that the work done on a body is equal to the change in its mechanical energy, in the case that mechanical energy is conserved; that is, where there is no loss of mechanical energy due to friction.

Let us consider the motion of an incompressible fluid of density ρ along an imaginary tube bounded by streamlines, as shown in Figure 9-4. We shall call such a tube a *streamtube*. Since each streamline represents the direction of motion of a particle of liquid in steady flow, no particle of liquid may cross a streamtube. At the left-hand end of the tube, the liquid has a velocity v_1, the tube has cross-sectional area A_1, the pressure is P_1, and the tube is at a height h_1 above some reference level. At the right-hand end of the tube, the velocity is v_2, the cross-sectional area is A_2, the pressure is P_2, and the height is h_2. When a small quantity of fluid of volume V is moved into the tube through the action of the external fluid, an equal volume of fluid must emerge from the streamtube against the force exerted by the pressure P_2 of the fluid outside the tube at the right-hand end of the streamtube. Let us imagine that the flow of fluid in the streamtube takes place as the result of the displacement s_1 of a piston of area A_1 which just fits the streamtube at its left-hand end such that the volume swept out by the

motion of the piston is $A_1 s_1 = V$, and of a corresponding displacement s_2 of a piston of area A_2 which just fits the streamtube at its right-hand end such that the volume swept out by the second piston is $A_2 s_2 = V$; that is, a quantity of fluid of volume V has just passed through the streamtube. The piston at position 1 has done work *on* the fluid equal to the product of the force exerted by the displacement of that force, hence the work done on

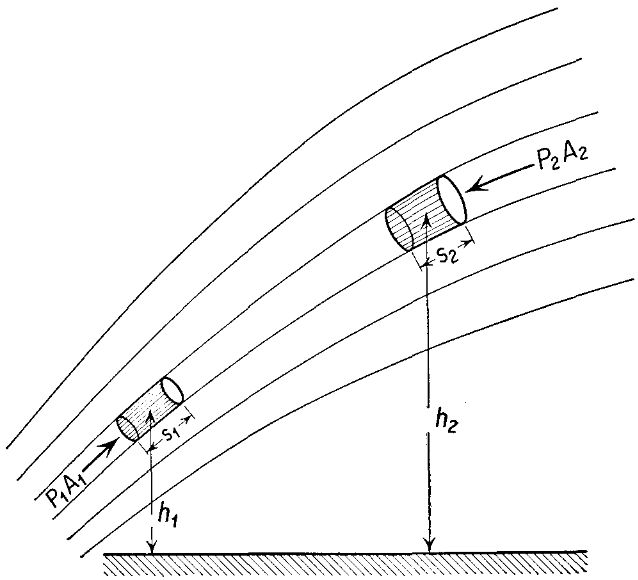


Fig. 9-4

the fluid is $P_1 A_1 s_1$, while the fluid has pushed back the piston at position 2 so that the fluid has done work on the second piston of amount $P_2 A_2 s_2$. The net work done on the fluid is therefore equal to

$$P_1 A_1 s_1 - P_2 A_2 s_2.$$

Since we have assumed the fluid to be incompressible,

we have $$A_1 s_1 = A_2 s_2 = V.$$

From the above equations, we find that the net work done on the fluid within the streamtube is

$$P_1 V - P_2 V.$$

If there was no loss of energy of the fluid due to frictional forces, this work done on the fluid in the streamtube must have resulted in a change of mechanical energy of the liquid which flowed into the tube at position 1 and out of the tube at position 2. The sum of the potential and kinetic

energies of the liquid flowing into the tube at position 1 is $\rho Vgh_1 + \frac{1}{2}\rho Vv_1^2$, while the sum of these energies of the liquid flowing out of the tube at position 2 is $\rho Vgh_2 + \frac{1}{2}\rho Vv_2^2$. Equating the work done on the fluid to the difference in its energy, we find

$$P_1V - P_2V = \rho Vgh_2 + \tfrac{1}{2}\rho Vv_2^2 - (\rho Vgh_1 + \tfrac{1}{2}\rho Vv_1^2). \tag{9-3a}$$

Dividing the equation through by V and transposing all quantities with subscript 1 to the left-hand side and all quantities with subscript 2 to the right-hand side of the equation, we find

$$P_1 + \rho gh_1 + \tfrac{1}{2}\rho v_1^2 = P_2 + \rho gh_2 + \tfrac{1}{2}\rho v_2^2. \tag{9-3b}$$

This equation expresses *Bernoulli's theorem,* which states that *at any two points along a streamline the sum of the pressure, the potential energy of a unit volume of fluid, and the kinetic energy of a unit volume of fluid has the same value.* As indicated in the derivation, Bernoulli's theorem holds rigorously only for frictionless, incompressible, streamline flow. Bernoulli's theorem is a statement of the principle of conservation of energy expressed in a form suited to the description of fluids in steady frictionless flow.

Although Bernoulli's theorem holds rigorously only for an incompressible fluid, experience indicates that it is valid for air in streamline flow at speeds up to about half the speed of sound. (The speed of sound is about 740 mi/hr.) Actual fluids such as water and air have internal fluid friction, or *viscosity,* so that, to be strictly true, the equal sign of Equation (9-3a) should be replaced by a greater than or equal sign ($\geqq$), meaning that the work done on the fluid in the streamtube is greater than, or at least equal to, the increase in mechanical energy. Some of the work done will be converted to heat energy through the action of internal friction. In examples and problems in the application of Bernoulli's theorem, we shall neglect the effects of viscosity.

9-3 Torricelli's Theorem

Let us apply Bernoulli's theorem to the flow of a liquid out of an orifice C at the base of a tank, as shown in Figure 9-5. We shall choose as the reference level for the measurement of potential energy the altitude of the emergent stream at D, where the cross-sectional area of the stream is A, and the velocity of the stream is v. At both positions B at the top of the tank and D at the emergent stream, the liquid is in free contact with the air and is therefore at atmospheric pressure P_{atm}. If the tank is sufficiently large so that the flow does not appreciably change the level of liquid in the tank, we may assume that the liquid at B is very nearly at rest; that is, the velocity of the liquid at B is zero. Applying Bernoulli's theorem in the form of

Equation (9-3b) to the flow of the fluid,

$$P_{\text{atm}} + \rho gh + 0 = P_{\text{atm}} + 0 + \tfrac{1}{2}\rho v^2,$$

from which

$$v^2 = 2gh,$$

or

$$v = \sqrt{2gh}. \tag{9-4}$$

If the orifice at the base of the tank has sharp edges, it may be observed that the stream narrows as it emerges from the tank at C. The portion of the stream with parallel sides is called the *vena contracta.* The narrowing of

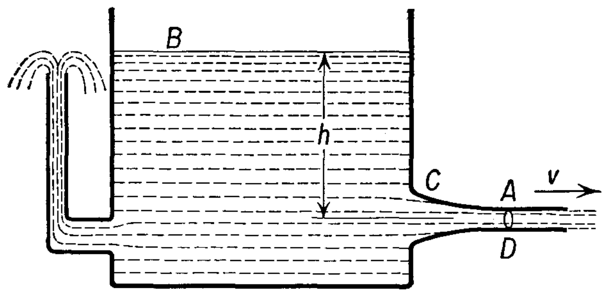


Fig. 9-5

the stream is due to the fact that the liquid is being accelerated and has not yet reached its final velocity. In accordance with Equation (9-2), the cross-sectional area of the stream must be larger where the velocity of flow is smaller.

If we wish to calculate the quantity of liquid flowing out of the tank per second, we may apply Equation (9-1) to the result obtained in Equation (9-4), obtaining

$$Q = Av,$$

so that

$$Q = A\sqrt{2gh}.$$

The speed with which a liquid emerges from a tank is the same as it would have had if it had been dropped over the top of the tank, for, in dropping through a height h, a unit volume of liquid would have lost potential energy ρgh and gained kinetic energy in equal amount. Thus

$$\rho gh = \tfrac{1}{2}\rho v^2,$$

so that

$$v = \sqrt{2gh}.$$

9-4 Venturi Tube

Let us consider a horizontal tube containing a constriction in which a fluid is flowing, as shown in Figure 9-6. We will refer to the wide section of the

tube by the subscript 1 and the narrow portion of the tube by the subscript 2. Applying Equation (9-3b) to the fluid flow, we find

$$P_1 + \rho g h_1 + \tfrac{1}{2}\rho v_1^2 = P_2 + \rho g h_2 + \tfrac{1}{2}\rho v_2^2.$$

Since the tube is horizontal, $h_1 = h_2$, so that

$$P_1 + \tfrac{1}{2}\rho v_1^2 = P_2 + \tfrac{1}{2}\rho v_2^2. \tag{9-5}$$

From Equation (9-1) the velocity of the liquid must be greater in the constricted portion of the tube, so that the pressure in the constriction must be lower than the pressure in the wider portion of the tube.

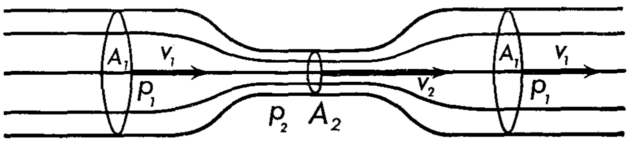


Fig. 9-6 The Venturi tube.

A Venturi tube finds application as a *flowmeter*. Referring to Equations (9-5) and (9-1), we have

$$P_1 + \tfrac{1}{2}\rho v_1^2 = P_2 + \tfrac{1}{2}\rho v_2^2,$$

and

$$A_1 v_1 = A_2 v_2.$$

If the quantity $P_1 - P_2$ is measured by a manometer or other pressure gauge, the two equations may be solved for v_1 and v_2, and this knowledge of the speed of flow determined by a pressure measurement, together with the known areas of the tube and the constriction, suffices to determine the rate of flow of fluid through the tube.

Although we have derived the relationship between pressure and velocity for a Venturi tube, the same relationship is true for any case of streamline flow. Thus, as air is entrained by a passing railroad train, the air close to the train is moving at considerably higher speed than the air some distance away. The pressure in the immediate vicinity of the train is low, and passengers on the station platform are cautioned to stand back from the edge of the platform when a rapidly moving train passes, lest the pressure differential created by the motion of the train force the passenger into the side of the train. Two passing ships or stunting airplanes which get too close together may be forced to collide through the low-pressure area between them which is created by their own motion.

There are many applications of the Bernoulli theorem utilizing variations of the Venturi tube. For example, the aspirator, sketched in Figure 9-7, is used as a vacuum pump. Water from a faucet flows through the horizontal tube and comes out of the constriction with a high velocity, thus

reducing the pressure P in this region below atmospheric pressure. The tube T may be connected to a chamber to be evacuated, and the pressure in it may be reduced to about the vapor pressure of water, usually about 2 or 3 cm of mercury. The same type of design is used for a chemical filter pump. A jet pump used in raising water from a well drives water through a

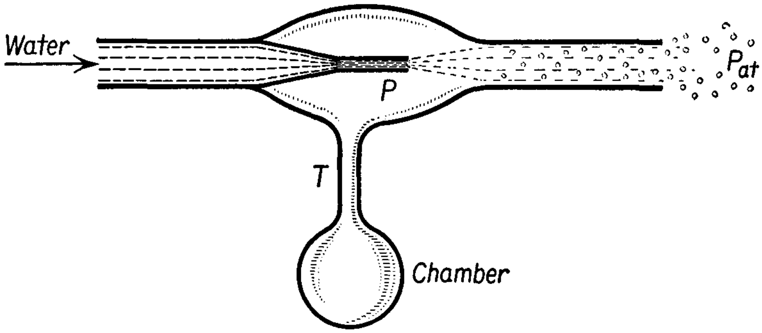


Fig. 9-7 An aspirator.

restricted orifice similar to that in Figure 9-7, and the pressure of the outside air forces the well water into the tube T and into the moving water stream. A similar design of tube, except that air is blown through the horizontal tube and constriction of Figure 9-7, is used in pneumatic conveyers of sand, grain, and other granular materials.

Illustrative Example. Water is flowing through a horizontal Venturi tube at the rate of 100 ft³/min. The pressure in the wide portion of the tube is 15 lb/in.², and its diameter is 6 in. Determine the pressure in the narrow portion of the tube, called the throat, whose diameter is 3 in.

The area of the tube A_1 is

$$A_1 = \frac{\pi}{16}\,\mathrm{ft}^2,$$

while the area of the throat A_2 is

$$A_2 = \frac{\pi}{64}\,\mathrm{ft}^2.$$

From Equation (9-2), we have

$$Q = A_1 v_1 = A_2 v_2.$$

Substituting numerical values, we get

$$\frac{100}{60}\,\frac{\mathrm{ft}^3}{\mathrm{sec}} = \frac{\pi}{16}\,\mathrm{ft}^2 \times v_1 = \frac{\pi}{64}\,\mathrm{ft}^2 \times v_2,$$

from which
$$v_1 = 8.5\,\frac{\mathrm{ft}}{\mathrm{sec}},$$

and $$v_2 = 34\,\frac{\text{ft}}{\text{sec}}.$$

In applying Equation (9-5), as in every other equation, it is essential that a consistent set of units be used. Now

$$P_1 = 15\,\frac{\text{lb}}{\text{in.}^2} = 15 \times 144\,\frac{\text{lb}}{\text{ft}^2} = 2{,}160\,\frac{\text{lb}}{\text{ft}^2},$$

and water, which weighs 62.4 lb/ft^3, has a density of

$$\rho = \frac{62.4}{32}\,\frac{\text{slugs}}{\text{ft}^3}.$$

Since the tube is horizontal, $h_1 = h_2$, and we may write Equation (9-5) as

$$P_1 + \tfrac{1}{2}\rho v_1^2 = P_2 + \tfrac{1}{2}\rho v_2^2,$$

and, substituting numerical values, we get

$$P_2 = 2{,}160\,\frac{\text{lb}}{\text{ft}^2} + \frac{1}{2}\,\frac{62.4}{32} \times (8.5)^2\,\frac{\text{ft lb}}{\text{ft}^3} - \frac{1}{2}\,\frac{62.4}{32} \times (34)^2\,\frac{\text{ft lb}}{\text{ft}^3}$$

$$= 1{,}100\,\frac{\text{lb}}{\text{ft}^2},$$

or $$P_2 = \frac{1{,}100}{144}\,\frac{\text{lb}}{\text{in.}^2} = 7.64\,\frac{\text{lb}}{\text{in.}^2}.$$

9-5 Lift of an Airfoil

The lift of an airplane wing or the action of a propeller of an airplane or a ship may be analyzed in terms of Bernoulli's theorem. The cross section of such an airfoil generally has a rather blunt nose and thin trailing edge to permit the fluid medium to flow around it in streamline flow. Ideally, an element of fluid volume parted by the airfoil will be reconstructed when the fluid passes the trailing edge of the airfoil, as shown in Figure 9-8. This

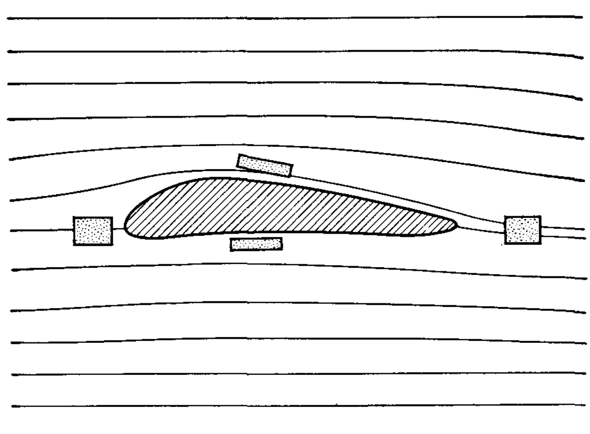


Fig. 9-8

implies that the fluid on top of the airfoil is moving more rapidly than the fluid at the bottom of the airfoil, so that the pressure at the top surface is lower than the pressure at the bottom surface, creating an upward force on the airfoil. It is for this reason that the greatest lift is often associated with an airfoil of rather marked asymmetry. A perfectly symmetrical wing had no lift unless it is turned at an angle with respect to the flow of fluid, termed the *angle of attack*, so that the fluid moving over the top surface travels a greater distance than the fluid moving over the bottom surface of the airfoil. The motion of air past a symmetrical airfoil for several angles of attack is shown in Figure 9-9.

Surfaces not normally considered to be airfoils may also experience lift. For example, the lift on a bridge in a high wind may be several times the drag force tending to blow the bridge off its piers.

Fig. 9-9 Smoke is used in the NACA smoke tunnel at Langley Field, Virginia, to make the flow of air visible, as illustrated in these photographs. Note the smoothness of the air flow in the lowest picture. When the angle of attack has been increased to 10°, the air flow begins to separate from the upper surface of the airfoil (center view); and when the angle is increased to 30°, the flow separates completely from the upper surface. Turbulence behind the trailing edge of the airfoil may be observed in this picture. (Reproduced from the *Journal of Applied Physics*, August, 1943, with permission of the National Advisory Committee for Aeronautics.)

9-6 The Siphon

A siphon is a bent tube used for transferring a liquid from one vessel to another one at a lower level. If the bent tube shown in Figure 9-10 is filled with liquid so that streamline flow can take place, liquid will emerge from the orifice with a speed given by Torricelli's theorem. By applying Bernoulli's theorem, we can determine the maximum height h of the bend

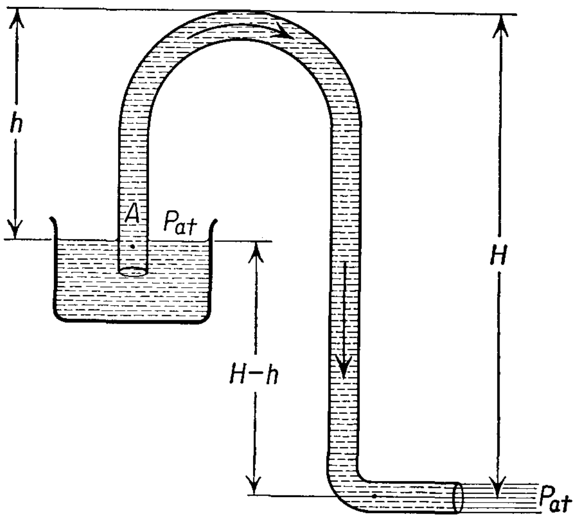


Fig. 9-10 Operation of a siphon.

in the tube above the level of the free liquid in the higher vessel, for a liquid cannot support a negative pressure, that is, a state of internal tension, except for extremely short time intervals. If the pressure of the liquid at its highest point is zero (or more accurately, if it is below the vapor pressure of the liquid), the liquid will pull apart, and bubbles will form, destroying the continuity of the liquid flow, and thus interrupting the siphon. The limiting height of the siphon may be found by setting P_2 equal to zero in Bernoulli's equation. Thus

$$P_{\text{atm}} = 0 + \rho g h + \tfrac{1}{2}\rho v^2,$$

from which

$$h = \frac{P_{\text{atm}} - \frac{1}{2}\rho v^2}{g}.$$

Thus the faster the fluid flows in the siphon, the lower the bend in the tube must be to maintain siphon flow.

9-7 Fluid Friction; Viscosity

When a fluid, either a liquid or a gas, is set in motion, different parts of the fluid move with different velocities. For example, if a jar of water is tilted so that the water starts flowing out, the top layer of the water moves over the lower part of the water. Just as there is friction when one surface of a solid slides over another, so there is friction when one layer of a fluid slides over another. This friction in fluids is called *viscosity*. When a fluid flows through a cylindrical pipe, the part of the fluid in contact with the pipe adheres to it and remains at rest. We may think of the rest of the fluid as divided into concentric cylindrical layers, the velocity of each succeeding inner layer increasing as we go to the center. A difference in pressure between the two ends of the pipe is needed to maintain a steady flow through it and oppose the force due to the viscosity of the fluid.

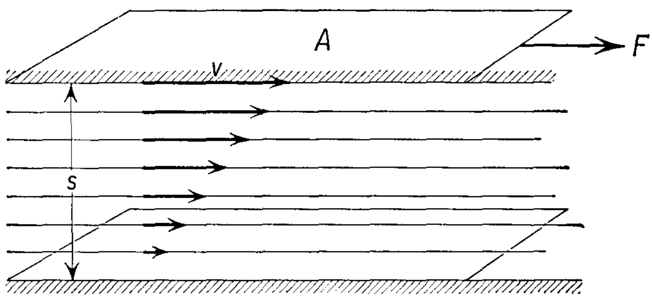


Fig. 9-11

The resistance experienced by a solid moving through a fluid is due essentially to the viscosity of the fluid. A certain amount of fluid adheres to the surface of the solid and moves with it, and this layer drags along an adjacent layer and so on, until, at sufficiently large distance from the solid, the fluid is at rest. Some of the momentum of the solid has been given up to setting the fluid in motion, and if no external force is applied, the solid will come to rest. An additional cause of the resistance experienced by objects moving through fluids is the turbulence set up in the fluid. When turbulence occurs, the fluid flow is no longer streamline and cannot be considered as taking place in layers, for the layers are sometimes broken up into eddies in which rotational motion takes place, and sometimes into waves. Such eddies are seen in the wake of the oars of a rowboat and occur in the air about the wingtips or the propeller tips of an airplane. If the paddle wheel of Figure 9-3 is placed at the center of an eddy, the paddle wheel will turn.

To formalize the concept of viscosity, we imagine a layer of fluid

between two flat plates, one of which is at rest and the other moving at velocity v, as shown in Figure 9-11. The fluid in contact with the stationary plate is at rest, while the fluid in contact with the moving plate is moving with velocity v, and the fluid in the space between the two plates is moving with intermediate velocities. To maintain the moving plate at a constant speed, it is found experimentally that a force F is required which is directly proportional to the velocity, inversely proportional to the separation of the two plates s, and directly proportional to the area of the moving plate A. Calling the constant of proportionality η (eta), the *coefficient of viscosity,* we find

$$F = \eta \frac{Av}{s} \cdot \tag{9-6}$$

From Equation (9-6) the unit of coefficient of viscosity is force times distance divided by area times velocity. By replacing the unit of force by its equivalent from Newton's second law, we see that the unit of viscosity is that of mass divided by the product of the unit of length by the unit of time. Thus in mks units, the coefficient of viscosity is expressed in kilograms per meter seconds, while in British gravitational units it is expressed in slugs per foot seconds. In the cgs system of units a coefficient of viscosity of 1 gm/cm-sec is called a *poise.*

The coefficient of viscosity varies considerably with temperature. For liquids the viscosity generally decreases with increasing temperature, while the viscosity of gases increases with increasing temperature. The viscosity of gases does not depend upon the pressure.

To achieve an experimental condition in which the coefficient of viscosity can be measured, the relative motion of two plane surfaces is approximated by two concentric cylinders, one of which is held fixed and the other is caused to rotate by the application of a constant torque. From a measurement of the angular speed, the dimensions of the apparatus, and the applied torque, the viscosity of the liquid between the two cylinders can be determined.

9-8 Reynolds' Number

To discuss problems in fluid flow, it is convenient to group the variables determining the motion into dimensionless parameters. One such parameter, called *Reynolds' number,* is especially useful in distinguishing between streamline or lamellar flow and turbulent flow. Reynolds' number R is given by the formula

$$R = \frac{\rho v d}{\eta},$$

where ρ is the density of the fluid, v is its speed relative to a pipe or to some obstacle to the flow, d is some dimension characteristic of the flow, and η is the coefficient of viscosity. In the flow of a fluid through a pipe, d

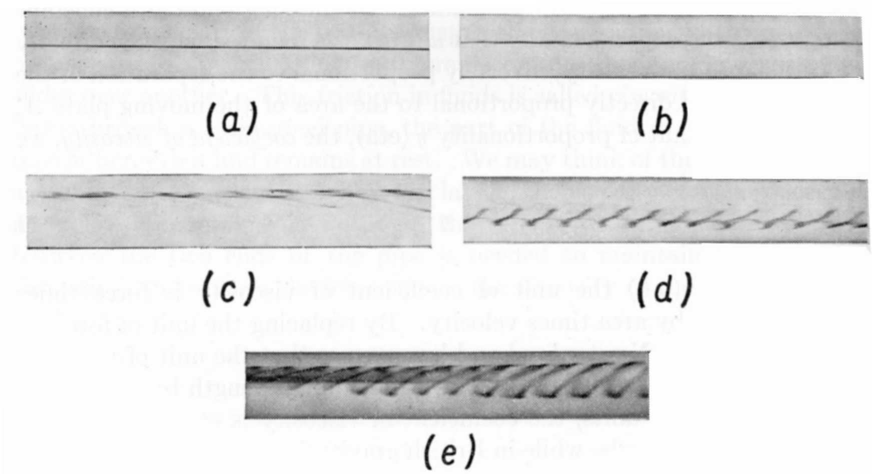


Fig. 9-12 Transition from lamellar to turbulent flow for water in a glass tube, with increasing value of Reynolds' number. Velocity lowest for (a), which shows lamellar flow, and highest for (e), which shows turbulent flow. (Reproduced by permission from *Fluid Mechanics*, 3rd ed., by R. C. Binder, p. 108. Copyright, 1943, 1949, 1955, by Prentice-Hall, Inc., Englewood Cliffs, N.J.)

represents the diameter of the pipe, while in the flow of a fluid past a cylinder, d is the diameter of the cylinder.

Experiments have shown that the stable flow of a fluid through a pipe is normally streamline or lamellar for values of R less than 2,000, whatever the density or viscosity of the fluid. The flow is generally turbulent if R is greater than 3,000, and the type of flow obtained in the transition region between these two values of R depends on such factors as surface roughness, the way the flow was started, and so on. The flow of water through a glass tube at increasing speeds, that is, at increasing values of R, is shown in Figure 9-12. The flow has been made visible by immersing a source of dye in the water stream.

The resisting force opposing the motion of an immersed body, such as an airfoil, a sphere, a cylinder, or a bridge, is also related to Reynolds' number. For example, eddies break off alternately on either side of a cylinder in a periodic manner, as shown in Figure 9-13, when R is greater than 20. Below this value of R, the flow is lamellar. The release of eddies from alternate sides of a cylinder exerts a periodic force on the cylinder. This is responsible for the generation of the vibrations of chimneys in a

high wind, the vibrations of the periscopes of submarines, the vibrations of towlines, the singing of telephone wires, and so on. Similar phenomena are responsible for the flutter of the wings and propellers of an airplane, and

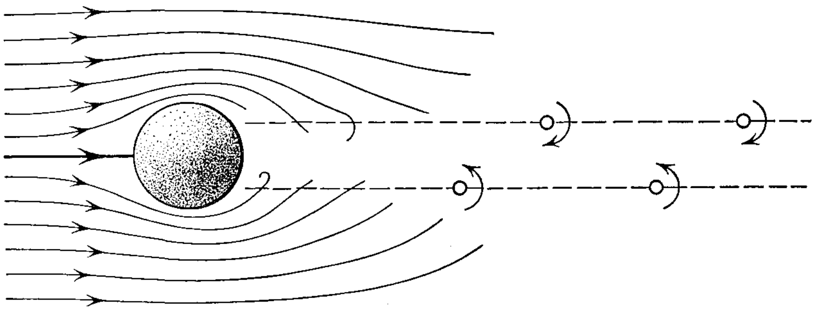


Fig. 9-13

for the failure of the Tacoma Narrows bridge in November, 1940. When the frequency of eddy formation approaches the natural frequency of vibration of the structure, excessively large vibrations may be excited, with resulting destructive failure.

Problems

9-1. Water flows through a horizontal pipe of varying cross section at the rate of 4 ft^3/min. Determine the velocity of the water at a point where the diameter of the pipe is (a) 1.5 in. and (b) 2 in.

9-2. Oil flows through a 12-in.-diameter pipeline with a speed of 3 mi/hr. How many gallons of oil are delivered each day by this pipeline? (One gallon = 231 $in.^3$.)

9-3. At a place in a pipeline where the diameter is 6 in., the speed of a steady stream of water is 12 ft/sec. (a) What will be the speed of the water in that portion of the pipeline where the diameter is 4 in.? (b) At what rate, in cubic feet per minute, is water being delivered by this pipeline?

9-4. A water storage tank is filled to a height of 16 ft. (a) With what speed will water come out of a valve at the bottom of the tank if friction is negligible? (b) To what height will this water rise if the opening is directed upward? (c) What quantity of water will emerge from the tank in each second? The area of the valve is $\frac{1}{3}$ $in.^2$.

9-5. A cylindrical water storage tank of diameter 10 ft is filled to a height of 16 ft. At the bottom of the tank, there is an opening 1 in. in diameter. How long will it take for the tank to drain itself empty? [Hint: Use the methods of the calculus].

9-6. Water flows steadily through a Venturi tube at the rate of 40 ft^3/min. At a place where the diameter of the tube is 4 in., the gauge pressure is 15 lb/$in.^2$. Determine the gauge pressure in the throat of the tube where the diameter is 2 in.

9-7. Gauges attached to a vertical tube in which water is flowing steadily show an absolute pressure of 25 $lb/in.^2$ where the diameter of the tube is 4 in. and a pressure of 15 $lb/in.^2$ where the diameter of the tube is 3 in., at a point 1 ft below the first gauge. Determine (a) the velocity of the liquid in the wider portion of the tube and (b) the quantity of water per second flowing through the tube.

9-8. Oil of density 0.9 gm/cm^3 flows through a horizontal tube 3 cm in diameter at a pressure of 1.5×10^6 $dynes/cm^2$. At one portion the tube narrows down to 2 cm in diameter, and the pressure drops to 10^6 $dynes/cm^2$. (a) Determine the velocity of the oil in the wider portion of the tube. (b) Determine the rate at which oil flows through this tube.

9-9. In a wind-tunnel experiment the pressure on the upper surface of a wing was 13.05 $lb/in.^2$, while the pressure on the lower surface was 13.15 $lb/in.^2$. Determine the lifting force of a wing of this design if it has a spread of 40 ft and a width of 9 ft.

9-10. A monoplane weighing 14,000 lb has a wing area of 600 ft^2. (a) What difference in pressure on the two sides of the wing surface is required to maintain this plane in level flight? (b) If the plane is flying at a level of 13,000 ft and the pressure on the lower wing surface is 9.0 $lb/in.^2$, determine the pressure on the upper wing surface.

9-11. The lower end of a siphon is 8 ft below the level of the water surface in the tank. (a) Determine the speed with which the water flows out of the open end of the siphon. (b) If the cross-sectional area of the siphon tube is 3.0 $in.^2$, determine the rate at which water is siphoned out. (c) If the bend in the siphon is 3 ft above the surface of the water in the tank, determine the pressure of the water in the bend of the siphon.

9-12. The pressure in the cylinder of a water pump is 45 $lb/in.^2$. Determine the height to which water may be lifted by this pump.

9-13. The level of water in a tank is 15 m above the ground. Water flows out of this tank in a horizontal direction through a valve located 5 m below the surface. Determine (a) the velocity with which the water escapes, neglecting friction, (b) the distance from the valve where the water strikes the ground, and (c) the velocity of the water when it reaches the ground.

9-14. Water falls from a height of 60 ft and drives a water turbine. If the rate of flow of water is 480 ft^3/min, determine the maximum power that can be developed by this turbine.

9-15. In a viscosimeter constructed of two concentric cylinders with an annular space between them, the outer cylinder is fixed, and the inner cylinder is made to rotate by the application of a torque. The length of the cylinders is 5 in., their mean radius is 3 in., and the space between them is 0.05 in. It is found that the inner cylinder will rotate with an angular speed of 50 rpm when a torque of 0.10 ft lb is applied. Using British gravitational units, determine the viscosity of the oil in the annular space between the two cylinders.

9-16. Find the conversion factor by which the poise may be converted from cgs to British gravitational units of viscosity.

9-17. Sea water, which weighs 64 lb/ft^3, is stored in an open tank. The water is piped to ground level by a vertical pipe of diameter 1 in., and emerges through a bend in the pipe in a stream of 1 in. diameter. The surface of the water in the

tank is 100 ft above the opening in the pipe. (a) Find the quantity of water which flows from the tank per second. (b) Find the pressure of the water in the pipe 10 ft above the ground.

9-18. An open tank is filled with sea water, which weighs 64 lb/ft^3. The water flows through a siphon tube of cross-sectional area 0.02 ft^2 which has a nozzle of area 0.01 ft^2 at its end. The open end of the nozzle is 5 ft below the free surface of the water in the tank. (a) What is the speed with which water flows from the nozzle? (b) What is the volume of flow from the nozzle, in ft^3/sec? (c) What is the speed of flow in the siphon tube at a point 3 ft above the nozzle? (d) What is the gauge pressure in the siphon tube at a point 3 ft above the nozzle?

10

Momentum and Impulse

10-1 Momentum

An extremely important concept in the development of mechanics is that of *momentum. The momentum of a body is defined as the product of its mass by its velocity.* We shall use the symbol $\mathbf{p}$ to denote the momentum of a body. The momentum of a body is a vector quantity, for it is the product of mass, a scalar, by velocity, a vector. While momentum and kinetic energy are compounded of the same two ingredients, mass and velocity, they are quite different concepts, and one aspect of their difference may be seen in the fact that momentum is a vector while energy is a scalar quantity.

Newton himself recognized the importance of momentum as a mechanical concept, for a free translation of his second law of motion expressed in modern terms would read: *the rate of change of momentum is proportional to the net force and is in the direction of that force.* Expressed in the form of an equation, Newton's second law would read

$$\mathbf{F} = \frac{d(m\mathbf{v})}{dt} = \frac{d\mathbf{p}}{dt}. \tag{10-1}$$

In our discussion of Newton's second law in Chapter 5, we treated mass as a constant and obtained the result

$$\mathbf{F} = m\frac{d\mathbf{v}}{dt} = m\mathbf{a}.$$

This form of Newton's second law is true for most problems in mechanics, when the speed of the body is small in comparison with the speed of light. Newton's original formulation, as represented in Equation (10-1), remains correct even for bodies which travel at speeds approaching the speed of light, when, according to Einstein's special theory of relativity, the mass of a body may be expressed as

$$m = \frac{m_0}{\sqrt{1 - v^2/c^2}} \tag{10-2}$$

where m_0 is the mass of the body at rest, v is its speed, c is the speed of light (3×10^{10} cm/sec), and m is the mass of the moving body. For our present purposes we shall not digress further into the relativistic aspects of mechanics, but shall focus our attention upon Equation (10-1).

From Equation (10-1) we see that if an unbalanced force is applied to a body, its momentum will change at a rate determined by the force. If the unbalanced force acting on a body is zero, the change of momentum is zero; that is, the momentum of the body remains constant. Thus, in the design of space ships for interplanetary flight, consideration must be given to the provision of fuel to accelerate the ship and to decelerate it for landing at the destination, but no fuel need be provided for propelling the ship over the major portion of its path, for, acted on by no appreciable external forces, the momentum of the ship will remain substantially constant.

Our study of mechanics has thus far concentrated its attention on the motion of a particle when acted upon by a force. Suppose we consider a stream of particles of mass m, each moving with velocity v, that strike a target and come to rest in it, and inquire about the average force exerted on the target to hold it in place. From another point of view we may ask what force the particles exert on the target. Each time a particle is stopped by the target, the momentum of the particle is changed from mv to zero. The change in momentum of the particle is mv. If n particles strike the target in each second, the average rate of change of momentum, that is, the change in momentum per second, is

$$\frac{\Delta p}{\Delta t} = nmv.$$

Rewriting Equation (10-1) in incremental form, we see that the average rate of change of momentum is the average force. Thus we have

$$\bar{F} = \frac{\Delta p}{\Delta t}, \tag{10-3}$$

so that

$$\bar{F} = nmv. \tag{10-4}$$

Illustrative Example. A pitcher throws baseballs at a target mounted on a helical spring at the rate of one ball every 2 sec. The baseballs strike the target at a speed of 80 ft/sec and come to rest in it. What is the average force exerted by the baseballs against the target? A baseball weighs 5 oz.

When the baseball collides with the target, all of the momentum of the ball is absorbed by the target. Each ball has a momentum of

$$p = mv = \frac{5}{16} \times \frac{1}{32}\,\text{slug} \times 80\,\frac{\text{ft}}{\text{sec}}$$

$$= 0.78\,\frac{\text{slug ft}}{\text{sec}}\cdot$$

Since one ball strikes the target every 2 sec, the number of balls striking the target per second is $n = \frac{1}{2}$. Substituting in Equation (10-4), we have

$$\bar{F} = nmv$$

$$= \tfrac{1}{2} \times 0.78 \,\frac{\text{slug ft}}{\text{sec}^2},$$

or

$$\bar{F} = 0.39 \text{ lb}.$$

Illustrative Example. A stream of water 10 cm² in area, moving horizontally with a speed of 25 m/sec, strikes the wall of a house and splatters to the ground, losing all of its forward motion. What is the force exerted on the wall of the house by the stream of water?

If Q is the volume of water that strikes the wall of the house per second, then

$$Q = Av,$$

where A is the cross-sectional area of the stream, and v is its velocity. If the water is of density ρ, the mass of water striking the house in each second is

$$m = Q\rho.$$

The water has momentum $mv = Q\rho v$ in the horizontal direction before striking the wall, and zero momentum in this direction afterward, so that the change of momentum in each second is

$$\frac{\Delta p}{\Delta t} = Q\rho v = Av^2\rho.$$

From Equation (10-3) we have, for the force in the horizontal direction,

$$\bar{F} = \frac{\Delta p}{\Delta t} = Av^2\rho. \tag{10-5}$$

Substituting numbers into Equation (10-5), we find

$$\bar{F} = 10 \text{ cm}^2 \times (2{,}500)^2 \,\frac{\text{cm}^2}{\text{sec}^2} \times 1 \,\frac{\text{gm}}{\text{cm}^3}$$

$$= 625 \times 10^5 \,\frac{\text{gm cm}}{\text{sec}^2}$$

$$= 6.25 \times 10^7 \text{ dyne}.$$

10-2 Impulse

In many mechanical problems the applied force is not steady, nor can the force be described in terms of simple mathematical functions. When a baseball bat strikes a ball, the force the bat exerts against the ball is zero at the initial instant of collision, then rises to some maximum value when the ball is violently deformed, and finally returns to zero when the ball leaves the bat. The behavior of materials and structures under such *impulsive*

forces is quite different from their behavior when subjected to steady forces, and we speak of materials as being brittle when they are not capable of withstanding impulsive loading. An impulsive force, such as that exerted by a baseball against a bat, might be described by the graph in Figure 10-1(a).

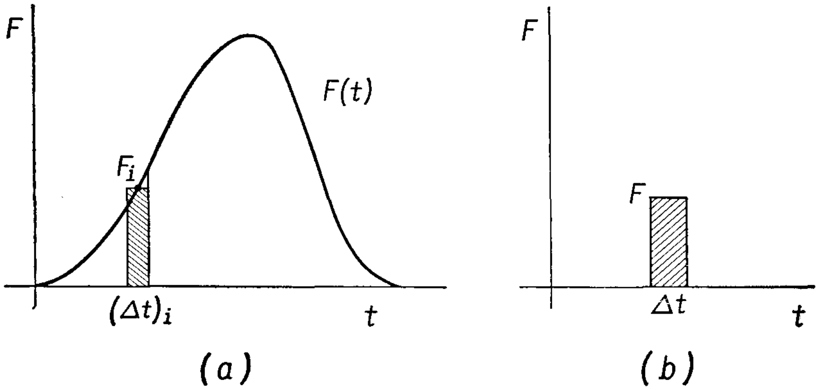


Fig. 10-1 Impulsive force.

In our earlier discussion of work and energy, we have seen the usefulness of considering the effect of a force acting through a distance. From such an analysis we derived our understanding of the concept of *work*, and we saw that the result of doing work on a particle was to change its *energy*. Another way to consider the effect of a force on a body is to study the effects produced when a force acts for a *time* interval. The product of a force by the time interval during which the force acts is called the *impulse*. When a force **F** acts for a time interval Δt, the impulse $\Delta \mathbf{J}$ is given by the formula

$$\Delta \mathbf{J} = \mathbf{F}\,\Delta t. \tag{10-6}$$

We see that *impulse is a vector quantity*, for it is given by the product of force, a vector, by time, a scalar. We shall show that the effect of an impulse acting upon a particle is to produce a change in its *momentum*.

Let us consider the incremental form of Newton's second law, as given by Equation (10-3). We have

$$\bar{F} = \frac{\Delta p}{\Delta t}.$$

If we multiply both sides of this equation by the time interval Δt and call the force acting in this time interval F, we find

$$F\,\Delta t = \Delta p,$$

and, substituting from Equation (10-6), we have

$$\Delta J = \Delta p, \tag{10-7a}$$

which relates the magnitude of the impulse to the magnitude of the change in momentum. In vector form this equation becomes

$$\boxed{\Delta \mathbf{J} = \Delta \mathbf{p}.} \tag{10-7b}$$

Illustrative Example. A body of mass 10 gm moves along the x axis with a speed of 3 cm/sec. A force of 400 dynes is applied in the positive y direction for a time interval of 0.1 sec. Find the velocity of the particle produced by the impulse.

Since the impulse is applied in the y direction, there is no change in the x momentum of the particle. The impulse in the y direction is

$$\begin{aligned} \Delta J_y &= 400 \text{ dynes} \times 0.1 \text{ sec} \\ &= 40 \text{ dyne sec.} \end{aligned}$$

From Equation (10-7) we have

$$\Delta p_y = 40 \text{ dyne sec.}$$

Since the momentum in the y direction was initially zero, the final momentum in the y direction is equal to 40 dyne sec, and we have

$$\begin{aligned} \Delta p_y &= mv_y = 40 \text{ dyne sec,} \\ v_y &= \frac{40 \text{ dyne sec}}{10 \text{ gm}} \\ &= 4 \, \frac{\text{cm}}{\text{sec}} \cdot \end{aligned}$$

Thus the body has a velocity of 4 cm/sec in the y direction and a velocity of 3 cm/sec in the x direction. Its resultant velocity is therefore 5 cm/sec, directed at an angle of 53° with the x axis into the first quadrant.

When the impulsive force is given by a simple rectangular pulse, as in Figure 10-1(b), the evaluation of the impulse is simple and straightforward. When the impulsive force is given as an arbitrary function of time, as in Figure 10-1(a), we may follow the procedures of the integral calculus and imagine the graph of $F(t)$ to be divided into a number of rectangular pulses of different heights, in which each impulse serves to change the momentum by a small amount. The total impulse is the area under the curve. To find the over-all effect we add the changes in momentum due to each impulse and write

$$J = \int_0^J dJ = \int_0^t F\, dt = \int_{p_i}^{p_f} dp,$$

$$J = p_f - p_i. \tag{10-8a}$$

Thus, *when an arbitrary impulsive force strikes a body, the impulse is equal to the difference between the final momentum* p_f *and the initial momentum* p_i. Although Equation (10-8a) was derived in scalar form, it is clear that we may consider Equation (10-8a) as an equation in one unspecified component of the more general vector equation

$$\mathbf{J} = \mathbf{p}_f - \mathbf{p}_i. \tag{10-8b}$$

In general, it is quite difficult to measure an impulse, but it is easy to observe a change in momentum. Thus the difference in momentum between a pitched baseball and a batted ball may be used to measure the impulse of the bat against the ball.

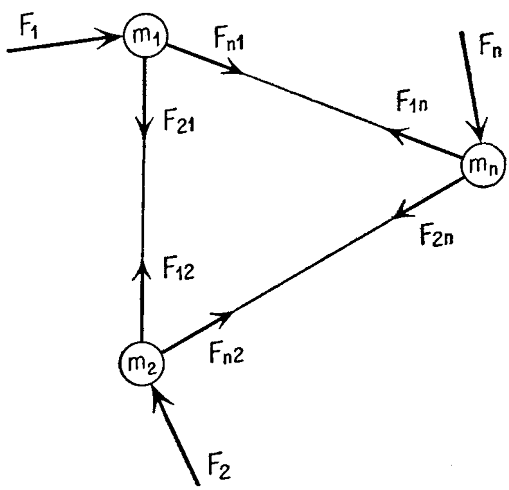


Fig. 10-2 System of particles.

10-3 Systems of Particles

Let us consider a system of particles such as the particles $m_1, m_2, \ldots m_n$ of Figure 10-2, and let us suppose that the forces the particles of the system exert on one another are always directed along the line joining them, in the manner of gravitational forces, or of the forces which might be exerted by strings. We call the forces which one particle of the system exerts on another particle of the system *internal forces*. The forces exerted on the particles of the system from outside the collection of particles are called *external forces*.

The division of the forces on the particles of a collection into internal forces and external forces is due to the French physicist Jean le Rond d'Alembert (1717–1783). Let us call $\mathbf{F}_{ab}$ the force exerted *by particle* a *on particle* b. By Newton's third law the force exerted by a on b must be equal in magnitude and opposite in direction to the force exerted by b on

a. Forces with *double subscripts* will always denote *internal* forces, exerted by one member of the system on another member of the system, while forces with *single subscripts* will always denote the *external* force on the particle indicated by the subscript.

Applying Newton's second law to each of the particles of the collection in turn, we find

$$\mathbf{F}_1 + \mathbf{F}_{21} + \cdots + \mathbf{F}_{n1} = \frac{d\mathbf{p}_1}{dt},$$

$$\mathbf{F}_2 + \mathbf{F}_{12} + \cdots + \mathbf{F}_{n2} = \frac{d\mathbf{p}_2}{dt},$$

$$\cdots$$

$$\mathbf{F}_n + \mathbf{F}_{1n} + \mathbf{F}_{2n} + \cdots + \mathbf{F}_{(n-1)n} = \frac{d\mathbf{p}_n}{dt}.$$

Adding these equations, and remembering that $\mathbf{F}_{12} = -\mathbf{F}_{21}$, and so on, we find that

$$\mathbf{F}_1 + \mathbf{F}_2 + \cdots + \mathbf{F}_n = \frac{d}{dt}(\mathbf{p}_1 + \mathbf{p}_2 + \cdots + \mathbf{p}_n).$$

Thus *the vector sum of all the external forces acting on the system of particles is equal to the rate of change of the total momentum of the system of particles.* If $\mathbf{p}$ is the sum of the momenta $\mathbf{p}_1 + \mathbf{p}_2 + \mathbf{p}_3 + \cdots + \mathbf{p}_n$, and $\mathbf{F}$ is the sum of the external forces acting on each of these particles $\mathbf{F}_1 + \mathbf{F}_2 + \mathbf{F}_3 + \cdots + \mathbf{F}_n$, we may write

$$\boxed{\mathbf{F} = \frac{d\mathbf{p}}{dt}.} \tag{10-9}$$

In the special case in which $\mathbf{F} = 0$, Equation (10-9) shows that $\mathbf{p}$ is constant.

Thus *if a system of particles is acted on by no external force, the total momentum of the system of particles is constant.* This statement is known as the *principle of conservation of momentum;* it is of very great importance in the analysis of mechanical systems. The principle of conservation of momentum, like the principle of conservation of energy, is valid in all realms of physics, from subatomic to astronomical.

A theorem which is extremely useful in understanding the behavior of many mechanical systems, and one which is very easily proved, states that *the momentum of a system of particles is equal to the product of the mass of the entire system by the velocity of its center of mass.* To prove the theorem

we recall that the coordinates of the center of gravity of a system of particles are given by Equations (4-4) as

$$x_0 = \frac{w_1x_1 + w_2x_2 + \cdots + w_nx_n}{w_1 + w_2 + \cdots + w_n}, \tag{4-4a}$$

and similarly for the y coordinate of the center of gravity y_0 and the z coordinate z_0. If we substitute $w = mg$ into the above equation and clear it of fractions, we find

$$(m_1 + m_2 + \cdots + m_n)x_0 = m_1x_1 + m_2x_2 + \cdots + m_nx_n.$$

Thus x_0 is the *center of mass* of the system of particles; it coincides with the center of gravity in any region where g is constant. Let us differentiate both sides of the equation with respect to time to obtain

$$(m_1 + m_2 + \cdots + m_n)\frac{dx_0}{dt} = m_1\frac{dx_1}{dt} + m_2\frac{dx_2}{dt} + \cdots + m_n\frac{dx_n}{dt},$$

which proves the theorem. The left-hand side of the equation is the mass of the system of particles times the x component of the velocity of the center of mass, while the right-hand side of the equation is the sum of the x components of the momenta of the particles of the system. Since analogous equations may be obtained immediately for the other two components of the momentum, we may combine these results into a single vector equation:

$$M\mathbf{v}_0 = \mathbf{p}_1 + \mathbf{p}_2 + \cdots + \mathbf{p}_n, \tag{10-10}$$

where M is the total mass of the system of particles and $\mathbf{v}_0$ is the velocity of its center of mass.

Let us examine the implications of Equation (10-10). If the sum of the external forces acting on a system of particles is zero, we see from Equation (10-9) that the momentum of the system must remain constant. From Equation (10-10), therefore, the velocity of the center of mass of the system must remain constant. If a shrapnel shell explodes in mid-air, the trajectory followed by the center of mass will be the same as it would have been if the shell had not burst, and the total momentum which the shrapnel fragments acquire in a direction normal to the path of the unexploded shell must be zero, as it was before the explosion, for the forces involved in the explosion are all internal forces. When a gun fires a shell, we observe that the center of mass of the system consisting of gun, shell, and explosive gases was at rest before the firing; hence the center of mass of this system must be at rest after the firing. If the projectile has acquired forward momentum, the gun must acquire an equal backward momentum; it must recoil.

10-4 Conservation of Momentum

The impact between two isolated bodies in space may be most easily understood in terms of the principle of conservation of momentum. In addition, many problems of propulsion may be most easily understood in terms of momentum conservation.

Consider the problem of an airplane moving through the air. We may think of the system consisting of the airplane and a volume of air around it as constituting a region isolated in space and acted on by no external forces. If the airplane's velocity in the forward direction is to be changed, its momentum in the forward direction must be changed. But the system, having no external forces acting on it, is not permitted to change its total momentum. If the momentum of the airplane is increased in the forward direction, the air must acquire an equal and opposite momentum in the backward direction. The function of the propellers is not to "screw the airplane through the air" but rather to deliver this backward component of momentum to the air. The jet engine of an airplane takes in a quantity of air, and, as a result of combustion processes, that air is expelled to the rear with greater velocity, in the form of exhaust gases. In this sense there is only a superficial mechanical difference between a propeller-driven airplane and a jet airplane, for both the propeller and the jet serve the same function—to give the air a backward component of momentum.

In exactly the same way the oars of a rowboat or the propellers of an ocean-going vessel propel the craft by delivering a backward momentum component to the water. The propellers of a ship are considerably smaller, in comparison to its mass, than the propellers of an airplane, because a smaller volume of water can carry off the required momentum, the water having far greater mass per unit of volume.

The problem of propelling a rocket is similar to that of propelling a jet airplane, but there is the additional complication that the rocket must carry along all the mass which must be ejected as momentum, while the airplane has a readily available supply of mass, in the form of the surrounding air, which it acquires at one speed and expels with greater speed as needed.

10-5 Elastic and Inelastic Impact

While in every collision between bodies the total momentum remains the same before and after the impact when the system is acted upon by no external forces, the mechanical energy of the system does not necessarily remain constant. Collisions in which the total kinetic energy remains constant are called *elastic*, while collisions in which the total kinetic energy is less after the collision than it was before the collision are called *inelastic*.

In an inelastic collision the kinetic energy lost in the impact is transformed to sound energy, to heat energy, and to the energy required to deform or fracture a body.

Consider the central collision between two spheres of masses M and m, moving with velocity $\mathbf{U}$ and $\mathbf{u}$ before collision and with velocity $\mathbf{V}$ and $\mathbf{v}$

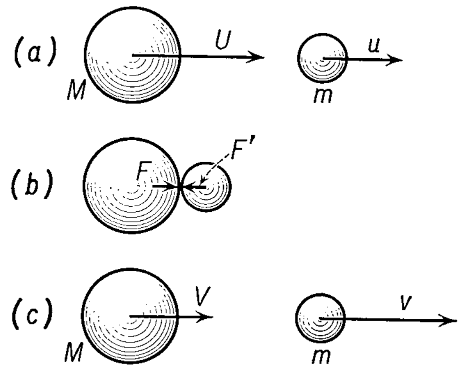


Fig. 10-3 Central collision of two spheres: (a) velocities before collision; (b) forces acting during collision, $F = -F'$; (c) velocities after collision.

after collision, respectively, as illustrated in Figure 10-3. Since the system, composed of the two spheres, is acted upon by no external forces, the total momentum of the system before collision will be equal to the total momentum of the system after collision. If, in addition, the collision is elastic, the total kinetic energy before collision will be equal to the total kinetic energy after collision. Writing the energy equation first, we have

$$\tfrac{1}{2}MU^2 + \tfrac{1}{2}mu^2 = \tfrac{1}{2}MV^2 + \tfrac{1}{2}mv^2.$$

The equation representing conservation of momentum is

$$MU + mu = MV + mv.$$

Transposing quantities for the sphere M to the left-hand side and quantities for the sphere m to the right-hand side of both equations, and simplifying, we get

$$M(U^2 - V^2) = m(v^2 - u^2),$$

and,

$$M(U - V) = m(v - u);$$

and, dividing the first equation by the second, we have

$$U + V = v + u,$$

or

$$U - u = v - V; \tag{10-11}$$

that is, the velocity with which the two spheres approach each other is numerically equal to the velocity with which they leave each other.

In general, one cannot say much about an arbitrary inelastic collision.

If the collision is completely *inelastic,* a solution may be obtained by noting that the two bodies stick together after the collision; that is, they have a common final velocity.

In some cases it is possible to characterize inelastic collisions by making use of a quantity called the *coefficient of restitution e,* defined by the equation

$$e = \frac{v - V}{U - u}. \tag{10-12}$$

The coefficient of restitution is a positive number whose value is given by the ratio of the velocity at which the two particles leave each other in a central collision to the velocity with which the two particles approach each other before the collision. From Equation (10-11) we see that $e = 1$ for an elastic collision. In a perfectly inelastic collision the particles stick together after the impact, so that their separation velocity is zero. Thus $e = 0$ in a perfectly inelastic collision.

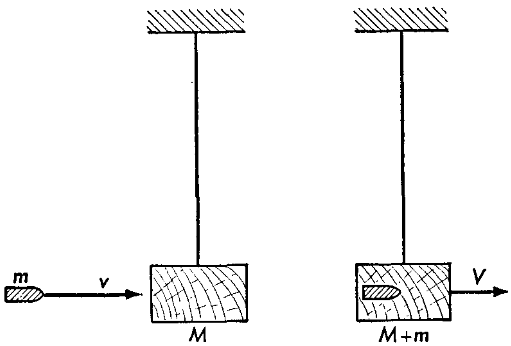


Fig. 10-4 (a) Initial momentum of system as bullet approaches target is mv; (b) final momentum of system just after bullet has penetrated target is $(M + m)V$.

Illustrative Example. A block of balsa wood whose mass is 600 gm is hung from a cord of negligible weight. A bullet whose mass is 2 gm and which has a muzzle velocity of 16,000 cm/sec is fired into this block at close range and becomes embedded in it. Determine the velocity with which the balsa wood is set in motion. What is the energy of the bullet before the collision? What is the energy of the system after the collision?

If we consider the bullet and the balsa-wood block as a single system, there is no external force acting on it, hence the momentum of the system is conserved in the collision process and is the same before the collision as after. The collision is a completely inelastic one, for, after the impact the bullet and the block of wood move as one, with common velocity V, and the final momentum of the system is $(M + m)V$, where M is the mass of the block and m is the mass of the bullet. The total momentum of the system before collision is mv, where v is the velocity of the bullet before the collision, as shown in Figure 10-4. From the principle of conservation of momentum, we write

$$mv = (M + m)V,$$

from which

$$V = \frac{m}{(M + m)} v,$$

and, substituting numerical values, we get

$$V = \frac{2 \text{ gm}}{602 \text{ gm}} \times 16{,}000 \frac{\text{cm}}{\text{sec}},$$

or

$$V = 52.2 \frac{\text{cm}}{\text{sec}}.$$

The kinetic energy of the system before collision is $\mathcal{E}_i$, where

$$\mathcal{E}_i = \tfrac{1}{2} m v^2$$
$$= \tfrac{1}{2} \times 2 \text{ gm} \times 256 \times 10^6 \frac{\text{cm}^2}{\text{sec}^2},$$

so that

$$\mathcal{E}_i = 256 \times 10^6 \text{ ergs},$$

and this is the total kinetic energy of the system before the collision, for the balsa block is at rest. The final kinetic energy of the system is $\mathcal{E}_f$, where

$$\mathcal{E}_f = \tfrac{1}{2}(M + m) V^2$$
$$= \tfrac{1}{2} \times 602 \text{ gm} \times \left(52.2 \frac{\text{cm}}{\text{sec}}\right)^2,$$

so that

$$\mathcal{E}_f = 0.84 \times 10^6 \text{ ergs}.$$

Note that mechanical energy has been lost in the collision process, a sign of an inelastic collision. From the principle of conservation of energy, we know that the energy has not vanished but has been transformed into heat, sound, deformation, and chemical change of the bullet and the balsa block.

Collision problems are of great importance in all of physics, from the realm of everyday experience to microscopic or atomic phenomena and to astronomical phenomena. Let us consider the collision of two perfectly elastic objects of the same mass, such as between two billiard balls, or between two nuclear particles such as two protons or two neutrons, or between a proton and a neutron (although the proton and neutron have slightly different masses). Let us suppose that before the collision one of these objects is at rest, and that the direction of the initial velocity U is along the line joining the centers of the two bodies. This will be called a central collision. Following the notation of the preceding discussion, we will suppose that $M = m$, and that u, the initial velocity of m, is equal to zero; that is, the sphere of mass m is at rest. From the momentum conservation equation, we find

$$MU = MV + mv,$$

but since

$$M = m,$$

we may write

$$U = V + v.$$

But from Equation (10-11), remembering that $u = 0$, we have

$$U = v - V.$$

Adding these two equations, we find that

$$U = v,$$

and that

$$V = 0.$$

Thus when two objects of equal mass collide in a central collision, all of the velocity of the moving object is transferred to the object which was formerly at rest. A similar analysis shows that a very light object bounces off a very heavy one with little transfer of energy, while, for objects of equal mass, all of the energy is transferred to the second object.

In ordinary experience these considerations help to determine the weight of a hammer used to drive a spike. If the hammer is too heavy, very little of the momentum and energy of the hammer will be transferred to the spike; the energy expended in driving the hammer will be inefficiently used, and, in addition, the work will be badly damaged by penetration of the hammer head itself. If the hammer is too light, it will bounce off the spike. In the ideal situation all of the energy and momentum of the hammer will be absorbed by the work, and the hammer will stop dead after the impact. Precisely the same considerations apply to the matching of the mass of the head of a croquet mallet to the mass of the ball and to countless other situations.

In one form of nuclear reactor, it is desired to use slow neutrons, although fast neutrons are emitted from uranium atoms which have undergone fission (see Chapter 46). If these neutrons are to be utilized efficiently in the reactor, they must be slowed down as rapidly as possible to very low speeds. This is accomplished by incorporating a *moderator* into the construction of the reactor. The moderator must contain nuclei whose mass is close to the mass of the neutron. Protons, having almost the same mass as neutrons, may be used to slow down the neutrons by elastic collision processes. Ordinary water is a convenient source of protons and can be used as a moderator; the kinetic energy and momentum of the neutrons are transferred to the protons in water in collision processes.

In atomic processes as well as in macroscopic processes, we may speak of inelastic collisions. In a collision between a rapidly moving proton and a molecule, the kinetic energy of the proton may dissociate the molecule, thus doing work against the forces which hold the molecule together. In the collision between an electron and an atom, the kinetic energy of the electron may be absorbed by the atom and converted into internal energy;

some of the internal energy may subsequently be emitted as light, as in the neon tube. In atomic and nuclear physics the principles of the conservation of energy and momentum and the concepts of elastic and inelastic collisions are extremely important in interpreting the data gathered from high-energy accelerators, from gaseous discharge tubes, from nuclear fission, and from many other atomic processes.

Problems

10-1. The hammer of a pile driver weighs 500 lb. The hammer is dropped from a height of 10 ft onto a pile, and drives the pile 3 in. (a) What is the momentum of the hammer before the impact? (b) What is the impulse delivered by the hammer to the pile? (c) Assuming that the pile exerts a steady force against the hammer, what is the magnitude of that force? (d) For how long a time interval does the hammer exert a force on the pile?

10-2. A 5-gm bullet is fired from a gun whose barrel is 60 cm long. The bullet leaves the gun with a muzzle velocity of 25,000 cm/sec. (a) What was the average force acting on the bullet? (b) What is the momentum of the bullet when it leaves the gun? (c) What impulse was delivered to the bullet? (d) How long was the bullet in the gun barrel?

10-3. A man driving a golf ball gives it a speed of 3,000 cm/sec. (a) If the mass of the golf ball is 50 gm, what impulse was imparted to it by the driver? (b) If the head of the driver has a mass of 75 gm, what fraction of the energy of the driver was delivered to the ball? Assume an elastic collision.

10-4. An ivory ball of mass 100 gm, moving with a velocity of 80 cm/sec, strikes a stationary ivory ball of equal mass. The velocity is parallel to the line joining the centers of the two balls. Assume a perfectly elastic collision. Determine the velocity of the two balls after collision.

10-5. An ivory ball of mass 400 gm, moving with a velocity of 90 cm/sec, strikes a stationary ivory ball having a mass of 100 gm. (a) Assuming perfect elasticity and a central collision, determine the velocity of each ball after collision. (b) What fraction of the energy of the first ball was transferred to the second ball?

10-6. An ivory ball of mass 200 gm, moving with a velocity of 100 cm/sec in the positive x direction, strikes a second ivory ball of mass 300 gm a glancing blow, so that the first ball is deflected by an angle of 30° from its initial direction and has a speed of 75 cm/sec after the collision. Find the speed and direction of motion of the second ball after the collision.

10-7. A man fires an automatic rifle, shooting a clip of 16 shells in 4 sec. Each bullet weighs 2 oz and has a muzzle velocity of 2,500 ft/sec. The man stands on a perfectly smooth floor. If the man and rifle weigh 160 lb, find the velocity with which the man is sliding backward after the 4-sec period.

10-8. A bullet whose mass is 2 gm is fired from a rifle with a muzzle velocity of 30,000 cm/sec into a piece of balsa wood mounted on a car with frictionless wheels. The total mass of the balsa wood and car is 1,500 gm. Determine (a) the initial momentum of the system, (b) the velocity of the balsa wood and car after the bullet was embedded in the wood, and (c) the mechanical energy lost in the impact.

10-9. An automobile weighing 2,400 lb, driving east at a speed of 50 mi/hr, collides with a heavily loaded truck driving west at a speed of 60 mi/hr. The loaded truck weighs 10,000 lb. If the collision is inelastic, find (a) the impulse received by the truck and (b) the impulse received by the car.

10-10. Four croquet balls are lined up in a frictionless trough. A fifth croquet ball, moving with speed v, strikes the end ball. Prove that only one ball will leave the other end with speed v in the same direction and of the same magnitude as the incident ball. Assume elastic collisions.

10-11. Show that a ball striking a wall in a perfectly elastic collision will bounce off the wall, making an equal angle with the wall. Is momentum conserved in this collision?

10-12. A fireman holds a fire hose which expels a stream of water from a 2-in.-diameter nozzle at a speed of 40 ft/sec. Find the force the fireman must exert to keep the nozzle stationary.

10-13. A ball is dropped onto an anvil from a height of 10 ft. If the coefficient of restitution is 0.1, to what height will the ball rise (a) after the first impact? (b) After the second impact?

10-14. If a ball of mass m is dropped from height h onto an anvil, and if the collision is perfectly elastic, find the time between successive impacts.

10-15. Repeat Problem 10-13 if the coefficient of restitution is e. Is the time between the first and second impact the same as the time between the second and third impact?

10-16. A ball A whose mass is 50 gm is moving to the right with a velocity of 80 cm/sec, and another ball B with a mass of 75 gm is moving to the left with a velocity of 120 cm/sec along the x axis. Determine (a) the initial momentum of the system, (b) the velocity of the center of mass of the system, and (c) the velocity of each ball after collision, assuming the collision to be elastic.

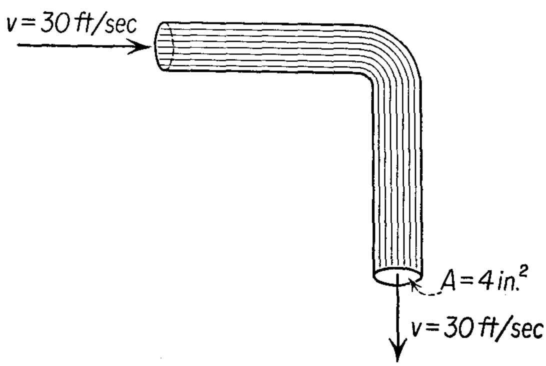


Fig. 10-5

10-17. A stream of water is moving through a horizontal pipe of uniform cross section of 4 in.^{2} with a speed of 30 ft/sec. The pipe has a right-angle bend in it, as shown in Figure 10-5. Determine the force that has to be exerted at the bend to hold the pipe in equilibrium.

10-18. A steel ball weighing 1 lb drops from a height of 16 ft, strikes a steel anvil, and rebounds to the same height. (a) Calculate the impulse on the ball. (b) If the time of contact between the ball and anvil is 0.002 sec, determine the impulsive force on the anvil.

10-19. The truck and car of Problem 10-9 are driven by men who weigh 160 lb. The drivers are fastened to their respective vehicles by seat belts. Find the impulse experienced by (a) the driver of the car, and (b) the driver of the truck in the collision.

10-20. A baseball of mass 250 gm leaves the bat at an angle of 30° above the horizontal with a velocity of 150 m/sec. Assume that the pitch was traveling horizontally with a velocity of 80 m/sec. If the bat was in contact with the ball for 0.05 sec, what was (a) the average horizontal component and (b) the average vertical component of the force exerted by the bat on the ball?

10-21. A ball of putty with mass 2 kg and velocity 25 m/sec strikes a wall and sticks. The collision lasts for 0.10 sec. (a) What is the magnitude of the average force exerted on the wall during the collision? (b) How much mechanical energy is lost during the collision? (c) What is the magnitude of the impulse transmitted to the wall?

11

Rotational Motion (The Dynamics of a Rigid Body)

11-1 Motion about a Fixed Axis

The motion of the flywheel of an engine and of a pulley on its axle are examples of an important type of motion of a rigid body, that of the motion of rotation about a *fixed axis*. Consider the motion of a uniform disk rotating about a fixed axis passing through its center of gravity C perpendicular to the face of the disk, as shown in Figure 11-1. The motion of this disk may be described in terms of the motions of each of its individual particles, but a better way to describe the motion is in terms of the *angle* through which the disk rotates. Calling two successive positions of a point in the plane of the disk P_1 and P_2, we find the angle of rotation by drawing radial lines from C to P_1 and to P_2. The angle θ between these two lines is the angle through which the disk has rotated; every point in the plane of the disk has rotated through the same angle θ in the same interval of time. The angle θ is called the *angular displacement* of the body. Both the angle θ and the direction of the axis of rotation must be given in order to specify properly a rotational displacement.

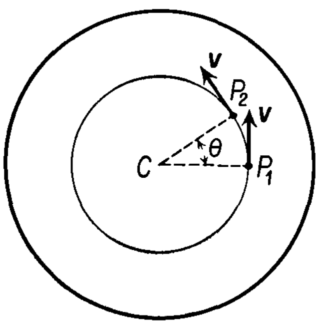


Fig. 11-1 Angle of rotation of a disk.

In spite of the apparent similarity between the specification of a rotational displacement and a linear displacement, an arbitrary rotational displacement is *not* a vector quantity, for one cannot add rotational displacements in the same way that linear displacements are added. Let us imagine that a blackboard eraser has its length initially directed along the x axis, and that the top face of the eraser is initially perpendicular to the

y axis. If we rotate the eraser first about the y axis by 90°, then about the z axis by 90°, the eraser lies on its side. If the rotation is first performed about the z axis, then about the y axis, the eraser will stand on end. The resultant of these two operations depends on the order in which they are

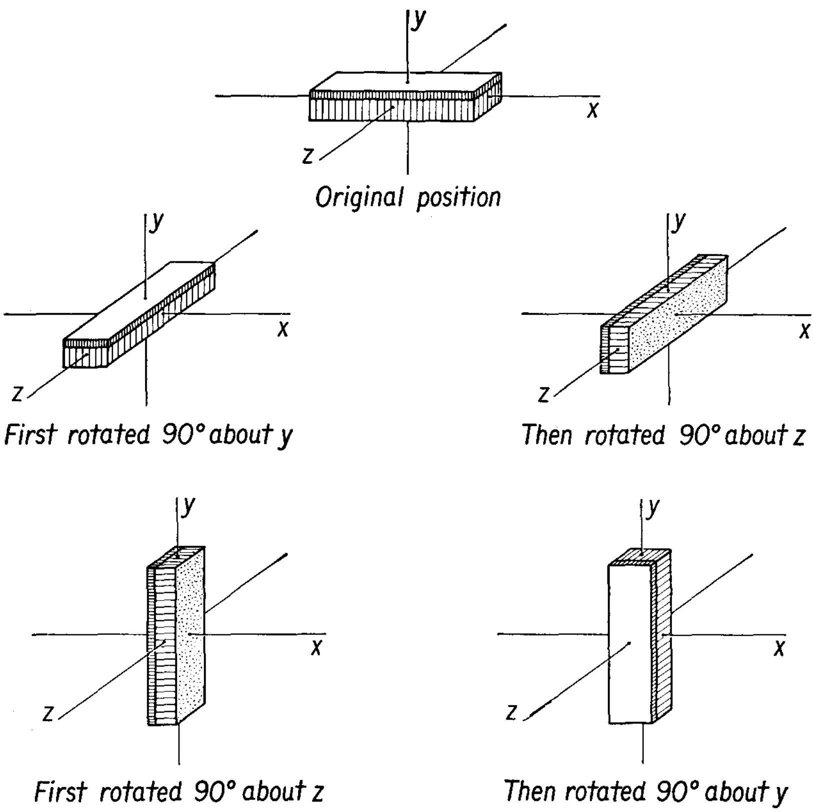


Fig. 11-2 The result of two finite rotations depends upon the order in which they are performed.

performed, as shown in Figure 11-2. As we have already seen, the resultant of two vectors, or of two linear displacements, does not depend on the order in which the sum is taken. Thus, although angular displacements involve both direction and magnitude, angular displacements of arbitrary magnitude cannot be true vectors.

If the rotational motion is restricted to rotation about a single *fixed* axis, it is possible to represent angular displacement as a vector quantity whose direction is parallel to that axis, in accordance with the right-hand rule previously given in the discussion of circular motion, for then the

resultant of two angular displacements does not depend on the order of rotations.

When the angular displacement of a body is restricted to *infinitesimal rotations*, these infinitesimal rotations may be thought of as vector angular displacements, for it may be shown that the sum of two infinitesimal rotations does not depend upon the order in which these rotations are performed. For this reason angular velocity is a vector quantity, for it is the result of dividing an infinitesimal angular displacement, a vector, by time, a scalar.

11-2 Kinetic Energy of Rotation

A rigid body rotating with uniform angular speed ω about a fixed axis possesses kinetic energy of rotation. Its value may be calculated by summing up the individual kinetic energies of all the particles of which the body is composed. A particle of mass m_1 located at distance r_1 from the axis of rotation has kinetic energy given by $\frac{1}{2}m_1v_1^2$, where v_1 is the speed of the particle. There will be a similar term for each particle making up the body, so that we may write, for the total kinetic energy $\mathcal{E}_k$,

$$\mathcal{E}_k = \tfrac{1}{2}m_1v_1^2 + \tfrac{1}{2}m_2v_2^2 + \cdots + \tfrac{1}{2}m_nv_n^2,$$

so that

$$\mathcal{E}_k = \textstyle\sum\frac{1}{2}m_iv_i^2.$$

Each particle of a rigid body rotates with uniform angular speed ω. Let us express the instantaneous linear speed of each particle in terms of the common angular speed. Remembering that $v = \omega r$, we substitute for v in the above equation to find

$$\mathcal{E}_k = \tfrac{1}{2}m_1r_1^2\omega^2 + \tfrac{1}{2}m_2r_2^2\omega^2 + \cdots + \tfrac{1}{2}m_nr_n^2\omega^2,$$

or

$$\mathcal{E}_k = \tfrac{1}{2}\omega^2(m_1r_1^2 + m_2r_2^2 + \cdots + m_nr_n^2).$$

Let us denote the factor in parentheses by the letter I; that is,

$$I = m_1r_1^2 + m_2r_2^2 + \cdots + m_nr_n^2,$$

or

$$\boxed{I = \textstyle\sum m_ir_i^2,} \tag{11-1}$$

so that the kinetic energy of the rotating body may be written as

$$\boxed{\mathcal{E}_k = \tfrac{1}{2}I\omega^2.} \tag{11-2}$$

The factor I is called the *moment of inertia* of the rotating body with respect to the particular axis of rotation. The moment of inertia depends upon the manner in which the mass is distributed with respect to the axis.

Clearly, the moment of inertia will be greatest when the mass is farthest from the axis of rotation. In the motion of rotating systems, the moment of inertia plays a role analogous to that of the mass in translational systems or in linear motion. Unlike the mass, which is a constant for a particular body, the moment of inertia depends upon the location and direction of the axis of rotation as well as upon the way the mass is distributed.

11-3 Moments of Inertia of Simple Bodies

The moment of inertia of a system of particles is given by Equation (11-1) as

$$I = \sum m_i r_i^2 = m_1 r_1^2 + m_2 r_2^2 + \cdots + m_n r_n^2.$$

Let us calculate the moment of inertia of several simple distributions of particles.

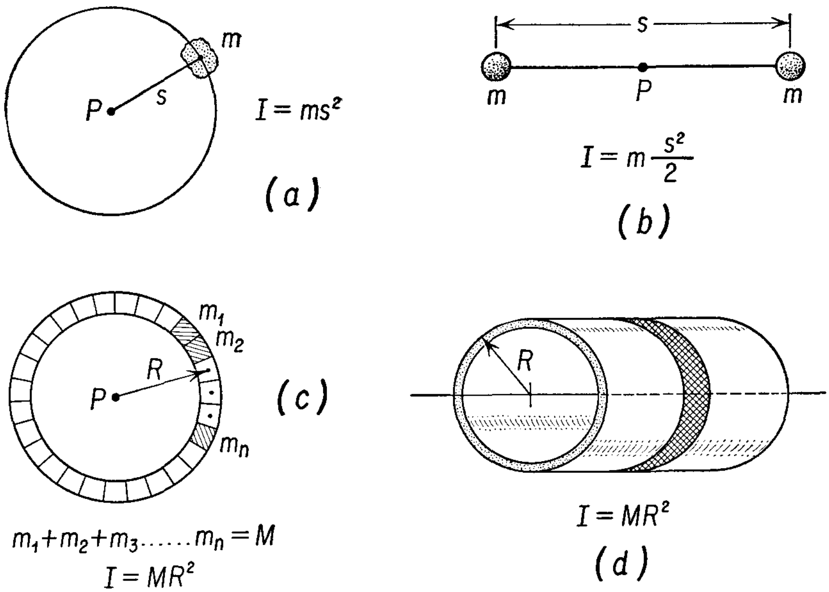


Fig. 11-3 Moments of inertia of some bodies of simple geometrical shapes. The axis is perpendicular to the paper and passes through P in (a), (b), and (c). In (d) the axis is the geometrical axis of the cylinder.

Consider a small stone of mass m attached to a long weightless string of length s, whose other end is fixed to a pivot P, as in Figure 11-3(a). Since there is only one mass to consider, the summation reduces to a single term, and the moment of inertia is given by $I = ms^2$.

A dumbbell, consisting of two equal masses m separated by a long weightless bar of length s free to rotate about its center of gravity at the

point P midway between the two masses, as shown in Figure 11-3(b), has a moment of inertia given by

$$I = m\left(\frac{s}{2}\right)^2 + m\left(\frac{s}{2}\right)^2 = \frac{ms^2}{2}.$$

A thin ring of mass M and mean radius R which is free to rotate about its center may be thought of as a collection of segments of mass m_1, m_2, m_3, and so on, as shown in Figure 11-3(c), each of which is located at a distance R from the axis of rotation. Applying Equation (11-1) to the ring, considered as a collection of particles, we find

$$\begin{aligned} I &= m_1R^2 + m_2R^2 + m_3R^2 + \cdots + m_nR^2 \\ &= (m_1 + m_2 + m_3 + \cdots + m_n)R^2, \end{aligned}$$

and since the summed mass of the segments is equal to the mass M of the ring, we find, for the moment of inertia of a hollow ring,

$$I = MR^2.$$

A hollow cylinder of mass M which is free to rotate about an axis through its center may be thought of as a stack of rings, as shown in Figure 11-3(d). From Equation (11-1) we see that the moment of inertia of a collection of matter about a given axis is simply the sum of the moments of inertia of each of the separate parts about the same axis. Thus the moment of inertia of a hollow cylinder of radius R about its axis is given by the same formula as the moment of inertia of a hollow ring, $I = MR^2$, where M now represents the mass of the cylinder.

A body which is composed of a distribution of matter rather than a collection of mass points must be imagined as segmented into small pieces approximating point masses. The moment of inertia is calculated by summing the quantity mr^2 over each of the imagined segments. Better approximations to the true moment of inertia of the body may be made by imagining the body to be broken up into finer and finer subdivisions. In the limit of an infinitesimally fine subdivision, the sum is replaced by an integral, and in the language of the calculus, if dm represents the mass of one element of the body of volume dv, located at a distance r from the axis of rotation, the moment of inertia of the body is given by

$$I = \int r^2\, dm. \tag{11-3a}$$

If the body is a homogeneous one of density ρ, the mass of a small element of volume dv is given by

$$dm = \rho\, dv,$$

and the moment of inertia may be written as

$$I = \int \rho r^2 \, dv. \tag{11-3b}$$

For bodies of complicated shape, the evaluation of the integral may be quite difficult, but for bodies of simple geometric shape, the evaluation of the integral is well within the reach of an introductory course in the calculus.

Illustrative Example. Calculate the moment of inertia of a rod of length L and cross-sectional area A about an axis perpendicular to the rod through one end, as shown in Figure 11-4. Suppose the density of the rod is ρ; the volume of an

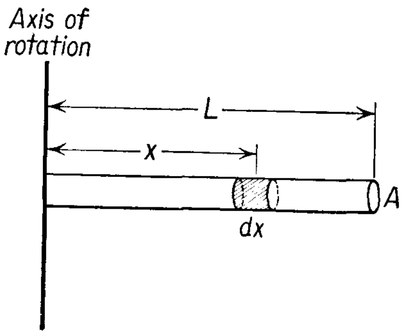


Fig. 11-4 Determining the moment of inertia of a rod.

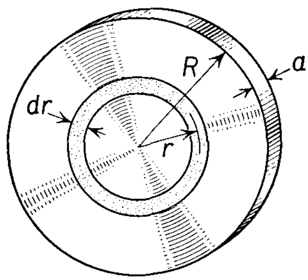


Fig. 11-5 Determining the moment of inertia of a solid disk.

element of length dx is given by $A\,dx$, and the mass dm of the element is $\rho A\,dx$. The moment of inertia of this element, located at a distance x from the axis of rotation, is $dI = \rho A x^2\,dx$, and the moment of inertia of the entire rod is obtained by summing, or integrating, the contributions to the moment of inertia of each element of the rod. Thus

$$I = \int_0^L dI = \int_0^L \rho A x^2\,dx = \left[\rho A \frac{x^3}{3}\right]_0^L$$

$$= \rho A \frac{L^3}{3}.$$

Remembering that AL is the volume of the rod, the mass of the rod M is given by ρAL. We write for the moment of inertia of the rod about a perpendicular axis through one end,

$$I = \frac{ML^2}{3}.$$

Illustrative Example. Calculate the moment of inertia of a flat solid disk of radius R and mass M about an axis through its center of mass and perpendicular to the plane of the disk. Let us suppose the disk to be of thickness a, and made of a homogeneous material of density ρ. To calculate the moment of inertia of the disk, we imagine the disk to be made up of a set of nested rings, as shown in Figure 11-5. The surface area of a ring of mean radius r and width dr is given by

$2\pi r\,dr$, and the mass dm of such a ring is the product of its volume by its density; that is, $dm = 2\pi r a \rho\,dr$. The moment of inertia of a ring is its mass times the square of its radius, and the contribution of the moment of inertia of any one such ring to the moment of inertia of the disk is

$$dI = (2\pi r\,dr\,a\rho)r^2 = 2\pi a r^3 \rho\,dr.$$

The moment of inertia of the entire disk is found by adding the moments of

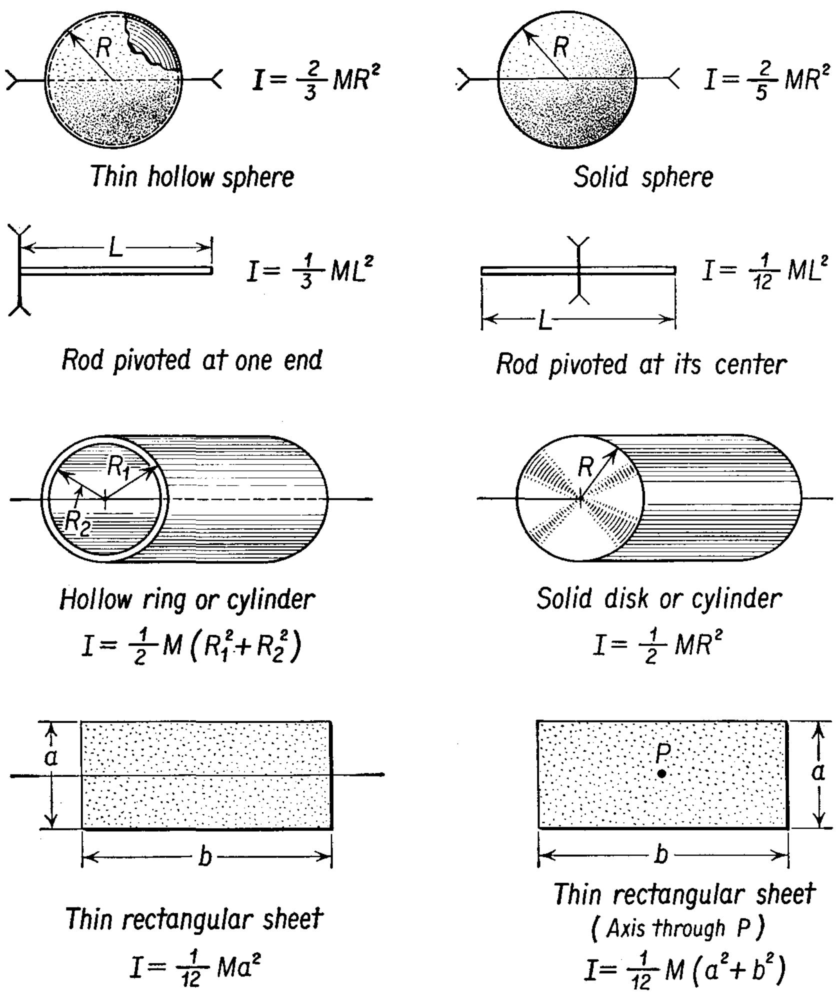


Fig. 11-6 Moments of inertia of several bodies of simple geometrical shapes.

inertia of all the rings which are imagined to constitute the disk. Thus

$$I = \int_0^R dI = \int_0^R 2\pi a\rho r^3\,dr = \left[2\pi a\rho\,\frac{r^4}{4}\right]_0^R$$
$$= \pi R^2 a\rho\,\frac{R^2}{2}\cdot$$

The factor $\pi R^2 a\rho$ may be recognized as the volume of the disk times density, or the mass M of the disk, and we have

$$I = \frac{MR^2}{2}\cdot$$

The moment of inertia of a solid disk about a central axis perpendicular to its face is half that of a hollow ring having the same mass and the same radius.

The moments of inertia of several bodies of simple geometric shapes are given in Figure 11-6. The moment of inertia is a useful and important concept in the study of the strength of materials, for it figures prominently in formulas for the strength of such members as angles and I beams.

The units for moment of inertia are those of mass multiplied by the square of a distance, for example, gm cm^2, or kgm m^2, or slugs ft^2.

11-4 The Parallel Axis Theorem

A theorem in mechanics which is very useful in the study of rotational motion is called the *parallel axis theorem* which states that *if the moment of inertia of a body about an axis through its center of mass is known, the moment of inertia of the body about any axis parallel to the first is given by the moment of inertia about the axis through the center of mass plus the product of the mass of the body by the square of the perpendicular distance between the two axes.* In the form of an equation we write

$$I = I_c + MR^2, \tag{11-4}$$

where I_c is the moment of inertia of the body about an axis through its center of mass, M is the mass of the body, and R is the perpendicular distance from the center of mass to the axis of rotation.

Illustrative Example. Find the moment of inertia of a hollow ring about an axis perpendicular to the plane of the ring which passes through a point on the circumference.

The moment of inertia of a hollow ring of mass M and radius R about an axis through its center of mass I_c perpendicular to the face of the ring has been shown to be equal to MR^2. The moment of inertia I of the ring about a parallel axis through its circumference is equal to

$$\begin{aligned} I &= I_c + MR^2 \\ &= MR^2 + MR^2 \\ &= 2MR^2. \end{aligned}$$

11-5 Torque and Angular Acceleration

In our discussion of the equilibrium of a rigid body, we found that, when the vector sum of all the torques acting on a body is zero, the body is in equilibrium as far as rotational motion is concerned. If an external torque acts on the body, it will acquire an angular acceleration $\boldsymbol{\alpha}$ given by

$$\boxed{\mathbf{G} = I\boldsymbol{\alpha},} \tag{11-5}$$

where **G** is the sum of all the external torques acting on the body about a fixed axis, and I is the moment of inertia of the body about the same axis. Equation (11-5) may be derived from Newton's laws of motion and represents a special form of Newton's equation applied to rotational motion.

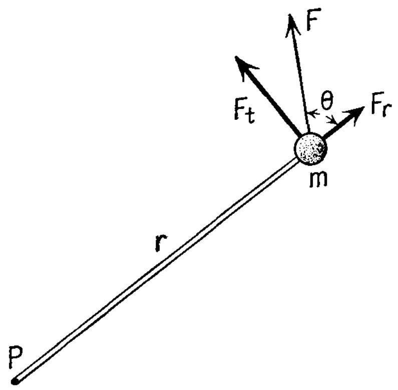


Fig. 11-7

Suppose that a particle of mass m is constrained to move in a circular path by a rigid weightless rod of length r about a point P, as shown in Figure 11-7. An arbitrary force F can only cause it to move in a tangential direction, for motion in the radial direction is not permitted by the rod. If the angle between the force F and the rod is given by θ, we can resolve the force into a radial component F_r and a tangential component F_t. The radial component produces no torque about the axis through P, hence we need consider only the effect of the tangential component.

From Newton's second law we may write, for the tangential component,

$$F_t = ma_t,$$

where a_t is the tangential component of the acceleration of the particle. We have already seen that the tangential acceleration of a particle moving in circular motion may be related to its angular acceleration α through the equation

$$a_t = \alpha r,$$

and, substituting into the equation above, we find

$$F_t = m\alpha r.$$

From Figure 11-7, the value of the tangential component of F is given by

$$F_t = F \sin \theta.$$

Thus

$$F \sin \theta = m\alpha r,$$

and, multiplying both sides of the equation by r, we find

$$Fr \sin \theta = mr^2\alpha.$$

The quantity $Fr \sin \theta$ on the left-hand side of the equation is exactly the torque G exerted by the force F about the axis through P, while the quantity mr^2 is the moment of inertia I of the particle about the same axis. Thus

$$G = I\alpha,$$

and we have verified Equation (11-5) for the simple case of a mass particle constrained to rotate about a fixed axis. Note that the directions of the vector $\mathbf{G}$ and of the vector $\boldsymbol{\alpha}$ are both perpendicular to the plane of the paper, pointing outward, in accordance with the right-hand rule.

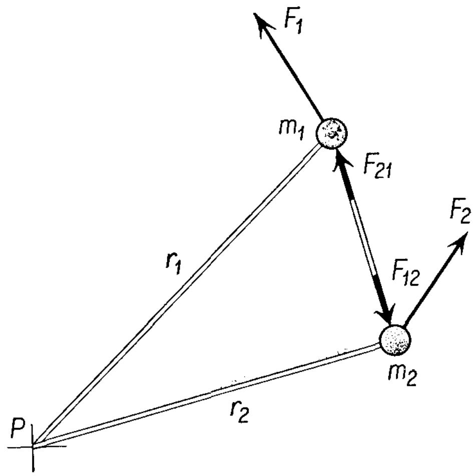


Fig. 11-8

Suppose we had a system of two particles m_1 and m_2 rigidly connected to each other and to the axis of rotation by a framework of weightless rods, as shown in Figure 11-8. The two particles and their framework are constrained to move in circular motion with a common angular velocity and common angular acceleration. Let the external force exerted on the particle of mass m_1 be F_1, the external force exerted on the second particle of mass m_2 be F_2, while the forces exerted *by* the second particle *on* the first is F_{21}, and *by* the first particle *on* the second is F_{12}. From Newton's third law these two forces must be equal and opposite to each other. Remembering that neither m_1 nor m_2 is free to move in the radial direction, we

apply Newton's second law to the tangential motion of each particle:

$$(F_1)_t + (F_{21})_t = m_1(a_1)_t,$$

$$(F_2)_t + (F_{12})_t = m_2(a_2)_t.$$

Let us multiply the first of these equations by r_1 and the second by r_2. The product $(F_1)_t r_1$ is the torque G_1 of the force F_1 about the axis of rotation. Similarly, the product $(F_{21})_t r_1$ is the torque G_{21} of the internal force F_{21} about the axis of rotation, and we may write

$$G_1 + G_{21} = m_1(a_1)_t r_1 = m_1 r_1^2 \alpha,$$

$$G_2 + G_{12} = m_2(a_2)_t r_2 = m_2 r_2^2 \alpha.$$

Since F_{21} and F_{12} are directed along the same straight line in opposite directions, and since they have equal magnitudes, the torques G_{21} and G_{12} are equal and opposite:

$$G_{12} = -G_{21}.$$

Adding the two equations, we find

$$G_1 + G_2 = (m_1 r_1^2 + m_2 r_2^2)\alpha,$$

or

$$G = I\alpha.$$

Once again we have verified Equation (11-5), that the sum of the external torques acting on a rigid body is equal to the product of the moment of inertia by the angular acceleration. Following the same procedure, the equation may be shown to be true for an arbitrary number of mass particles connected by a rigid framework and hence for a continuous rigid body.

The equation $\mathbf{G} = I\boldsymbol{\alpha}$, is the rotational counterpart of the equation $\mathbf{F} = M\mathbf{a}$. The only assumption which has been introduced in the study of rotational motion is that the rotating body is a rigid one in which the force exerted by one part of the body on another lies along the line joining the two parts. We see also that the rotational equilibrium of a rigid body is a special case of Equation (11-5). A rigid body is in rotational equilibrium when its angular acceleration is zero, and hence when the sum of the external torques acting upon it is zero. This is exactly analogous to the translational equilibrium of a rigid body, which may be considered as a special case of the equation $\mathbf{F} = M\mathbf{a}$, for a body is in translational equilibrium when its acceleration is zero, and hence when the sum of the external forces is zero.

Illustrative Example. A disk 30 cm in diameter and having a mass of 900 gm is mounted so that it can rotate about a fixed axis passing through its center, as in Figure 11-9. A mass of 200 gm hangs from a string which is wound around the disk. Determine (a) the acceleration of the 200-gm mass, (b) the angular acceleration of the disk, and (c) the tension in the string.

To solve the problem we first examine the figure to establish a simple sign convention. The weight will be accelerated in the downward direction by the force of gravity, and at the same time the disk will be caused to rotate in a clockwise direction. Let us call the clockwise direction the positive direction of rotation, and the downward direction the positive direction of translation. From the nature of the force which can be exerted by a string, the upward force exerted by the string on the falling mass must be equal in magnitude to the downward force exerted on the disk by the string. We shall call the magnitude of this force S and indicate the directions on the diagram. From the nature of the constraints, we note that the only way the mass can receive a given downward displacement is for a length of string to unroll from the disk. If the radius of the disk is R, the disk must receive an angular displacement θ when the mass is displaced by a distance s such that $s = \theta R$. Note that the directions of the angular and linear displacements as well as their magnitudes are related by this equation, for a positive value of s implies a positive value of θ. Thus we have $v = \omega R$ and $a = \alpha R$, where v and a are the velocity and acceleration of the mass, positive downward, and ω and α are the angular velocity and angular acceleration of the disk, positive clockwise. With the relationships of constraint and the sign conventions established, we proceed to a formal solution of the problem.

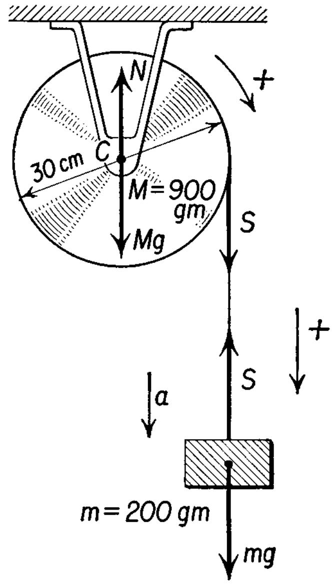


Fig. 11-9

Let us first consider the forces acting on the 200-gm mass. From Newton's second law we can write

$$mg - S = ma.$$

Now let us consider the rotation of the disk. The only force acting on the disk that exerts a torque about the axis of rotation through C is the pull of the cord S. If R is the radius of the disk and I is its moment of inertia, we find, on substituting in Equation (11-5),

$$G = SR = I\alpha,$$

and, since

$$a = \alpha R,$$

we have

$$SR = \frac{Ia}{R}\cdot$$

Substituting the value of S into the first equation, we find

$$mg - \frac{Ia}{R^2} = ma,$$

and, solving for a, we find

$$a = \frac{mg}{m + \frac{I}{R^2}}.$$

The moment of inertia of a uniform disk about an axis through C is $I = \frac{MR^2}{2}$,

so that
$$a = \frac{mg}{m + M/2}.$$

The numerical values are $m = 200$ gm, $M = 900$ gm, $R = 15$ cm, and $g = 980$ cm/sec^2, and we find

(a) $$a = 309 \frac{\text{cm}}{\text{sec}^2};$$

(b) $$\alpha = \frac{a}{R} = 20.5 \frac{\text{radians}}{\text{sec}^2};$$

(c) $$S = m(g - a)$$
$$= 200 \text{ gm} \times (980 - 309) \frac{\text{cm}}{\text{sec}^2},$$

so that
$$S = 134{,}800 \text{ dynes}.$$

The tension in the cord is less than the weight of the 200-gm mass hanging from its end.

11-6 Rotational Energy, Work, and Power

Whenever a rigid body is set into rotation about an axis, work is done by the torques acting on it to increase its kinetic energy of rotation. Suppose

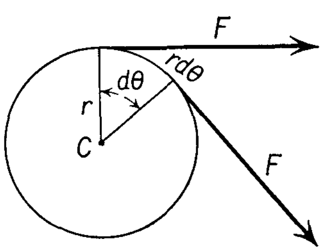


Fig. 11-10 The work done by a constant torque G when it acts through an angle $d\theta$ is $G\,d\theta$.

that a force F acts on the rim of a wheel of radius r and rotates the body through an angle $d\theta$, as shown in Figure 11-10. The displacement of the wheel is $r\,d\theta$, and the work done by this force is

$$d\mathcal{W} = Fr\,d\theta;$$

but
$$Fr = G;$$

therefore
$$d\mathcal{W} = G\,d\theta, \tag{11-6a}$$

or the work $d\mathscr{W}$ done by a torque G is equal to the product of the torque and the angle $d\theta$ through which it acts.

Just as in the case of the work done by a force, the work of an applied torque is done by the component of the torque parallel to the axis of rotation. In vector notation, if a torque **G** produces a rotation $d\boldsymbol{\theta}$, the work done is

$$\boxed{d\mathscr{W} = \mathbf{G} \cdot d\boldsymbol{\theta}.} \tag{11-6b}$$

If a constant torque acts on a rigid body which is rotating about a fixed axis, then, from the principle of conservation of mechanical energy, assuming no loss due to friction, the work done by the torque will produce a change in the kinetic energy of the body given by

$$\mathscr{W} = G\theta = \tfrac{1}{2}I\omega_f^2 - \tfrac{1}{2}I\omega_i^2, \tag{11-7}$$

in which ω_f is the final angular speed of the body, ω_i is the initial angular speed of the body, and θ is angular displacement through which the torque has acted.

Illustrative Example. The flywheel of a steam engine whose moment of inertia is 72 kg m^2 is given an angular speed of 150 rpm in 90 rev, starting from rest. Determine the torque, assuming it to be constant, which acted on the flywheel.

The angle θ through which the torque acted is

$$\theta = 90 \times 2\pi = 180\pi \text{ radians.}$$

The final speed of the flywheel is

$$\omega_f = 150\,\frac{\text{rev}}{\text{min}} \times \frac{1\text{ min}}{60\text{ sec}} \times \frac{2\pi\text{ radians}}{1\text{ rev}}$$

$$= 5\pi\,\frac{\text{radians}}{\text{sec}}\cdot$$

Applying Equation (11-7) to the solution of the problem,

$$G \times 180\pi = \tfrac{1}{2} \times 72\text{ kg m}^2 \times 25\pi^2\,\frac{1}{\text{sec}^2},$$

$$G = 15.7\text{ nt m.}$$

If the constant torque G is applied for a time dt to the rotation about a fixed axis, then we may find the power $\mathscr{P}$ by dividing both sides of Equation (11-6b) by the time dt,

$$\mathscr{P} = \frac{d\mathscr{W}}{dt} = \mathbf{G} \cdot \frac{d\boldsymbol{\theta}}{dt}$$

$$\mathscr{P} = \mathbf{G} \cdot \boldsymbol{\omega}. \tag{11-8a}$$

When the torque is parallel to the axis of rotation,

$$\mathcal{P} = G\omega, \tag{11-8b}$$

for, by definition, the angular velocity is equal to the angular displacement divided by the time. Thus the power $\mathcal{P}$ expended by a constant torque G applied for a time t is equal to the product of the torque by the angular velocity. Equation (11-8b) is the rotational analogue of the equation $\mathcal{P} = Fv$.

Illustrative Example. A $\frac{1}{2}$-hp motor is designed to operate at a speed of 1,750 rpm. What is the torque which the shaft of the motor can exert when operating at the rated speed?

The power delivered by the motor is $\frac{1}{2} \times 550 \dfrac{\text{ft lb}}{\text{sec}}$, while its angular speed is $1{,}750 \times \dfrac{2\pi}{60} \dfrac{\text{radians}}{\text{sec}}$. Substituting in Equation (11-8b), we find

$$\frac{550}{2} \frac{\text{ft lb}}{\text{sec}} = G \times 1{,}750 \times \frac{2\pi}{60} \frac{\text{radians}}{\text{sec}},$$

$$G = 1.50 \text{ lb ft.}$$

11-7 Angular Momentum and Angular Impulse

A rigid body rotating with angular velocity $\boldsymbol{\omega}$ about a fixed axis has an angular momentum $\mathbf{p}_\theta$ about this axis given by

$$\mathbf{p}_\theta = I\boldsymbol{\omega}, \tag{11-9}$$

where I is the moment of inertia of the body about this axis. Note that since the angular velocity about a given axis is a vector quantity which lies parallel to the axis of rotation, in a direction given by the right-hand rule, and the moment of inertia about this axis is a scalar quantity, the *angular momentum is a vector quantity.* To change the angular momentum of a body, an external torque must be applied to it. Remembering that the instantaneous angular acceleration is given by the derivative of the angular velocity with respect to the time, we may write Equation (11-5) as

$$\mathbf{G} = I\frac{d\boldsymbol{\omega}}{dt} = \frac{d(I\boldsymbol{\omega})}{dt}.$$

Substituting $\mathbf{p}_\theta$ for $I\boldsymbol{\omega}$ from Equation (11-9), we have

$$\mathbf{G} = \frac{d}{dt}(\mathbf{p}_\theta), \tag{11-10}$$

or the torque acting on a rigid body is equal to the rate of change of the angular momentum. Although Equation (11-10) was here derived for a rigid body, it may be shown that a system of particles obeys the same rule;

that the rate of change of the total angular momentum of the system of particles is equal to the sum of the external torques acting on the system of particles. *In the absence of external torques, the angular momentum of a rigid body must be constant;* that is, there is no change in the angular momentum of a rigid body when the sum of the external torques is zero. This is known as the principle of *conservation of angular momentum,* and, like the principle of conservation of energy and the conservation of linear momentum, is one of the most important general principles of mechanics.

Just as in the case of linear motion, we may treat impulsive motion in the case of rotation by examining the incremental form of the equation relating the torque to the rate of change of angular motion. We may write

$$\mathbf{G} = \frac{\Delta \mathbf{p}_\theta}{\Delta t},$$

and, multiplying through the equation by the time interval Δt during which the torque $\mathbf{G}$ is applied, we find the angular impulse $\Delta \mathbf{J}_\theta$ to be

$$\Delta \mathbf{J}_\theta = \mathbf{G}\,\Delta t = \Delta \mathbf{p}_\theta. \tag{11-11}$$

Thus the change in angular momentum is equal to the angular impulse.

An example of angular-momentum changes due to an angular impulse is the operation of the clutch in an automobile where a rotating disk connected to the engine engages a second disk connected to the rear wheels.

Since the angular momentum is a vector quantity, a rigid body set spinning on its axis will maintain its direction of rotation as well as its angular speed, providing no external torque acts on it.

Examples of the operation of the principle of conservation of angular momentum are numerous, in everyday life as well as in astronomy and in atomic and nuclear physics. The force exerted by the sun and by other celestial bodies on the earth is directed through the center of the earth (to a good approximation) in accordance with Newton's law of universal gravitation. Since the axis of rotation of the earth passes through its center, these forces exert no torque on the earth about its axis of spin. Consequently, the angular momentum of the earth and the length of the day are constant.

Consider a stone attached to the end of a string being whirled in a horizontal circle. If the string is made to wind itself around a vertical stick, becoming shorter with each revolution, the stone is observed to whirl with increasing angular speed as the string winds itself up. As the string becomes shorter, the moment of inertia of the stone about its axis of rotation is decreased. Since the force exerted on the stone by the string is in the radial direction, there is no external torque exerted by the string on the stone. Its angular momentum remains constant, but the decrease in moment of inertia must be accompanied by an increase in its angular speed.

A cat manages to fall on its feet, a diver can land in the water headfirst, an ice skater can execute a pirouette on the toe of one skate, all through the action of the principle of conservation of angular momentum. The principle of conservation of angular momentum explains why the changing mass distribution of the earth, as the result of volcanoes, tides, and winds, affects the instantaneous speed of rotation of the earth on its axis. In atomic and nuclear physics the atom or nucleus is acted upon by external forces which act through the center of mass of the system. The angular momentum of an atom or nucleus about its center of mass is constant, and, in fact, the value of the angular momentum of an atom or a nucleus is one of the more important pieces of information which can be used to describe atomic or nuclear systems.

11-8 Rolling Motion

The motion of a wheel which is rolling along the ground, without slipping, can be considered in one of two ways: either as a rotation of the wheel about an axis through its center of gravity C and an additional translational motion of the entire wheel with the same velocity as the center of gravity,

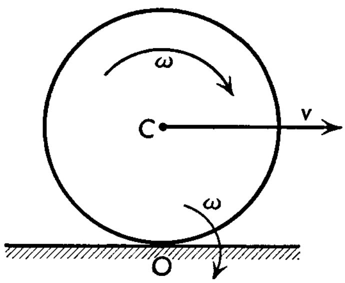


Fig. 11-11 Wheel rolling on the ground. ω is the angular velocity and v is the linear velocity of the center of gravity.

as shown in Figure 11-11, or as a rotation of the wheel about an *instantaneous axis* through the point of contact O between the wheel and the ground. A point in the body which is on the instantaneous axis is momentarily at rest. The instantaneous axis itself moves forward as the wheel moves forward, but it always remains parallel to itself and to the axis through the center of gravity. The angular velocity of the wheel about the instantaneous axis is the same as that about the axis through the center of gravity.

The general motion of a rigid body may be thought of as made up of two parts: one a motion of translation of the center of gravity, with the entire mass of the body acting as though it were concentrated there, and the other a motion of rotation of the body about an axis through the center of gravity. The angular velocity and angular acceleration are calculated

by taking the torques and the moment of inertia about an axis through the center of gravity. The linear velocity and the linear acceleration of the center of gravity are then calculated by considering all the forces as though they acted through the center of gravity, and by applying Newton's second law of motion to this case.

Illustrative Example. A disk of radius R and mass M rolls without slipping down an inclined plane of height h. Discuss the motion of the disk.

There are three forces acting on the disk, its weight Mg, the normal force N perpendicular to the plane and passing through the center of the disk, and the

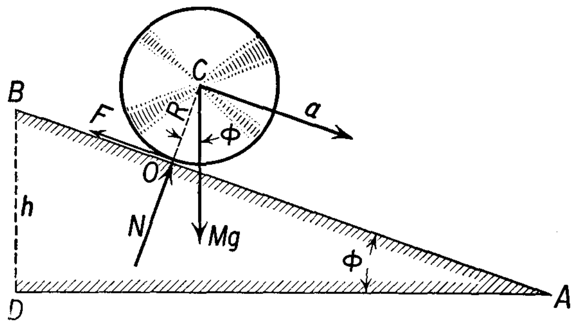


Fig. 11-12 Disk rolling down an inclined plane.

frictional force F acting parallel to the plane at the point of contact O, as illustrated in Figure 11-12. The instantaneous axis of rotation passes through point O and is perpendicular to the plane of the disk.

Let us determine the torques about the instantaneous axis through O. Since both F and N pass through O, they contribute nothing to the torque about the axis through O. The moment arm of the force Mg with respect to the axis through O is $R \sin \phi$, where ϕ is the angle of the inclined plane. Hence the torque equation becomes

$$MgR \sin \phi = I\alpha.$$

From the parallel axis theorem, the moment of inertia of the disk about an axis through O is

$$I = I_c + MR^2,$$

where I_c is the moment of inertia about the center of gravity. Solving the torque equation for α, we find

$$\alpha = \frac{MgR \sin \phi}{I},$$

and the acceleration of the center of gravity a is given by

$$a = \alpha R = \frac{MgR^2 \sin \phi}{I}.$$

It is interesting to compare the linear acceleration of the center of gravity of a solid disk of mass M with a hollow ring of the same mass. Note that the analysis thus far does not specify whether the disk is solid or hollow. For a solid disk we have seen that $I_c = \frac{1}{2}MR^2$, while for a hollow ring of the same mass and radius, $I_c = MR^2$, so that for a solid disk $I = \frac{3}{2}MR^2$, while for a hollow ring $I = 2MR^2$. Thus for a solid disk

$$a = \tfrac{2}{3}g \sin \phi,$$

while for a hollow ring

$$a = \tfrac{1}{2}g \sin \phi,$$

so that a solid disk will roll down an incline with greater acceleration than a hollow disk of the same radius. Note that neither the mass of the disk nor its radius appears in the expression for the acceleration. Thus all solid disks will roll down an inclined plane faster than all hollow rings. A similar relationship may be found for spheres.

It may be observed that the linear acceleration a of the center of gravity is less than the acceleration of a body which slides down a similar but frictionless inclined plane.

It is instructive to consider this problem from the energy point of view. If the disk rolls down the incline without slipping, there is no energy lost in doing work against the frictional force. If the disk starts at the top of the incline with zero kinetic energy, its total energy is its potential energy Mgh. At the bottom of the incline its energy is all kinetic. We may calculate the kinetic energy in two ways. First, let us consider that the motion of the disk at the bottom of the incline consists of rotation about the instantaneous point of contact. Let us consider the motion of a solid disk, whose moment of inertia about a point on its rim is $I = \frac{3}{2}MR^2$. The kinetic energy is all rotational about the point of contact, and we have

$$Mgh = \tfrac{1}{2}I\omega^2,$$

$$Mgh = \tfrac{3}{4}MR^2\omega^2,$$

$$\omega^2 = \frac{4gh}{3R^2},$$

$$\omega = \sqrt{\frac{4gh}{3R^2}},$$

where ω is the angular speed of the disk at the bottom of the incline.

Now let us consider the rolling disk, alternatively, as having its energy made up of two parts. The motion may be thought of as a translational motion of the center of gravity with kinetic energy $\frac{1}{2}Mv^2$, and a rotational motion about the center of gravity with kinetic energy $\frac{1}{2}I\omega^2$. The moment of inertia of a solid disk about its center of gravity is $I = \frac{1}{2}MR^2$. The total kinetic energy of the disk at the bottom of the incline is therefore given by

$$\mathcal{E}_k = \tfrac{1}{2}Mv^2 + \tfrac{1}{2}I\omega^2,$$

and, remembering that the linear velocity of the center of gravity is related to the

angular velocity of the disk by

$$v = \omega R,$$

the kinetic energy becomes

$$\mathcal{E}_k = \tfrac{1}{2}M\omega^2R^2 + \tfrac{1}{2} \times \tfrac{1}{2}MR^2 \times \omega^2,$$

$$\mathcal{E}_k = \tfrac{3}{4}MR^2\omega^2,$$

exactly as before.

11-9 Rotational Motion and Linear Motion

In Section 6-6 we have already seen that many of the equations developed for linear motion could be transcribed to the problem of angular motion simply by replacing the symbols used to describe linear motion by a set of symbols appropriate to angular motion. From the present chapter it is possible to display some additional analogues. Thus in rotational motion the moment of inertia I plays a role analogous to that of the mass M in linear motion. The angular momentum $I\boldsymbol{\omega}$ plays a role analogous to that of the linear momentum $M\mathbf{v}$. The torque $\mathbf{G}$ and the force $\mathbf{F}$ play analogous roles. It is instructive to transcribe a number of the equations which have been thus far developed to display the similarity of the equations frequently used in mechanics for both linear and angular motions, as in Table 11-1. The formulas in the table have been somewhat simplified over the formulas developed in the text. In particular, the formulas are all indicated as relating scalar quantities, and the displacement and velocity are assumed parallel to the force in the equations for work and power.

Although there are analogous quantities used to describe linear motion and rotational motion, it must be recognized that these quantities represent quite different things. The mass is a constant quantity and is independent of the position of any coordinate frame, but this is distinctly not true of the moment of inertia. The moment of inertia is a different number for every different axis of rotation, whether the difference is due to position or to orientation. Similarly, it is difficult to conceive just how one might transport mechanical energy in the form of linear kinetic energy in a package, yet every engine has a flywheel which carries rotational kinetic energy. The behavior of a system having linear momentum in response to an applied force is very similar to the behavior of the system at rest to the same applied force. We shall see in the next section that the way a system having angular momentum responds to an applied torque is far more complex and more interesting.

11-10 The Gyroscope

We have stressed the fact that the angular momentum of a rigid body about an axis remains constant unless acted on by an external torque. In the

TABLE 11-1

Linear Motion	Angular Motion
$x = \bar{v}t$	$\theta = \bar{\omega}t$
$v = u + at$	$\omega = \omega_i + \alpha t$
$x = ut + \frac{1}{2}at^2$	$\theta = \omega_i t + \frac{1}{2}\alpha t^2$
$v^2 = u^2 + 2ax$	$\omega^2 = \omega_i^2 + 2\alpha\theta$
$\bar{v} = \dfrac{u + v}{2}$	$\bar{\omega} = \dfrac{\omega_i + \omega}{2}$
$F = ma$	$G = I\alpha$
$\mathscr{W} = Fs$	$\mathscr{W} = G\theta$
$\mathscr{P} = Fv$	$\mathscr{P} = G\omega$
$\mathscr{E}_k = \frac{1}{2}Mv^2$	$\mathscr{E}_k = \frac{1}{2}I\omega^2$
$p = mv$	$p_\theta = I\omega$
$F = \dfrac{dp}{dt}$	$G = \dfrac{dp_\theta}{dt}$
$J = \Delta p$	$J_\theta = \Delta p_\theta$

x	corresponds to	θ
v		ω
a		α
F		G
M		I
p		p_θ
J		J_θ

previous sections we considered the change in angular momentum about the *same axis* as the original direction of the angular momentum. The direction of the torque vector was parallel to the direction of the angular momentum vector, and the change in the angular momentum vector was always an increase or a decrease in its length. Let us consider now the implications of a change in the *direction* of the angular-momentum vector.

Suppose that a bicycle wheel is mounted on an axle which protrudes beyond the end of the wheel, as shown in Figure 11-13 and that it is spinning in such a direction that its angular-momentum vector is nearly in the direction of the positive x axis. Let us rotate the axle through a positive angle $\Delta\theta$ in the x-y plane, as shown in Figure 11-14. Since the rotation is in a direction perpendicular to the direction of spin, the angular speed ω with which the wheel is spinning on its axle will not be altered, and the angular-momentum vector after the wheel has been rotated will be of the same length as before but will now point in a different direction. The change in the angular momentum $\Delta\mathbf{p}_\theta$ will be a vector directed from the head of the initial angular-momentum vector to the head of the final angu-

lar-momentum vector. Remembering that the torque is equal to the rate of change of the angular momentum,

$$G = \frac{d\mathbf{p}_\theta}{dt},$$

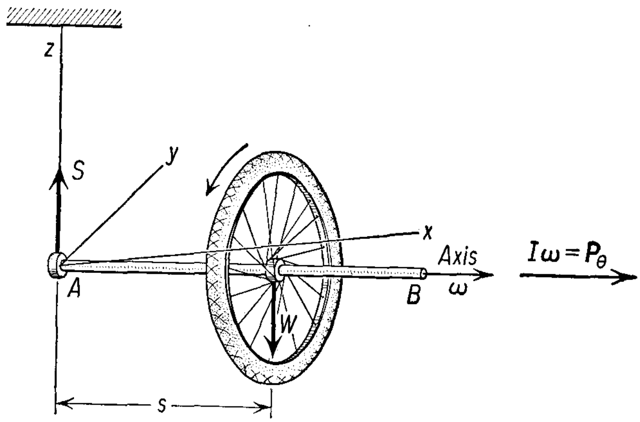


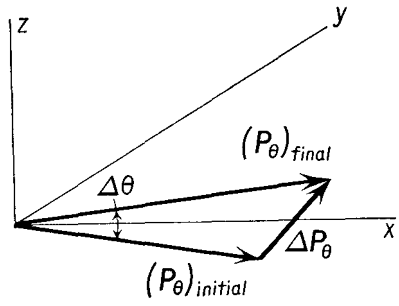
Fig. 11-13 Gyroscope.

which, for small increments in time Δt, may be written as

$$G = \frac{\Delta \mathbf{p}_\theta}{\Delta t}.$$

We see that the torque required to produce the rotation $\Delta\theta$ must be parallel to the change in the angular momentum and must be in the y direction.

Fig. 11-14



In other words, to produce a rotation of the bicycle wheel about the z axis requires that a torque be applied about the y axis. Such a torque would produce a rotation about the y axis *if the body were not spinning.* If the rotation $\Delta\theta$ is produced in a small time interval Δt, at a uniform rate Ω (capital omega), then

$$\Delta\theta = \Omega\,(\Delta t).$$

Referring again to Figure 11-14, for small angular displacement,

$$\Delta\theta = \frac{\Delta p_\theta}{p_\theta},$$

and, substituting from the above equation for $\Delta\theta$, we have

$$p_\theta \Omega\,(\Delta t) = \Delta p_\theta,$$

$$p_\theta \Omega = \frac{\Delta p_\theta}{\Delta t} = G,$$

$$I\omega\Omega = G. \tag{11-12a}$$

Thus the applied torque G is equal in magnitude to the product of the spin angular momentum by the angular speed with which the axis of rotation is itself rotated. A rotation of the spin axis is called *precession*. In this example the direction of spin is parallel to the x axis, the direction of the rotation of the axis is in the z direction, while the direction of the torque which must be applied to produce the rotation is in the y direction. We might write this in terms of the vector product

as
$$\mathbf{G} = \boldsymbol{\Omega} \times \mathbf{p}_\theta,$$

or
$$\mathbf{G} = I\boldsymbol{\Omega} \times \boldsymbol{\omega}, \tag{11-12b}$$

showing that the torque vector is perpendicular to both the spin vector $\boldsymbol{\omega}$ and the precession vector $\boldsymbol{\Omega}$.

Let us re-examine the problem by considering the effect of a torque on the bicycle wheel. Suppose that the left end of the axle of the bicycle wheel is supported by a string which is hung from the ceiling, as shown in Figure 11-13. Once the wheel has been set in motion, spinning about its axle, the forces acting on it are its weight W, acting through the center of gravity, and the tension in the string S. The torque produced by these forces about a horizontal axis through A perpendicular to the plane of the figure (in the positive y direction) is of magnitude Ws, where s is the distance from the center of gravity to A. This torque will produce a change of angular momentum per unit time in accordance with Equation (11-10). The vector representing the change of angular momentum per unit time will be parallel to the torque vector and hence in the positive y direction, directed into the plane of the paper.

Suppose we consider a very small time interval Δt during which this torque acts. The wheel must receive an increment of angular momentum $\Delta \mathbf{p}_\theta = \mathbf{G}(\Delta t)$, and in Figure 11-14 we add this increment of angular momentum to the initial angular momentum to find the final angular momentum of the wheel. We see that in the time Δt the final angular-momentum vector has been rotated by an angle $\Delta\theta$ about the vertical or z axis. The

average angular speed with which the axis has been rotated $\Omega = \Delta\theta/\Delta t$ is the speed of *precession*. The forces S and W continue to act, and the rotational motion of the axle of the bicycle wheel will continue in a horizontal plane. Note that our first impression would be to assert that the wheel would tend to be twisted by the forces W and S so that the axle should tend to hang vertically. The rotating wheel is sometimes called a *gyroscope*, and the motion analyzed above is called *gyroscopic motion*.

Any rotating body can be considered as a gyroscope. When a torque acts to change the direction of spin, precessional motion will occur. The earth's axis of rotation precesses in the heavens in a circle of 23.5° radius with a period of 26,000 years. The torques which cause the axis of rotation to precess are due primarily to the gravitational attraction of the sun and moon on the earth's equatorial bulge. The motion of a top whose axis of spin is inclined to the vertical is a common example of precessional motion. When the top is not spinning, it is in unstable equilibrium when resting on its point. When the top is spinning, its most stable position is one in which its center of gravity is directly above the point of support, and when a spinning top is thrown, it climbs to this position in apparent disregard for equilibrium conditions; this is owing to the friction between the peg of the top and the surface. The propeller of an airplane acts like a gyroscope, and when the airplane turns the gyroscope will precess, unless the airplane has its propellers in pairs which rotate in opposite directions, so that the total angular momentum due to spin is zero.

Whenever any piece of rotating machinery is mounted on a moving platform, such as a ship or an airplane, the bearings of that machine must exert a torque on the shaft of the machine so that it will precess in the direction in which the platform is turning. The greater the rotational speed and the spin angular momentum, the greater is the torque required, according to Equations (11-12). For this reason, motors mounted on board ships or aircraft must have specially designed bearings, capable of withstanding far greater loads than would be required of the same appliance if the machine were used on a stationary platform.

If a gyroscope is mounted on earth so that its axis is parallel to the axis of rotation of the earth, that is, in a north-south direction, the rotating earth does not change the direction of the axis of rotation of the gyroscope in space, and there is no tendency for the gyroscope to precess. If the gyroscope is mounted with its axis in some other direction, the rotation of the earth will cause the gyroscope to precess, so that the direction of the true north can be recognized from the behavior of a rotating gyroscope. This is the basis of the gyrocompass.

Gyroscopic motion is the basis of the behavior of the bicycle. If a rider leans to the left, the front wheel of a bicycle will turn to the left as though to catch the rider in his fall. The discussion based on Figure 11-13

is suitable to a discussion of the behavior of a bicycle advancing in the $-y$ direction in which the rider leans to his left. From Figure 11-14 we see that the front wheel will turn into the rider's fall. A novice learning to ride must learn to let the bicycle do his thinking for him, while the skilled cyclist can ride without touching the handle bars by shifting his weight from side to side.

Problems

11-1. The flywheel of a gasoline engine is built in the form of a uniform disk of radius 1 ft and weighs 75 lb. The flywheel is rotating with an angular speed of 3,300 rpm. Determine the kinetic energy of the flywheel.

11-2. A small copper disk of 15 cm radius and 350 gm mass is rotating with an angular speed of 12 radians/sec about an axis through its center. Determine (a) the kinetic energy of the disk and (b) its angular momentum.

11-3. A pulley 6 in. in diameter is mounted so that it can rotate about a fixed axis through its center. The pulley weighs 12 lb and has a moment of inertia of 0.02 slug ft^2. A constant force of 3 lb is applied to the rim of the pulley by means of a cord wrapped around it. Determine (a) the angular acceleration of the pulley and (b) the angular speed it has at the end of 10 sec, assuming that the pulley was initially at rest.

11-4. Find the moment of inertia of a dumbbell consisting of two spheres of radius 10 cm connected by a cylindrical rod 1 cm in radius and 50 cm long about an axis through the center of gravity perpendicular to the rod. The dumbbell is made of iron of density 7.8 gm/cm^3.

11-5. Prove the parallel axis theorem for the case of two equal point masses.

11-6. A wheel in the form of a uniform disk of mass 900 gm and radius 8 cm is mounted so that it can rotate about a fixed horizontal axis passing through its center. A cord is wrapped around the circumference of the wheel, and a mass of 50 gm is attached to its free end. (a) Determine the angular acceleration of the wheel when it is released. (b) Determine the linear acceleration of the 50-gm mass. (c) Determine the tension in the cord. (d) Determine the angular velocity of the wheel at the end of 5 sec. (e) Determine the kinetic energy of the entire system when the mass has fallen through a distance of 10 cm.

11-7. A wheel having a radius of 6 cm is mounted so that it can rotate about a fixed horizontal axis passing through its center. A cord wrapped around the circumference of the wheel has a mass of 250 gm attached to its free end. When allowed to fall, the mass takes 5 sec to fall a distance of 100 cm. Determine (a) the angular acceleration of the wheel and (b) its moment of inertia.

11-8. A rod 50 cm long and weighing 5 lb is pivoted at one end. The rod is raised to a horizontal position and released. (a) What is the angular velocity of the rod when it is at an angle of 45° with the horizontal. (b) What is the angular acceleration of the rod at the same deflection?

11-9. A solid cylinder 2 ft in diameter and weighing 64 lb starts at the top of a rough plane 24 ft long and inclined at an angle of 30° with the horizontal and rolls down without slipping. (a) How much energy did the cylinder have at the

top of the hill? (b) How much energy will it have at the bottom of the hill? (c) Determine its angular velocity at the bottom of the hill.

11-10. Show that Kepler's law of areas (Section 6-14) is consistent with the law of conservation of angular momentum, for circular orbits.

11-11. A steel hoop rolls without sliding down a plane inclined at an angle of 30° with the horizontal. The mass of the hoop is 600 gm, and its radius is 8 cm. Determine (a) its moment of inertia about its instantaneous axis of rotation, (b) its angular acceleration, (c) the force of friction between the hoop and the plane, and (d) the linear velocity of the center of the hoop when the hoop has rolled 1 m down the incline.

11-12. A gasoline engine develops 75 hp when turning at 3,300 rpm. Determine the torque delivered by this engine to the drive shaft.

11-13. A solid spherical ball of radius 1 ft is rolled toward a curb of height 4 in. What must be the speed of the center of gravity of the ball if the ball is to jump the curb?

11-14. A boy weighing 100 lb stands at the center of a brass turntable 6 in. thick and 10 ft in diameter. The turntable is rotated with an angular speed of 1 radian/sec. The boy walks out along a radius to a point 5 ft from the center of the turntable. What is the angular speed with which the disk is now rotating? The density of brass is 8.6 gm/cm^3.

11-15. A solid disk having a mass of 1 kg and a radius of 2 cm is wrapped with string. The free end of the string is supported from a point in the ceiling, and the disk is released. Determine the angular speed of the disk when the center of gravity of the disk has fallen 1 m.

11-16. A top having a moment of inertia of 5,000 gm cm^2 is spinning at a speed of 25 rev/sec at an angle of 30° with the vertical. The top has a mass of 500 gm, and its center of gravity is 4 cm from its point. The spin is counterclockwise, as seen from above. (a) What is the angular velocity of precession of the top axis? (b) As seen from above, is it clockwise or counterclockwise?

11-17. A unicycle has a wheel 36 in. in diameter and a mass of 1 slug. Neglect the weight of the seat and frame. A man weighing 150 lb sits on the unicycle, so that his center of gravity is 4 ft from the bottom of the wheel, and pedals the unicycle until it is moving with a speed of 20 ft/sec. The man leans to his right so that man and cycle make an angle of 1° with the vertical. What is now the direction and angular speed of precession of the axle of the unicycle?

11-18. A satellite of mass 10 kg is launched at the equator due north with a velocity of 10^6 m/hr. The satellite is to fly along the surface of a sphere of radius 6.5×10^6 m concentric with the earth. (The earth's radius is approximately 6.4×10^6 m.) When the satellite reaches 37° north latitude, what are the components of its velocity with respect to the earth (a) in the north-south direction, and (b) in the east-west direction? Assume that the satellite experiences no drag. Take into account the necessity to conserve angular momentum.

11-19. When a car is going forward, the engine and flywheel are rotating counterclockwise as viewed by the driver. In which direction will the car tend to go if the front wheels are suddenly lifted by a bump in the road?

11-20. Discuss the motion of an airplane whose propellers suddenly stop rotating from the point of view of the conservation of angular momentum.

11-21. A man sits on a piano stool holding a spinning bicycle wheel in his hands. The axis of the bicycle wheel is directed vertically. He turns the wheel end for end while remaining seated, and finds that he begins to rotate in the same direction that the wheel was spinning originally. Why? What happens if he once again reverses the wheel?

11-22. A man sits on a piano stool which is at rest. The man and stool have a weight of 200 lb, and may be approximated by a vertical cylinder of radius 1 ft. The man catches a 5-oz baseball moving horizontally with a speed of 80 ft/sec at a distance of 1 ft from the axis of rotation of the stool. (a) What is the angular momentum of the baseball with respect to the axis of the stool at the instant it is caught? (b) What is the angular velocity of the system consisting of the man, stool, and ball after the ball is caught?

11-23. Show that the moment of inertia of a thin rectangular sheet of sides a and b is $I = \frac{1}{12} Ma^2$, when the axis of rotation lies in the plane of the sheet, through the center of mass, and is parallel to the side b, as in Figure 11-6. [HINT: Start with the expression for the moment of inertia of a rod and integrate.]

11-24. Find the moment of inertia of a thin flat sheet cut in the form of the quadrant of a circle of radius R with respect to an axis normal to the plane of the quadrant and passing through the center of the circle.

11-25. A 2-kg mass is suspended from a string which is wound over the axle of a wheel. It is observed that the mass has a downward acceleration of 2 m/sec^2. The radius of the axle is 0.05 m and the radius of the wheel is 1.5 m. At $t = 0$ the system is at rest. (a) What is the angular acceleration of the wheel? (b) What is the angular velocity of the wheel at $t = 5$ sec? (c) What is the radial acceleration of a point on the rim of the wheel at $t = 5$ sec? (d) What is the moment of inertia of the wheel and axle?

11-26. A solid sphere of radius 10 cm and mass 250 gms rolls without slipping down an inclined plane which makes an angle of 37° with the horizontal. (a) What is the linear acceleration of the center of mass of the sphere parallel to the plane? (b) What is the angular acceleration of the sphere about an axis through its center of mass? (c) What is the angular velocity of the sphere about an axis through its center of mass when the sphere has rolled a distance of 5 m down the incline (measured along the incline)?

11-27. An electric motor which turns at a speed of 3600 rpm has an armature of mass 10 kg. The armature may be approximated by a solid cylinder of radius 5 cm. The motor is mounted on an airplane which turns to the right through a 90° arc in 15 sec. Assume that the airplane turns without banking. What is the magnitude and direction of the torque exerted on the spinning armature by the bearings if (a) the axis of the motor is mounted vertically? (b) the motor is mounted horizontally with its axis parallel to the wings? (c) the motor is horizontal with its axis perpendicular to the wings?

12

Periodic Motion

12-1 Introduction

One of the more important problems in mechanics is the study of periodic motions, that is, *motions which repeat themselves in regular intervals of time, called the period.* An example of periodic motion which we have already encountered is uniform circular motion, in which the velocity and acceleration of the body at a given angular position were always the same. If a particle was found at a given position at a time t, we could be sure that it would return to that position at time $t + T$ later, where T was the period of the rotational motion. A body in uniform circular motion moves under the influence of a varying force, the centripetal force, which, though constant in magnitude, varies in direction. In general, an object moving in periodic motion must move under the influence of a varying force which is directed to some equilibrium position or to some neutral position. In uniform circular motion this position is the center of the circle.

There are many other types of periodic motion. Among these are the *oscillations* of a weight attached to a vertical spring, the motion of a simple pendulum, and the oscillations of the balance wheel of a watch. A glass marble bouncing up and down on a hard steel anvil represents such a motion; see Figure 12-1(a). If the collision between the marble and the anvil is perfectly elastic, the marble rises after each collision to the height h_0 from which it was dropped. The motion is perfectly repetitive, as shown in Figure 12-1(b), where the height h of the marble has been plotted as a function of time. We see that the graph is really a succession of parabolas displaced along the time axis by the period T. In actual practice it is not possible to achieve a perfectly elastic collision between the marble and the anvil. The marble loses a fraction of its energy on each collision, and the motion is not perfectly periodic, as shown by the dotted lines in Figure 12-1(b). The interval between successive impacts becomes shorter and shorter, and the motion is known as an *aperiodic* motion, that is, nearly but not quite periodic.

Other examples of oscillating motion include the movement of a piston in a gasoline engine, the vibrations of a string of a musical instrument, of the wingtips of an airplane in flight, and of a building or a suspension bridge

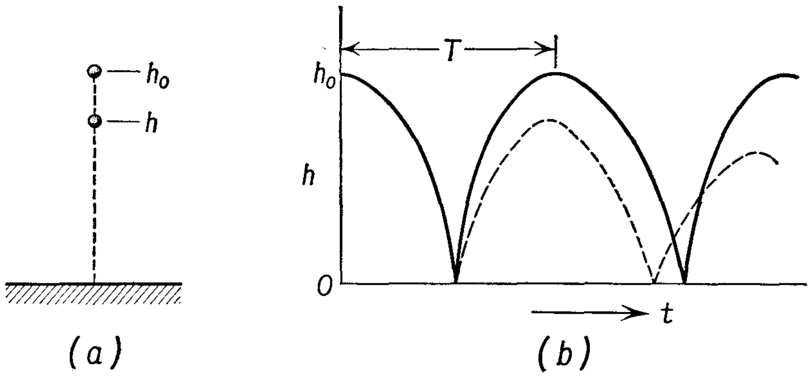


Fig. 12-1 (a) Ball bouncing off an anvil. (b) Periodic motion (solid line) of a perfectly elastic marble bouncing on a rigid anvil, and the aperiodic motion (dotted line) of a glass marble on a steel anvil.

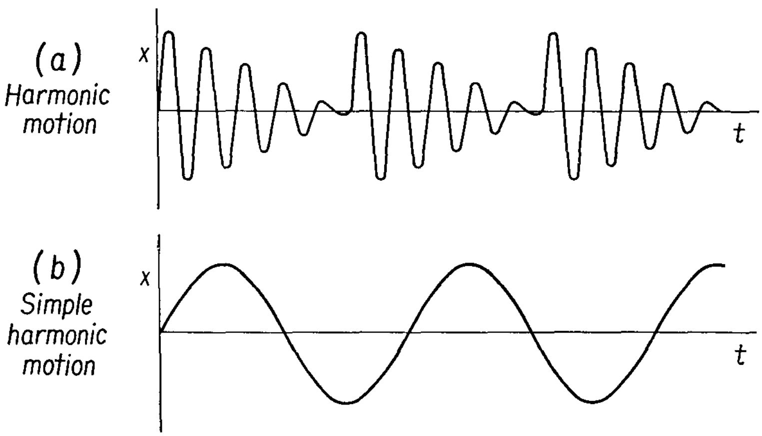


Fig. 12-2 Harmonic motion. (a) Graph of a series of damped oscillations. (b) Graph of simple harmonic motion.

in a high wind, and the bobbing of a ship or of a bell buoy. Many of these motions can be described by the use of a combination of sine or cosine functions of time and are consequently called *harmonic* motions. The simplest of these motions follows a single sine or cosine curve and is called *simple harmonic motion.* Examples of harmonic motion and simple harmonic motion are shown in Figure 12-2.

12-2 Elasticity

In an earlier discussion of collision problems, we used the word *elastic* to describe a process in which mechanical energy was conserved. In the present section the word "elastic" will be used to describe a different but related property of matter. When a force is applied to a solid body made of any one of a great variety of materials, the shape of the body is altered in a nonpermanent way. When the force is removed, the body returns to its original shape, as in the example of a coil spring. In general, the body deforms until the restoring force exerted by the body is equal and opposite to the applied force. Experiments conducted by Robert Hooke (1635–1703) showed that the nonpermanent deformation of many bodies was directly proportional to the force which created the deformation. The name *Hooke's law* is applied to all cases where *the restoring force exerted by a deformed object is proportional to the deformation.* Many materials follow Hooke's law over a limited range of deformation, and one finds that, over this range of applicability of Hooke's law, an object returns to its initial configuration without permanent deformation when the applied force is released. Such materials are called *elastic.* In neither sense of the uses to which we have put the word "elastic" is rubber a perfectly elastic material, for the deformation of a rubber band is not proportional to the applied force, nor is the impact between a rubber ball and an anvil a collision in which mechanical energy is conserved.

We have already seen that a helical spring stretches in proportion to the applied force, but we need not restrict ourselves to linear motion. In a clock spring, or a rod clamped at one end, the angular displacement is proportional to the applied torque tending to twist the rod or the spring.

To describe the elastic restoring force exerted by a stretched spring or by a beam which is deflected from its equilibrium position, we may write

$$\boxed{\mathbf{F} = -k\mathbf{x},} \tag{12-1}$$

which states that the force **F** exerted by the spring or the beam is proportional to the displacement **x** and is in a direction opposite to that displacement, hence the minus sign. The spring constant k is a constant of proportionality which has the dimensions of force per unit length.

To describe the elastic restoring torque of a clock spring or a twisted shaft, we may write

$$\boxed{\mathbf{G} = -\kappa\boldsymbol{\theta},} \tag{12-2}$$

where **G** is the restoring torque exerted by the spring or shaft when it has been twisted from its equilibrium position through an angle **θ**. We represent **θ** as a vector quantity in this case, for the axis of rotation is fixed in space. The torque exerted by the spring is opposite in direction to the displacement. The constant κ (kappa) is a constant of proportionality which has the dimensions of torque per unit angular displacement.

When an object obeying Hooke's law, as represented in Equations (12-1) and (12-2), *is displaced from its equilibrium position and released, the subsequent motion is simple harmonic.* An elastic system in which the restoring force is directly proportional to the displacement is said to obey Hooke's law. The forces exerted by the materials from which engineering structures are constructed are elastic in character. The forces exerted by adjacent atoms in a molecule may be approximated by Hooke's law. The study of simple harmonic motion is thus of considerable importance in engineering as well as in the physics of atoms and molecules.

12-3 Equations of Simple Harmonic Motion

Let us consider simple harmonic motion along the x axis. In the previous paragraph we have indicated that motion in which the restoring force was proportional to the displacement was simple harmonic; that is, the motion could be described in terms of sine or cosine functions. If a particle of mass m is subject to an elastic restoring force, we may write, from Newton's second law and from Equation (12-1),

$$ma = -kx,$$

and, writing for a its value using the notation of the calculus, $a = \dfrac{d^2x}{dt^2}$,

we have

$$m\frac{d^2x}{dt^2} = -kx.$$

Transposing, we find

$$m\frac{d^2x}{dt^2} + kx = 0.$$

Let us divide the equation by m and set

$$\frac{k}{m} = \omega^2, \tag{12-3}$$

so that the equation becomes

$$\frac{d^2x}{dt^2} + \omega^2 x = 0. \tag{12-4}$$

Equation (12-4) is a second-order differential equation, for it involves not only the variable x but also its second derivative with respect to the time, $\frac{d^2x}{dt^2}$. Such equations are commonly encountered in physics and engineering, and their solutions are often quite complex. For Equation (12-4) the solution is straightforward. First we observe that the equation contains a second derivative and will have to be integrated twice to find a solution for x as a function of time. The solution will therefore contain two constants of integration which will have to be evaluated in terms of the initial conditions of the problem. Knowing these facts in advance, we propose as a trial solution

$$x = A \cos(\omega t + \phi). \tag{12-5}$$

Remembering that $d(\cos x) = -\sin x\, dx$, and that $d(\sin x) = \cos x\, dx$, we find that

$$v = \frac{dx}{dt} = -A\omega \sin(\omega t + \phi) \tag{12-6}$$

and,

$$a = \frac{dv}{dt} = \frac{d^2x}{dt^2} = -A\omega^2 \cos(\omega t + \phi), \tag{12-7}$$

from which

$$a = -\omega^2 x. \tag{12-8}$$

The constants A and ϕ are the two constants of integration. The constant A is the *amplitude* of the simple harmonic motion; that is, it represents the largest value the displacement x can attain. The constant ϕ is called the *phase angle;* that is, the value of x when t is zero is given by $A \cos \phi$.

On substituting the values of x and $\frac{d^2x}{dt^2}$ from Equations (12-5) and (12-7) into the Equation (12-4), we find

$$-A\omega^2 \cos(\omega t + \phi) + \omega^2 A \cos(\omega t + \phi) = 0,$$

$$0 = 0,$$

so that Equation (12-5) is a correct solution to the differential Equation (12-4).

The period T for one complete oscillation is given by

$$T = \frac{2\pi}{\omega}, \tag{12-9}$$

for if the time is increased by T, the angle at which the cosine is to be evaluated is increased from $[\omega t + \phi]$ to $\left[\omega\left(t + \frac{2\pi}{\omega}\right) + \phi\right]$, an increase of 2π, which is the angular interval in which the sine or cosine repeats itself.

The number of oscillations per second is called the *frequency* f, which is given by

$$f = \frac{1}{T}, \tag{12-10}$$

and, substituting from Equation (12-9), we find

$$\omega = 2\pi f, \tag{12-11}$$

which is exactly the same as the relationship between angular velocity and the frequency for uniform circular motion.

From Equation (12-8) we note that the acceleration of a particle in simple harmonic motion is proportional to the displacement but is always in the opposite direction.

Illustrative Example. A particle attached to a spring has a frequency of 4 vibrations per second (abbreviated vib/sec) and an amplitude of 6 cm. Determine (a) the period of the vibration, (b) the maximum velocity of the particle, and (c) the maximum acceleration of the particle.

(a) From Equation (12-10) we have

$$T = \frac{1}{f} = \frac{1}{4\text{ vib/sec}}$$
$$= \frac{1}{4}\,\frac{\text{sec}}{\text{vib}},$$

and since vibration is a dimensionless quantity, a pure number,

$$T = \tfrac{1}{4}\text{ sec.}$$

(b) The maximum velocity may be obtained from Equation (12-6) by observing that the largest numerical value of v occurs when $\cos(\omega t + \phi) = 1$. Thus we have

$$v_{\max} = A\omega.$$

From Equation (12-11)

$$\omega = 2\pi f = 8\pi\,\frac{\text{radians}}{\text{sec}}.$$

Thus, substituting for A its value of 6 cm, we find

$$v_{\max} = 6\text{ cm} \times 8\pi\text{ sec}^{-1}$$
$$= 48\pi\,\frac{\text{cm}}{\text{sec}}.$$

(c) The greatest numerical value of the acceleration is

$$a_{\max} = A\omega^2$$
$$= 6\text{ cm} \times 64\pi^2\text{ sec}^{-2}$$
$$= 384\pi^2\,\frac{\text{cm}}{\text{sec}^2}.$$

If, instead of the cosine function a sine function had been used as a trial solution, we would have found that the sine function would have proved equally satisfactory. Formally there is no essential difference between a sine-function solution and a cosine-function solution; one can be changed into the other by simply adding $+90°$ or $-90°$ to the angle, for

$$\sin(\theta + 90°) = \cos\theta,$$

and

$$\cos(\theta - 90°) = \sin\theta.$$

While a description of simple harmonic motion has been achieved in the present section through the use of the calculus, another way to gain insight into simple harmonic motion is to compare it to uniform motion in a circle, called the *reference circle.*

12-4 The Reference Circle

When an object moves in uniform circular motion, its projection onto the x or y axis moves in simple harmonic motion. The projected position corresponds to the position of a particle moving in simple harmonic motion; the projected acceleration vector corresponds to the acceleration of a particle moving in simple harmonic motion.

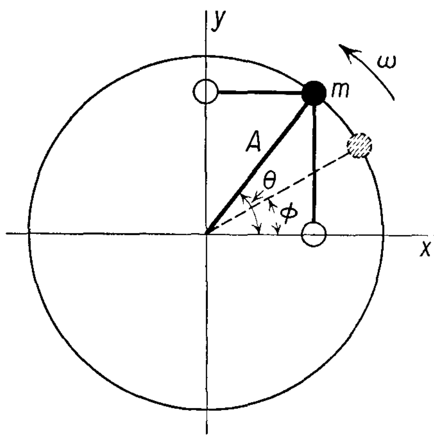


Fig. 12-3 The reference circle. A is the radius of the circle.

Referring to Figure 12-3, let us suppose that at time $t = 0$, an object in uniform circular motion (solid circle) is located at a position on the circle of radius A given by the angle ϕ, and that it rotates in the counterclockwise direction with uniform angular speed ω. The angular position of the object at a subsequent time t is given by the angle θ such that $\theta = \omega t + \phi$. The x coordinate of the object is

$$x = A\cos(\omega t + \phi), \tag{12-12}$$

while the y coordinate is given by

$$y = A\sin(\omega t + \phi), \tag{12-13}$$

but these are also the coordinates of the projection of the circular motion onto the x and y axes, respectively.

Comparing Equations (12-12) and (12-5), we see that the projection of the motion in the reference circle onto the x axis is precisely the same as the

simple harmonic motion, provided that the radius of the reference circle is made equal to the amplitude of the simple harmonic motion, and the rotational speed of the object in the reference circle is made equal to the *angular frequency* ω of the simple harmonic motion. The reference-circle analogue clarifies the meaning of the phase angle ϕ in the simple harmonic motion.

We see also that either a sine function or a cosine function is suitable for the description of simple harmonic motion. The projected motion along the y axis is of the same frequency and amplitude as that along the x axis,

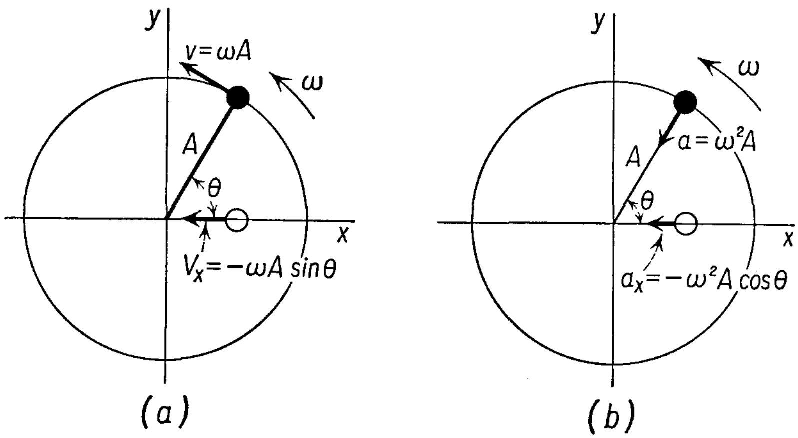


Fig. 12-4 Use of the reference circle to determine (a) the velocity of the particle (white) in simple harmonic motion, and (b) the acceleration of this particle.

but the two motions are 90° out of phase with each other. When the x projection is at the origin, the y projection has its maximum value A.

To make the case for the reference circle complete, let us find the velocity and acceleration of the projected particle, shown as a white circle in Figure 12-4. The black particle moving in uniform circular motion with angular speed ω has a linear speed $v = \omega A$ directed tangentially. The projection of this vector onto the x axis is the velocity of the white projected particle moving in simple harmonic motion in the x direction. From the figure

$$v_x = -\omega A \sin\theta = -A\omega \sin(\omega t + \phi),$$

which is identical with Equation (12-6). An object moving in uniform circular motion must experience a centripetal acceleration $a = \omega^2 A$. From Figure 12-4(b) the acceleration of the projected (white) particle is the projection of the acceleration of the real (black) particle onto the x axis.

Thus we have

$$a_x = -\omega^2 A \cos\theta = -A\omega^2 \cos(\omega t + \phi),$$

which is identical with Equation (12-7).

Remembering that $A \cos\theta = A \cos(\omega t + \phi) = x$, and substituting in the above equation, we find that the acceleration of the white particle on the x axis is

$$a = -\omega^2 x, \tag{12-14}$$

which is identical with Equation (12-8).

The basic condition for simple harmonic motion was that the restoring force was proportional to the displacement, or that $F = -kx$. Substituting for F from Newton's second law, we found that for a particle of mass m subject to an elastic restoring force,

$$ma = -kx,$$

or

$$a = -\frac{k}{m}x, \tag{12-15}$$

and again we see that the acceleration of the projected particle on the x axis, given by Equation (12-14), is identical with that experienced by a real particle which is subject to an elastic restoring force as given in Equation (12-15), provided that,

$$\omega^2 = \frac{k}{m},$$

which is identical with Equation (12-3). The device of the reference circle is a very useful method for solving problems in simple harmonic motion at the level of this text. In using the reference circle, it must be remembered that the radius of the reference circle is equal to the amplitude of the simple harmonic motion it is chosen to represent; the frequency of the rotational motion in the reference circle must be equal to the frequency of the simple harmonic motion; and the phase angle ϕ must be chosen so that the projected particle is at the proper location at time $t = 0$.

Illustrative Example. A particle attached to a spring has a frequency of 4 vib/sec and an amplitude of 6 cm. Determine (a) the period of the vibration, (b) the maximum velocity of the particle, (c) the velocity of the particle when its displacement is 2 cm, (d) the acceleration of the particle when its displacement is 2 cm, (e) the maximum acceleration of the particle, and (f) the time required by the particle to move from a displacement of +2 cm to a displacement of +4 cm.

(a) The period of the vibration is the reciprocal of the frequency of vibration. Hence

$$T = \frac{1}{f} = \frac{1}{4}\ \text{sec}.$$

To solve the remainder of the problem, we make use of the reference circle. Let us suppose that the simple harmonic motion takes place along the x axis, and let us construct a reference circle of radius equal to the amplitude of the simple harmonic motion; thus the radius of the reference circle is 6 cm. The real particle in simple harmonic motion will be referred to as the white particle, while

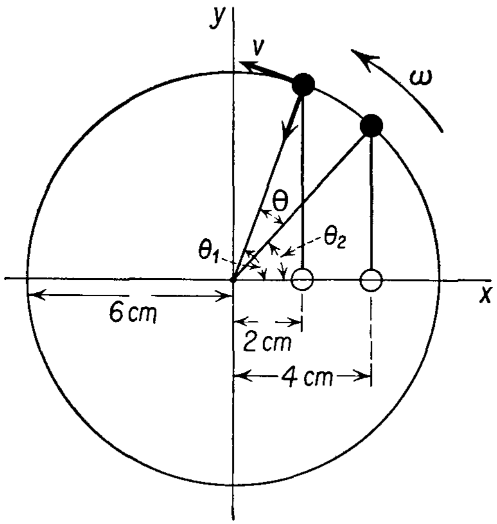


Fig. 12-5

the imagined particle moving in the reference circle will be referred to as the black particle, corresponding to the manner in which they have been drawn in Figure 12-5. The frequency with which both particles move is given by the frequency of the simple harmonic motion which is 4 vib/sec. This implies that the angular speed of the black particle is $\omega = 2\pi f = 8\pi$ radians/sec, which, for convenience, we imagine to be in the counterclockwise direction.

(b) The black particle moves with constant angular speed and therefore with constant linear speed, but its velocity vector is constantly changing in direction. Since the speed of the white particle is the projection onto the x axis of the velocity vector of the black particle, the white particle will move with greatest speed when the black particle is moving parallel to the x axis. This occurs at the two positions where the circle intersects the y axis, that is, when the white particle passes through its equilibrium position. At this position the speeds of the two particles are identical, so that the maximum speed of the white particle is equal to the constant linear speed of the black particle. Thus

$$v = \omega A = 8\pi \times 6 \frac{\text{cm}}{\text{sec}} = 48\pi \frac{\text{cm}}{\text{sec}},$$

$$v = 151 \frac{\text{cm}}{\text{sec}}.$$

(c) When the white particle is displaced 2 cm from the equilibrium position,

the position angle of the black particle is θ_1 such that

$$\theta_1 = \text{arc}\cos\left(\tfrac{2}{6}\right) = 1.23 \text{ radians}.$$

To find the velocity of the white particle, we must find the x component of the velocity of the black particle, which is given by

$$\begin{aligned} v &= v \sin\theta_1 \\ &= 1.51 \times 0.942 \,\frac{\text{cm}}{\text{sec}} \\ &= 142 \,\frac{\text{cm}}{\text{sec}}. \end{aligned}$$

(d) The acceleration of the white particle is the x component of the acceleration of the black particle and is therefore given by

$$a_x = -\,a\cos\theta_1,$$

where the minus sign indicates the direction of the acceleration vector. For the black particle in circular motion,

$$\begin{aligned} a &= \omega^2 A = (8\pi)^2 \times 6 \,\frac{\text{cm}}{\text{sec}^2} \\ &= 3{,}789 \,\frac{\text{cm}}{\text{sec}^2}, \end{aligned}$$

so that

$$\begin{aligned} a_x &= -3{,}789 \times \frac{2}{6}\,\frac{\text{cm}}{\text{sec}^2} \\ &= -1{,}263 \,\frac{\text{cm}}{\text{sec}^2}. \end{aligned}$$

(e) The maximum acceleration of the white particle will occur when the centripetal acceleration of the black particle is in the direction of the x axis; that is, where the circle intersects the x axis. At these points the acceleration of the two particles is identical. Thus the maximum acceleration of the white particle in simple harmonic motion is

$$a = 3{,}789 \,\frac{\text{cm}}{\text{sec}^2},$$

occurring at the position of maximum displacement.

(f) When the white particle is at a displacement of 2 cm, we have already seen that the black particle is located at angle θ_1. When the displacement of the white particle is 4 cm, the black particle is located at angle θ_2. The angular displacement of the black particle in rotating from θ_2 to θ_1 is Θ (capital theta).

$$\Theta = \theta_1 - \theta_2.$$

We have already seen that $\theta_1 = 1.23$ radians, while

$$\theta_2 = \text{arc}\cos\left(\tfrac{4}{6}\right) = 0.84 \text{ radian}.$$

Thus

$$\begin{aligned} \Theta &= 1.23 \text{ radians} - 0.84 \text{ radian} \\ &= 0.39 \text{ radian}. \end{aligned}$$

Knowing that ω, the angular speed of the black particle, is 8π radians/sec, we find the time for the black particle to travel from the first to the second position as

$$t = \frac{\Theta}{\omega}$$

$$= \frac{0.39 \text{ radian}}{8\pi \text{ radians/sec}},$$

so that

$$t = 0.016 \text{ sec.}$$

This problem may also be solved by the substitution of appropriate numbers for A, ω, and ϕ in Equations (12-5), (12-6), and (12-7), as in a previous example. From the statement of the problem, $A = 6$ cm, $\omega = 2\pi f = 8\pi \text{ sec}^{-1}$, and if we wish to have the initial position of the particle at the undisplaced or equilibrium position when $t = 0$, we would set $\phi = \pi/2$ radians. The maximum values of the velocity and acceleration are given by $v = A\omega$ and $a = A\omega^2$. The value of the velocity and acceleration when the displacement was 2 cm could be found by substituting $x = 2$ cm in Equation (12-5), solving for t, and substituting that value of the t in Equations (12-6) and (12-7). To find the time at which the displacement is 4 cm, we set $x = 4$ cm in Equation (12-5), and to find the elapsed time in traveling between a displacement of 2 cm and a displacement of 4 cm, we would subtract the two times. The geometric procedure using the reference circle and the procedure involving the use of the formulas derived by mathematical analysis yield equivalent results. Note that in the analytic treatment and in the reference circle the particle in simple harmonic motion could be thought to move along either the x or the y axis, at our convenience.

12-5 Angular Harmonic Motion

A system capable of rotating about a fixed axis will move with angular harmonic motion when the torque G which acts on it is proportional to its angular displacement θ and opposite in direction to it, as given by

$$G = -\kappa\theta. \tag{12-2}$$

If I is its moment of inertia about this axis, we can write

$$G = I\alpha,$$

where α is its angular acceleration. Solving the above equations for α, we get

$$\alpha = -\frac{\kappa}{I}\theta, \tag{12-16}$$

which shows that the angular acceleration α is proportional to the angular displacement and opposite in direction to it. Equation (12-16) can be compared with Equation (12-8) for linear simple harmonic motion; Equation (12-16) is its analogue and is the equation for *angular harmonic motion.*

The period for angular harmonic motion is then given by

$$T = 2\pi \sqrt{\frac{I}{\kappa}} \tag{12-17}$$

for the period of a system of moment of inertia I and restoring constant κ moving with angular harmonic motion about a fixed axis.

The equations for the angular displacement and the angular velocity as functions of the time can be obtained in exactly the same manner as those for the linear displacement and the linear velocity, and will be their respective analogues.

12-6 The Pendulum

The motion of a pendulum is another example of periodic motion. When the amplitude of oscillation of a pendulum is small, the pendulum motion may be approximated as simple harmonic. This property was discovered by Galileo, and was first applied to the construction of a clock by the Dutch physicist Christian Huygens (1629–1695) in 1657.

The pendulum appears in two forms—the *simple pendulum* consisting of a string of negligible weight, one end of which is attached to some fixed support while the other end is attached to a small ball called a *pendulum bob*, and the *physical pendulum*, a rigid body which is supported at some point above its center of gravity. When at rest, the bob is vertically beneath the point of support, as shown in Figure 12-6, and is in equilibrium under the action of two forces, its weight mg and the tension S in the string. When pulled aside to some position A and released, it travels in a circular arc through its equilibrium position C to a point B on the other side. When the pendulum was moved to A, it was actually lifted through a height h. From the principle of conservation of energy, the points A and B must be at equal heights h above the point C. In the absence of frictional forces, the motion would continue indefinitely, but of course, no device can be built which completely eliminates frictional forces, hence the amplitude gradually diminishes as mechanical energy is converted to other forms of energy.

To derive an expression for the period of the simple pendulum, we observe that the torque on the simple pendulum which tends to rotate it about the point of support is due to the weight of the pendulum bob and is given by

$$G = -mgL \sin \theta. \tag{12-18}$$

Comparing the torque on a simple pendulum with the condition for an elastic restoring torque in Equation (12-2), we see that the torque acting on a simple pendulum is proportional not to the angular displacement but to the sine of the angular displacement, so that the motion is not simple

harmonic. But if the displacement is restricted to small angles, the sine of θ is practically equal to θ, and we may write

$$G = -mgL\theta. \tag{12-19}$$

For small displacements from the equilibrium position, the pendulum moves under the influence of restoring torque proportional to the angular dis-

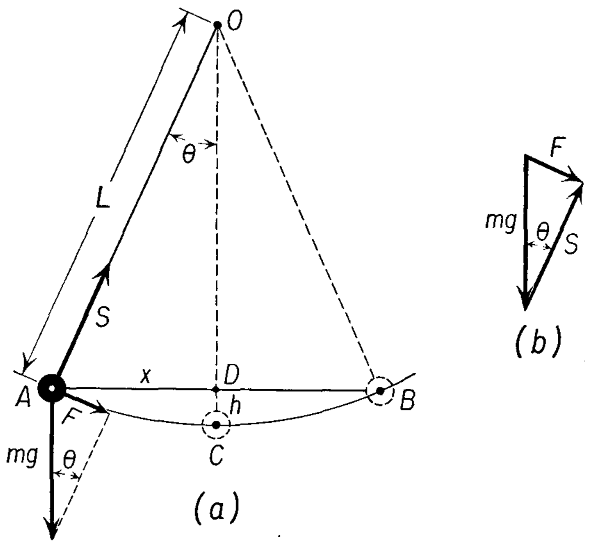


Fig. 12-6 The simple pendulum.

placement, and its motion is simple harmonic. Considering the pendulum as a rigid body, of moment of inertia $I = mL^2$, we may write

$$G = I\alpha,$$

and, eliminating G from the above equations, we find

$$\alpha = -\frac{mgL}{I}\theta. \tag{12-20}$$

By combining Equation (12-20) with Equation (12-16), we can see that the period is

$$T = 2\pi\sqrt{\frac{I}{mgL}}, \tag{12-21}$$

and, substituting for I its value mL^2, we find that

$$T = 2\pi\sqrt{\frac{L}{g}}. \tag{12-22}$$

Thus the period of a simple pendulum does not depend upon the mass of the pendulum bob but only upon its length and the value of g. The simple pendulum therefore provides one method of determining the value of g.

In Figure 12-7 a physical pendulum has been drawn in which the point O is the point of support and the point C is the position of the center of gravity. The distance between C and O is represented by h. If the mass of the physical pendulum is m and its moment of inertia about an axis through O is I, the equations for the torque and the angular acceleration are identical with those for the simple pendulum except that h replaces L. The period of a physical pendulum is then given by

$$T = 2\pi\sqrt{\frac{I}{mgh}}\cdot \tag{12-23}$$

Comparing Equations (12-22) and (12-23), we find that a physical pendulum will have the same period as a simple pendulum of length L when

$$L = \frac{I}{mh}\cdot \tag{12-24}$$

L is called the *equivalent simple pendulum length* of the physical pendulum and is shown in Figure 12-7 as the distance between O and a point O' below the center of gravity. The point O' is called the *center of oscillation* and has the interesting property that, if the pendulum is suspended about O' as an axis, its period will be exactly the same as it is when suspended about O as an axis. An accurate determination of the locations of the two points O and O' makes possible an accurate determination of the equivalent simple pendulum length L and is one of the best methods for determining g.

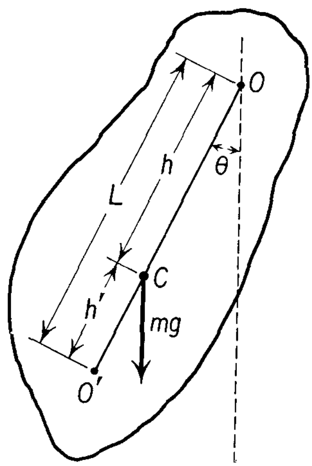


Fig. 12-7 The physical pendulum. The distances h and h' are measured from C to O and O' respectively. $L = h + h'$ is the equivalent simple pendulum length.

The physical pendulum is often used for pendulum clocks. In order for such a pendulum clock to keep accurate time, it is necessary that the pendulum be so constructed that the moment of inertia and the length L are constant and remain the same in spite of temperature variations. Such pendulums are said to be *compensated.*

It is appropriate to note here that a clock or watch consists of two separate mechanisms. One of these beats off equal time intervals, and the other counts them and moves the hands. In a pendulum clock a physical pendulum is the means of measuring equal time intervals, while in a watch

or chronometer the torsional vibrations of the balance wheel perform this necessary function. The pendulum clock is adjusted to keep correct time by altering both h and I, while in the watch the coarse adjustment is made by varying the moment of inertia of the balance wheel, and the fine adjustment is made by altering the torsion constant κ of the watch spring. In both cases the primary characteristic of simple harmonic motion that is essential to the measurement of time is the fact that the period does not depend on the amplitude of vibration. Such motion is called *isochronous*, meaning that the oscillation requires the same time for all possible amplitudes.

12-7 Resonance

The frequency with which a pendulum or an elastic object vibrates when it is displaced from its equilibrium position and released is called the *natural frequency* of the system. Thus a piano string or a diving board will vibrate with its natural frequency after being displaced from an

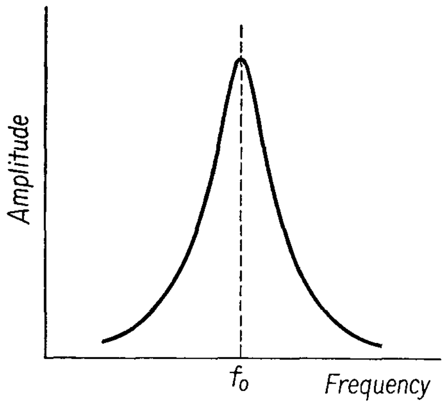


Fig. 12-8 The amplitude of vibration of an oscillator of natural frequency f_0 depends upon the frequency f with which it is driven. At resonance, $f = f_0$, and the oscillator vibrates with maximum amplitude.

equilibrium position. If such a mechanical oscillator is not simply struck a blow, or displaced from an equilibrium position and then released, but is driven by a force which varies periodically, it is required to move with the frequency of the driving force. The amplitude with which the oscillator will vibrate under the influence of such a force will depend markedly upon the driving frequency. When the frequency with which the oscillator is driven is far from its natural frequency, the amplitude of vibration will be quite small, but when the driving frequency is the same as the natural frequency of the oscillator, it will be excited to large amplitudes, as shown in Figure 12-8. This condition is known as *resonance*. At resonance the amplitude of the oscillations will depend upon the degree of *damping* in the oscillator itself, that is, on the amount of internal friction in the oscillator.

Although we shall not undertake the analysis of resonant motion, we shall mention some examples of resonance of importance in engineering. Any elastic structure which is deflected to large amplitudes by a periodic driving force is thereby exposed to large alternating stresses and is likely to fail at far smaller loads than the same structure could safely withstand under static loading, through a process called *fatigue failure*.

As an illustration, consider an airplane engine that is designed to operate at N rev/sec. Because it is impossible to build a perfectly balanced engine, the rotation of the engine acts as a driving force which drives the

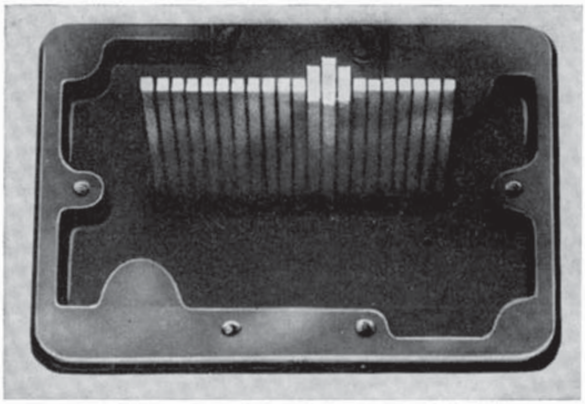


Fig. 12-9 Reed comb of a Frahm tachometer. The reeds are tuned in intervals of 50 vib/min from 3,000 at the left to 4,000 at the right. The assemblage was in contact with the case of a motor rotating at 3,600 rpm. (Courtesy of James G. Biddle Co.)

entire airplane at a frequency of N vib/sec. Any part of the airplane whose natural frequency is close to N will be excited to large amplitudes. If, by inadvertence, the natural frequency of the wing, or the propeller, were the same as that of the engine, failure of these parts would occur in a much shorter time than might otherwise be the case. In such circumstances a part can often be made stronger, that is, can be made to last longer, by removing material from it in such a way as to alter its natural frequency.

The phenomenon of resonance can be used in the measurement of the rotational frequency of a motor. Figure 12-9 shows a reed tachometer consisting of a comb of "harmonica" reeds which have been adjusted so that their natural frequencies are close to each other. If this *tachometer* is placed on the floor or table adjacent to a motor, some reeds will be driven by the vibration of the motor; the natural frequency of the reed driven to largest amplitude will measure the frequency of rotation of the motor. Such a tachometer may be mounted almost anywhere on board a single-engine airplane to indicate the engine speed in revolutions per minute.

Problems

12-1. A body whose mass is 500 gm hangs from a vertical spring whose constant is 200,000 dynes/cm. The body is pulled down a distance of 6 cm and released. Determine (a) the period of the motion, (b) the resultant force on the body when at the 6-cm point, and (c) the acceleration at this position.

12-2. In Problem 12-1 determine (a) the velocity of the body, when its displacement is 3 cm, and (b) its maximum velocity. Obtain your answers by considerations of simple harmonic motion and also by applying the principle of conservation of energy.

12-3. A body which has a mass of 60 gm is attached to a helical spring 25 cm long and, when lowered gently, stretches the spring 5 cm. The body is then pulled down an additional 8 cm and released, thus setting the spring in vibration. (a) What is the constant of the spring? (b) What is the period of oscillation?

12-4. A body which has a mass of 40 gm is attached to a spring, and the system is then set into vibration. The measured value of the period of vibration is 0.50 sec. (a) Determine the constant of the spring. (b) Determine the velocity of the body at the equilibrium position if the amplitude is 6 cm. (c) Determine its maximum acceleration.

12-5. When a cylinder whose mass is 4.0 kg is hung from a spring and set into motion, the frequency is 2.4 vib/sec. When another cylinder is substituted for the first one, the frequency of vibration is 3.2 vib/sec. Determine the mass of the second cylinder.

12-6. Determine the period of a simple pendulum, oscillating with small amplitude, when the length of the pendulum is 75 cm.

12-7. A simple pendulum 1.0 m long, having a mass of 250 gm, is displaced through an angle of 10° and released. Determine (a) the resultant force acting on the pendulum bob at this position of maximum displacement, (b) its maximum angular acceleration, (c) its maximum angular velocity, (d) the tension in the pendulum string when the displacement is 5°, and (e) the velocity and acceleration of the pendulum bob when the displacement is 5°.

12-8. (a) Determine the period of vibration of a pendulum 80 cm long at a place where $g = 980$ cm/sec². (b) What length of pendulum at the same place will have half this period?

12-9. A "seconds" pendulum has a period of 2 sec. A seconds pendulum which kept accurate time at a place where g was 980 cm/sec² is found to lose 2 min/day at a new location. Find g at this new location.

12-10. An object moves in simple harmonic motion with period of 4 sec and amplitude 1 m. (a) What is the frequency of the motion? (b) What is the velocity when the displacement is 30 cm? (c) What is the acceleration when the displacement is -60 cm? (d) How long a time is required for the object to move from the point where the displacement is 30 cm to the point where the displacement is -60 cm?

12-11. An object moves in rotational simple harmonic motion with period 4 sec and amplitude 1 radian. (a) What is the frequency of the motion? (b) What is the angular velocity when the angular displacement is 0.30 radian? (c) What is the angular acceleration when the angular displacement is -0.60 radian? (d) How long a time is required for the object to move from an angular

displacement of 0.30 radian to a point where the angular displacement is -0.60 radian?

12-12. Two simple harmonic oscillators have the same frequency of 4 cycles/sec and the same amplitude of 10 cm. At a particular time it is observed that the first oscillator is moving to the right and the second oscillator is moving to the left along the same line when both oscillators have the same displacement of 6 cm. (a) How far were they apart 0.25 sec later? (b) How far were they apart 0.10 sec after the initial observation?

12-13. Two simple harmonic oscillators, having the same amplitude of 10 cm in the y direction, are displaced so the first has a positive initial displacement of 10 cm, while the second has a negative initial displacement of 10 cm. The first oscillator has a frequency of 8 cycles/sec, while the second has a frequency of 4 cycles/sec. The oscillators are released at the same instant. Find the time at which both oscillators are at the same position.

12-14. A spring of spring constant 50 lb/ft has one end fastened to the wall and the other end fastened to a rectangular block of wood weighing 32 lb. The spring is horizontal, and the block rests on a horizontal table. The coefficient of friction between the block and the table is 0.1. The block is initially displaced so that the spring is stretched 1 ft from its equilibrium position and is released. What will be the speed of the block when it passes through the equilibrium position? [NOTE: The motion is not simple harmonic, for there is loss of mechanical energy through friction. Solve from energy considerations.]

12-15. A meter stick of mass 400 gm is pivoted 30 cm from one end and is allowed to oscillate as a physical pendulum. The width of the meter stick is 2 cm, and its thickness is $\frac{1}{2}$ cm. Find the period of the oscillation.

12-16. A circular disk 1 cm thick and of radius 20 cm, having a mass of 4 kg, is suspended from a pivot 2 cm from the edge of the disk and is allowed to oscillate while it hangs downward. Find the period of the disk for small oscillations.

12-17. A particle moves in simple harmonic motion in the x direction with amplitude 10 cm and frequency 5 cycles/sec. The particle is moving in the positive x direction at a displacement of $x = +5$ cm when $t = 0$. Find the position of the particle when $t = 0.02$ sec.

12-18. Two identical springs are laid side by side on a horizontal frictionless table, each having a spring constant of 1 lb/in. A sphere of mass 0.1 slug is connected to the free end of one spring, while a second sphere of mass 0.2 slug is connected to the free end of the second spring. Both masses are drawn aside a distance of 6 in. and released. What is the displacement of the sphere of mass 0.2 slug when the displacement of the other sphere is -3 in.?

12-19. Find the period of vibration of a cylinder of radius r, height h, and density ρ which is floating upright, partially immersed in a fluid of density ρ_0.

12-20. A wire is bent in the form of the arc of a circle of radius R, and is mounted so that it is in the vertical plane. A bead is placed upon the wire and released. In the absence of friction between the bead and the wire, show that the bead will oscillate with period

$$T = 2\pi(R/g)^{1/2}$$

provided that its initial displacement is sufficiently small.

13

Properties of Matter

13-1 Internal Forces

When a system is subjected to external forces, it generally undergoes a change in size or shape or both. We have thus far touched very lightly on such changes; for example, we have considered the change in length of an elastic spring and the change in volume of a gas when such systems were subjected to varying pressures. The changes produced in a system by the action of external forces depend upon the physical properties of the material of which the system is composed. A study of the properties of matter leads to information which is of practical value to both the physicist and the engineer, and also gives us some information about the internal forces which act between the constituent parts of the substance. In the final analysis these physical properties must be explicable in terms of the forces between the molecules of the substance and, in some cases, between the atoms of the substance.

We have so far discussed only one type of force which exists between particles—the gravitational attraction of two particles because of their masses. However, gravitational forces are much too small to account for the observed properties of substances. Furthermore we frequently encounter cases in which a force of repulsion is needed to explain the phenomena, whereas gravitational forces are always forces of attraction. One other fact worth noting here is that the forces which act between molecules produce their effect only over very short distances, that is, distances of the order of molecular diameters. These are called *short-range forces.* These short-range forces are undoubtedly of electrical origin. As we proceed with our study of physics, we shall study the forces between electrically charged particles and show how these are thought to be related to the structure of matter.

At present, there are 102 different elements known. A chemical analysis of any substance will show that it is composed of one or more of these elements. If the substance is a chemical compound, the elements

composing it always occur in a definite ratio of their weights. More complex substances consist of several or many compounds. The smallest constituent of a chemical compound is the *molecule;* it is the fundamental structural unit of the compound. A molecule is composed of one or more atoms held together by the short-range forces due to their electrical charges. As we shall show, an atom of an element consists of a very small but massive *nucleus* surrounded by a suitable number of *electrons.* The nucleus consists of two kinds of particles: *neutrons,* which are neutral particles, and *protons,* which are positively charged electrically. Every atom of any one element has exactly the same number of protons in its nucleus. This number is the *atomic number* of the element and can have one of the values from 1 to 102, inclusive. In the normal state of the atom, the atomic number also represents the number of electrons outside the nucleus. Most of the chemical properties of an element can be explained in terms of the arrangement and behavior of the electrons in the atoms. The electrons exert forces of repulsion on each other and forces of attraction on positively charged nuclei. When a molecule is formed with two or more atoms, there is a change in the electronic arrangement, and a new arrangement of charges is produced. This rearrangement is brought about by the short-range forces which act between the atoms.

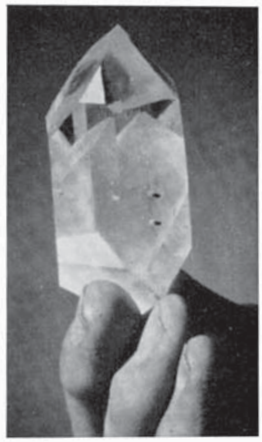


Fig. 13-1 Photograph of a large quartz crystal grown at the Bell Telephone Laboratories. (Courtesy of Bell Telephone Laboratories.)

When the substance is in the solid phase, the forces which exist between atoms and molecules cause them to form definite geometrical patterns; these show up as the crystalline structure. Sometimes these crystals grow to a fairly large size, as in the case of rock salt or quartz (see Figure 13-1). Sometimes the crystals are very small and can be seen only with the aid of a microscope. Most metals consist of such microcrystals. Even in the liquid phase, there is a definite grouping of the atoms and molecules, although individual atoms and molecules often change places. The arrangement of the atoms and molecules of a substance can be determined by means of x-ray analysis (see Chapter 44).

In the gaseous phase, the molecules are comparatively far apart, so that the forces they exert on each other are extremely small. It is this aspect of the structure of gases which makes it possible to analyze the behavior of gases almost without regard to their chemical nature, although it is not possible to neglect the chemical differences in either the liquid or the solid phase.

13-2 Elasticity

The subject of elasticity has already been introduced in connection with our discussion of periodic motion. In order to treat the elastic properties of matter more quantitatively, and in a manner which is not dependent on the configuration or shape of the body being studied, we find it convenient to introduce two new terms: these are (a) *stress* and (b) *strain.*

Stress is defined as *the internal force* F, *brought into play when the substance is distorted in any way, divided by the area* A *over which it acts.*

Thus
$$\text{Stress} = \frac{F}{A}\cdot \tag{13-1}$$

The stress is therefore the force per unit area. Clearly, the stress indicated in Equation (13-1) is an average stress, for the stress in a member may vary from point to point within that member, as in the case of a bent beam, where one surface of the beam is in tension, while the other surface of the beam is in compression. We shall restrict this discussion to the simplest case, where the stress is uniform, and shall not attempt to treat cases of variable stress for which the stress at a point would be defined through a limiting process.

In the cgs system the stress is expressed in dynes per square centimeter; in the British gravitational system, in pounds per square foot; in the mks system, in newtons per square meter. In most engineering practice the stress is expressed in pounds per square inch.

Strain is defined as *the ratio of the change in size or shape to the original size or shape.* As a ratio, strain has no physical dimensions; that is, it has numerical value only. Methods of expressing the strain will be given in the discussion of the various cases.

The relationship between stress and strain was first given by Robert Hooke (1635–1703) and is known as *Hooke's law.* This states that *for an elastic body the stress divided by the strain is a constant,* or,

$$\frac{\text{Stress}}{\text{Strain}} = K, \tag{13-2}$$

where K is called the *modulus of elasticity.* The units for K are the same as those for stress, since strain is expressed as a pure number.

13-3 Tensile Stress and Strain

As an example of the stress set up inside a substance, let us consider the increase in the length of a rod produced by the action of two forces, each equal to F, applied at the ends of the rod, as shown in Figure 13-2(a).

These forces are applied by means of clamps C_1 and C_2 attached to the ends of the rod.

If L is the original length of the rod, and if Δl is the increase in length produced by the application of the forces F, then the strain produced is

$$\text{Strain} = \frac{\text{increase in length}}{\text{original length}} = \frac{\Delta l}{L}. \tag{13-3}$$

To determine the stress in the rod, let us take any cross-sectional area A through the rod and consider the forces which act on it, as in Figure 13-2(b). The entire rod is in equilibrium under the action of the two external forces, each of magnitude F but acting in opposite directions, so as to extend the rod. These forces produce a *tension* in the rod. Let us consider an area A near the clamp C_2; it is acted on by the external force F downward, and, since it is in equilibrium, it must also be acted upon by a force F upward exerted by that part of the rod which is immediately above this area. If we take any other parallel cross-sectional area A, it will be acted upon by a force downward equal to F exerted by that part of the rod below it, and another force equal to F exerted by that part of the rod above it. The effect of these two forces is to tend to separate the rod across this section; it is opposed by the forces of attraction between the molecules on the two sides of this section. The stress in the rod is the quotient of one of these forces F by the cross-sectional area A, or

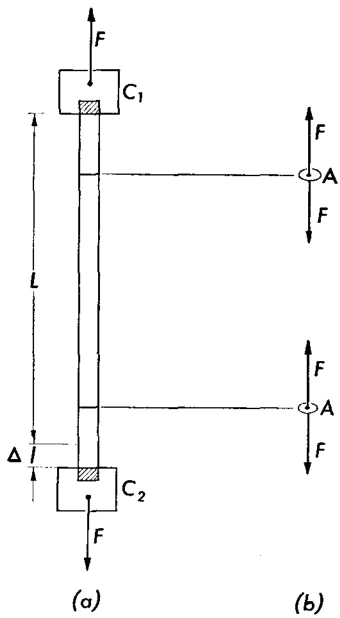


Fig. 13-2 Tensile stress in a rod.

$$\text{Stress} = \frac{F}{A}.$$

This type of stress is called a *tensile stress.*

If the material of the rod is elastic, then we know, from Hooke's law, that

$$\frac{\text{Stress}}{\text{Strain}} = K.$$

Putting in the values of stress and strain found above, and replacing the

letter K by Y, we get

$$\frac{F/A}{\Delta l/L} = Y. \tag{13-4}$$

Y is called *Young's modulus* after Thomas Young (1773–1829), an eminent British physician, physicist, and linguist who made significant contributions to the study of vision and the theory of light, and who deciphered Egyptian hieroglyphics, in addition to his studies in elasticity. The values of Y for several substances are listed in Table 13-1.

TABLE 13-1 ELASTIC CONSTANTS OF SOME SOLIDS

Material	Young's Modulus		Shear Modulus		Bulk Modulus	
	in dynes/cm^2 $\times 10^{11}$	in lb/in.^{2} $\times 10^6$	in dynes/cm^2 $\times 10^{11}$	in lb/in.^{2} $\times 10^6$	in dynes/cm^2 $\times 10^{11}$	in lb/in.^{2} $\times 10^6$
Aluminum, rolled	6.96	10.1	2.37	3.44	7.	10.
Brass	9.02	13.1	3.53	5.12	6.1	8.5
Copper, rolled	12.1 –12.9	17.5 –18.6	4.24	6.14	14	21
Duralumin	6.89	10.0	2.75	3.98		
Iron, cast	8.4 – 9.8	12–14	...	...	9.6	14
Glass, crown	6.5 – 7.8	9.5 –11.3	2.6 –3.2	3.8 – 4.7		
Lead	1.47– 1.67	2.13– 2.42	0.54	0.78	0.8	1.1
Nickel	20.0 –21.4	29.0 –31.0	7.06–7.55	10.24–10.95		
Platinum	16.67	24.18	6.42	9.32		
Silver, hard drawn	7.75	11.24	2.00	3.77		
Steel, annealed	20.0	29.0	8.11	11.76	16	23
Tin	3.92– 5.39	5.69– 7.82	1.67	2.42		
Tungsten, drawn	35.5	51.5	14.8	21.5		

The extent to which a substance remains elastic as the tensile stress is increased can be determined only by experiment. Figure 13-3 shows the results of a typical experiment on a metallic rod. In this figure the stress is plotted as ordinate and the strain as abscissa. The curve is obtained by exerting a force, measuring the strain, increasing the force, measuring the new strain, and so on. The straight-line portion of this curve, from O to E, represents the values of the stress and the strain for which the rod is elastic. If a stress of value CD is applied to the rod and then removed, the rod will no longer return to its original length. It is said to have a per-

manent set. The point E is called the *elastic limit* of the material. At point B the stress was great enough to break the rod. This value of the stress is known as the *ultimate stress* or the *breaking stress* of the material.

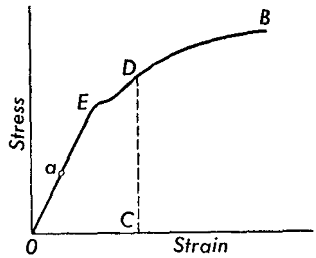


Fig. 13-3 Stress-strain curve for a ductile material.

Illustrative Example. A piece of copper wire 0.0508 in. in diameter and 3 ft long is suspended from a rigid support and supports a load of 8 lb. Determine (a) the stress in the wire, (b) the increase in length produced by the 8-lb load, and (c) the strain produced.

(a) The wire has a cross-sectional area of 0.00203 in.^{2}. The stress in this wire is

$$\text{Stress} = \frac{F}{A} = \frac{8 \text{ lb}}{0.00203 \text{ in.}^2} = 3{,}950 \frac{\text{lb}}{\text{in.}^2}\cdot$$

(b) The increase in length can be found from Equation (13-4), provided that Y is known. From Table 13-1 the value of Young's modulus for copper is given as $17.5 \times 10^6 \dfrac{\text{lb}}{\text{in.}^2}$. Since both Y and the stress are expressed in the same units, there is no need to convert either quantity into units appropriate to the British gravitational system, and we write

$$Y = \frac{F/A}{\Delta l/L};$$

hence

$$\Delta l = \frac{F}{A} \times \frac{L}{Y},$$

$$\Delta l = 3{,}950 \frac{\text{lb}}{\text{in.}^2} \times \frac{36 \text{ in.}}{17.5 \times 10^6 \dfrac{\text{lb}}{\text{in.}^2}},$$

$$\Delta l = 0.008 \text{ in.}$$

(c) The strain produced in the wire is

$$\text{Strain} = \frac{\Delta l}{L} = \frac{0.008 \text{ in.}}{36 \text{ in.}} = 0.00022,$$

$$\text{Strain} = 2.2 \times 10^{-4}.$$

The strain, or the fractional change in length, is about 2 parts in 10,000.

13-4 Compressive Stress and Strain

If the ends of a rod of some material are subjected to the action of two forces, each of magnitude F but directed so as to diminish its length, the rod is said to be under *compression,* and the stress inside the rod is a *compressive stress,* as illustrated in Figure 13-4(a). If we consider the forces which act across any cross-sectional area A, that part of the rod to the right

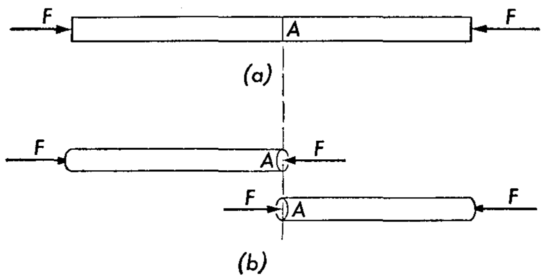


Fig. 13-4 (a) Rod put under compression by action of two external forces each equal to F. (b) Any cross section A is acted upon by an internal force F to the left due to the section of the rod on the right, and by an internal force F to the right due to the section of the rod on the left.

of this area exerts a force F to the left, while that part to the left of this area exerts a force to the right, as shown in Figure 13-4(b). The compressive stress in the rod is the quotient of one of these forces divided by the area over which the force acts, exactly as in the case of the tensile stress. The molecular forces brought into play by the action of the external forces must be forces of repulsion. The strain produced by the compressive stress is the ratio of the decrease in length Δl to the original length, again exactly as in the case of tensile strain. If the material of which the rod is made is elastic, then experiment shows that, within the elastic limit, the compressive stress divided by the compressive strain, or Young's modulus for compression, is identical with Young's modulus for tension for the same material. In consequence, no distinction is made in tables between Young's modulus for tension or for compression, and tabular values of Young's modulus are intended for use in both types of stress.

13-5 Compressibility of Gases: Boyle's Law

The compressibility of gases was first studied by Robert Boyle (1627–1691). Suppose we have a mass of gas in a cylinder with a tight-fitting piston, on which a force F is exerted producing a pressure $P = F/A$, where A is the area of the piston, as shown in Figure 13-5. The gas will be subject to this pressure and will occupy a volume V determined by the distance of

the piston from the end of the cylinder. By increasing the force on the piston to a new value F_1, the pressure on the gas will be increased to a new value P_1. If the process takes place slowly, so that the gas remains at constant room temperature, it is found that

$$\frac{V_1}{V} = \frac{P}{P_1},$$

or, stated in words, at constant temperature the volume of a gas varies inversely as the pressure. A more convenient way to express the same result is

$$P_1V_1 = PV = \text{a constant}; \tag{13-5}$$

that is, *the product of the pressure and volume of an enclosed gas remains constant* if the temperature is constant. This statement is known as *Boyle's law.*

Since the mass of gas within an enclosed system remains constant, we can determine the manner in which the density of a gas varies with pressure by dividing both sides of Equation (13-5) by the mass M of the gas within the cylinder. We obtain

$$\frac{P_1}{M/V_1} = \frac{P}{M/V},$$

or

$$\frac{P_1}{P} = \frac{\rho_1}{\rho}, \tag{13-6}$$

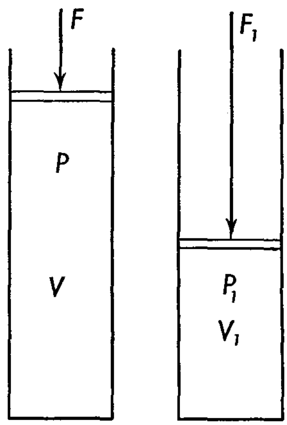


Fig. 13-5 Gas in a cylinder is compressed by increasing the force on the piston; that is, by increasing the pressure of the gas.

in which ρ is the density of the gas at pressure P, and ρ_1 is the density of the gas at pressure P_1.

Illustrative Example. A steel tank contains 2 ft^3 of oxygen at a gauge pressure of 200 lb/in.^{2}. What volume will this gas occupy at the same temperature at atmospheric pressure?

The gauge pressure is the difference between the pressure of the gas in the cylinder and the pressure of the atmosphere. Hence, the pressure P_1 of the oxygen in the cylinder is 214.7 lb/in.^{2}; the pressure P of the atmosphere is taken as 14.7 lb/in.^{2}. Using Boyle's law in the form of Equation (13-5) and substituting values, we get

$$214.7\,\frac{\text{lb}}{\text{in.}^2} \times 2\text{ ft}^3 = 14.7\,\frac{\text{lb}}{\text{in.}^2} \times V,$$

$$V = 29.2\text{ ft}^3.$$

13-6 Volume Change: Bulk Modulus

In both compressive and tensile stresses, the stress acts along one direction in the body and produces a change in only one dimension. The change produced in the cross-sectional area of a rod under compression or tension is practically negligible. To produce equal strains in all three dimensions of a homogeneous solid, it is necessary to have equal stresses along these three directions. The simplest method of doing this is to subject the solid to a uniform hydrostatic pressure. Since the dimensions of the solid are normally measured at some pressure P_0 which is generally the atmospheric pressure, we are usually interested in the change in the dimensions of the solid when it is subjected to a change in pressure from P_0 to a different pressure $P_0 + \Delta P$. If we call V the volume of the solid and ΔV the change in volume produced by the change in pressure ΔP, then, from Hooke's law,

$$\frac{\text{Stress}}{\text{Strain}} = \frac{\Delta P}{\Delta V/V} = K.$$

Since an increase in pressure always produces a decrease in volume, the modulus K for volume change will always be a negative number. To avoid having a negative number, let us define the *bulk modulus* $B = -K$, so that

$$B = -\frac{\Delta P}{\Delta V/V}. \tag{13-7}$$

Not only solids but liquids and gases undergo volume changes when subjected to changing pressures, and Equation (13-7) can be applied to fluids as well as to solids. Since the denominator of Equation (13-7) is a ratio of the change in volume to the volume, it is a pure number; hence the bulk modulus is expressed in units of pressure, or force per unit area.

TABLE 13-2 BULK MODULUS OF LIQUIDS

Material	in dynes/cm^2 $\times 10^{11}$
Carbon disulphide	0.15
Ethyl alcohol	0.09
Glycerin	0.45
Mercury	2.6
Nitric acid	0.03
Water	0.23

The bulk modulus of a gas which obeys Boyle's law at constant temperature is given by the pressure of the gas, for we may state Boyle's law as

$$PV = \text{constant},$$

and, taking differentials of this equation, we find

$$P\,dV + V\,dP = 0,$$

and, solving for P, we find

$$P = -\frac{dP}{dV/V}.$$

If we replace the differentials by small increments so that the differential dP is replaced by ΔP and the differential dV is replaced by ΔV, we have

$$P = -\frac{\Delta P}{\Delta V/V} = B.$$

From the above analysis it can be seen that, if the pressure of a gas is changed by any amount while the temperature is kept constant, the bulk modulus will vary and, at any stage of the process, will be equal to the pressure of the gas at that stage. This is the reason that tires or basketballs inflated to high pressures seem hard, or difficult to deform, while the same object inflated to low pressure is easy to deform, or soft.

13-7 Shearing Stress and Strain

It is possible to produce a change in the shape of a solid without changing its volume. Such a distortion is called a *shear*. A simple method of producing a shear is illustrated in Figure 13-6(a). If we take a rectangular solid and apply a force F along its top surface, and an equal force F acting in the opposite direction along its bottom surface, the rectangular surfaces such as $BCDE$ at right angles to the top and bottom surfaces will be distorted into parallelograms such as $B'C'DE$, whose angles are not right angles. If we imagine the solid as made up of a series of layers parallel to the top and bottom surfaces, each of area A, then the effect of the shear is to cause one layer to slide with respect to another layer, much as in a deck of cards.

The shearing stress set up on the solid is the force F divided by the area A of the surface over which it acts. Each layer parallel to the top and bottom surfaces is acted upon by two forces, as shown in Figure 13-6(b), the layer above it exerting a force to the right, and the layer below it exerting a force to the left. These forces between the molecules of the layers oppose the sliding of one layer with respect to another. If the applied force becomes too great, the solid will be sheared; that is, it will be separated into two parts, with the surfaces of separation parallel to the direction of the applied force.

The shearing strain produced is measured by the ratio of the distance Δx through which the top surface has been moved relative to the bottom

surface, to the height h, that is, the distance between the two surfaces. Thus

$$\text{Shearing strain} = \frac{\Delta x}{h}.$$

It will be noted that the tangent of the angle $B'EB$ is $\Delta x/h$.

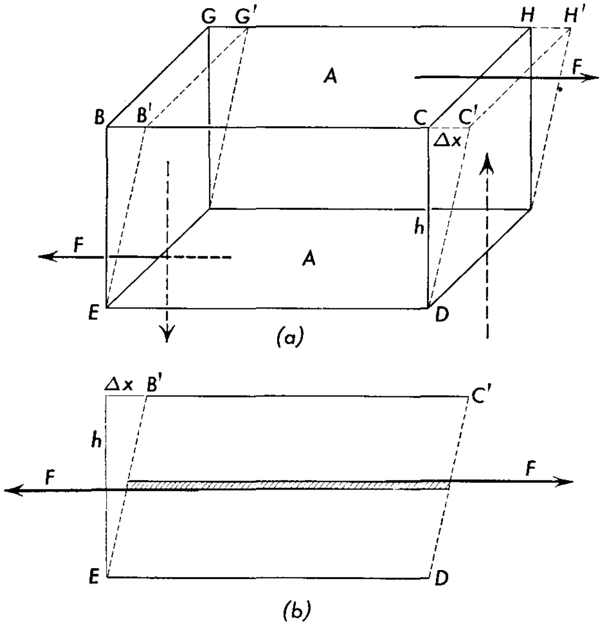


Fig. 13-6 (a) Shearing stress set up in a solid by parallel forces F acting on top and bottom surfaces each of area A. Rectangle $BCDE$ is distorted into parallelogram $B'C'DE$. There is no change in the volume of the solid. (b) Shaded area is section of layer in the solid parallel to top and bottom surfaces. Shearing strain is $\Delta x/h$.

Applying Hooke's law, we get for the *shear modulus* M,

$$M = \frac{F/A}{\Delta x/h}, \tag{13-8}$$

and once again we note that the units of the shear modulus are the units of force per unit area.

When a thin-walled tube or a rod is twisted, its deformation depends on the shear modulus. The force constant of a helical spring may be computed from the dimensions of the spring and the shear modulus of the wire from which it is made. When a spring is elongated or compressed, the wire from which it is made does not stretch, but rather, any short length of the wire acts like a rod which is being twisted.

While we have defined three constants of elasticity as though they were completely independent, it may be shown that, for a homogeneous material in which the elastic properties are the same in every direction, only two of these constants are independent, and the third may be expressed in terms of the other two.

While all three of the elastic constants are meaningful for solids, only the bulk modulus has meaning for fluids. This is because neither a liquid nor a gas is capable of supporting a tensile or a compressive stress, nor can either of these fluids support a shearing stress under ordinary conditions. When a fluid is subjected to a shearing stress, the upper surface acquires a uniform velocity with respect to the lower surface of Figure 13-6, as we have seen in the discussion of viscosity. Only the bulk modulus is appropriate to a liquid or gas, for these are capable of withstanding only hydrostatic pressures.

Illustrative Example. Two clamps are fastened near the ends of a rectangular steel rod 5 in. long. The rectangular cross section of the rod has an area of 2.5 in.^{2}. A force of 800 lb is exerted on each of these clamps parallel to this area but in opposite directions. Determine (a) the shearing stress, (b) the shearing strain, and (c) the relative displacement of the top surface with respect to the bottom surface.

(a) The shearing stress is

$$\frac{F}{A} = \frac{800 \text{ lb}}{2.5 \text{ in.}^2} = 320 \frac{\text{lb}}{\text{in.}^2}.$$

(b) The shearing strain can be found by solving Equation (13-8) for $\Delta x/h$, obtaining

$$\frac{\Delta x}{h} = \frac{320}{11.8 \times 10^6} = 27.2 \times 10^{-6}.$$

(c) Since $h = 5$ in.,

$$\Delta x = 5 \text{ in.} \times 27.2 \times 10^{-6}.$$

so that

$$\Delta x = 1.36 \times 10^{-4} \text{ in.}$$

for the relative displacement of the two surfaces.

13-8 Cohesion and Adhesion

The fact that molecular forces have a short range would lead us to expect some distinctive types of phenomena to be observable at the surfaces of substances. Conversely, the appearance of these surface phenomena should lead to information about these molecular forces. For example, if we take two pieces of metal, each with an accurately plane surface, and bring them together, there will be no observable force between them until the two surfaces are placed in contact. Once they are placed in contact,

a very great force will be required to pull them apart. This experiment, which can readily be performed with two pieces of steel or two pieces of lead, shows that the forces between the molecules in the two surfaces have a very short range of effectiveness. It is for this reason that a bearing in which a shaft rotates is never made of the same material as the shaft itself. The force of attraction between molecules of the same substance is sometimes called *cohesion.*

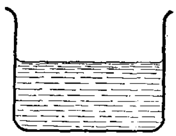


Fig. 13-7 Free surface of water in a glass jar is level (horizontal) except near the glass.

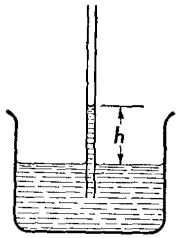


Fig. 13-8 Level of water in a capillary tube is at a height h above level in the large vessel.

If some water is poured into a glass vessel, the free surface of the water will be a level surface, except at the region of contact with the glass; at this region, the water will be seen to cling to the glass for a short distance above the level surface, as shown in Figure 13-7. This phenomenon can be accentuated by immersing a glass tube with a narrow bore, a *capillary tube,*

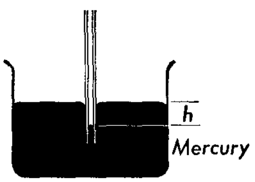


Fig. 13-9 Level of mercury in a glass capillary tube is at a level h below that in a large vessel.

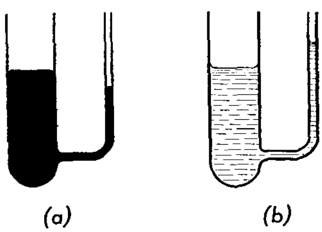


Fig. 13-10 (a) **U** tube containing mercury. (b) **U** tube containing water.

into the water, as in Figure 13-8. The level of the water inside the capillary tube will be found to be considerably higher than the level inside the larger jar, and an examination of the surface of the water in the capillary tube shows that it is not plane but is concave, and spherical in shape. Not all liquids behave like water. For example, if a glass capillary tube is immersed in mercury contained in a larger dish, as in Figure 13-9, the level

of the mercury will be lower in the capillary tube than in the dish. If a glass **U** tube is constructed with one arm about 1 cm in diameter and the other about 0.2 cm in diameter, and mercury is poured into the tube, the level in the narrower tube will be lower than in the wider tube, while if water is poured into such a **U** tube, the level of the water will be higher in the narrower tube, as shown in Figure 13-10.

One method of accounting for the behavior of liquids in capillary tubes is to assume that there are forces of attraction, also of short range, between the molecules of the liquid and the molecules of the solid at the surfaces

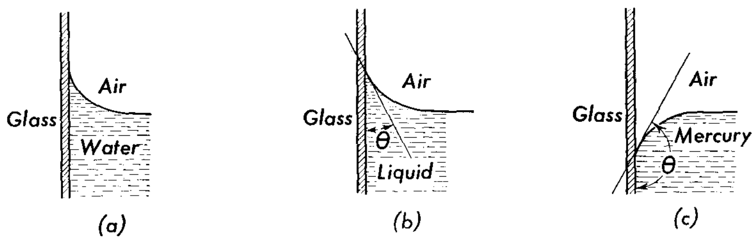


Fig. 13-11 Angles of contact. (a) 0° between water and glass; (b) angle of contact $\theta < 90°$ between a liquid and glass; (c) angle of contact $\theta > 90°$ between mercury and glass.

of contact. This type of attractive force between molecules in the surface of one substance for those in the surface of another substance is sometimes called *adhesion*, to distinguish it from the force of cohesion between like molecules. If the force of adhesion is greater than the force of cohesion, the liquid will cling to the solid surface; that is, it will wet the solid. If the force of cohesion is greater than the force of adhesion, the liquid will cling to itself; that is, it will tend to form droplets when placed on a smooth surface, rather than spreading out to wet the surface.

The angle between the liquid surface and the solid surface at the region of contact is an indication of the relative values of the forces of adhesion and cohesion. This angle is known as the *angle of contact.* For water and glass the angle of contact is practically 0°, while for some other liquid, the angle of contact will have some value θ, as shown in Figure 13-11. If the force of cohesion is much greater than the force of adhesion, as in the case of mercury and glass, the angle of contact θ is greater than 90°. For mercury and glass, $\theta = 139°$. Values of some contact angles are given in Table 13-3.

13-9 Surface Tension

We have seen that the liquid inside a capillary tube has a curved surface, and, if the tube is cylindrical, the surface of the liquid may be nearly

TABLE 13-3 CONTACT ANGLES

Liquid	Tube	Angle, Degrees
Alcohol	Glass	0
Ether	Glass	0
Glycerin	Glass	0
Mercury	Glass	139
Water	Glass	0
Water	Paraffin	107

spherical. The interesting phenomena associated with liquid surfaces can be most easily explained by introducing the concept of *surface tension.*

That a liquid surface behaves as though it is under tension can be demonstrated in a variety of experiments. Let us construct a rectangular wire frame having one side movable; this can be done by curving the ends of a wire AB so that it slides easily on two legs of the frame, as in Figure

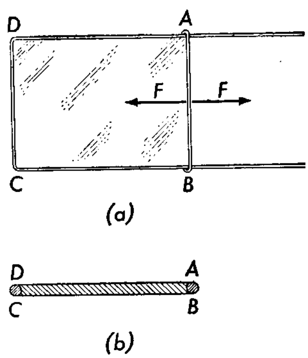


Fig. 13-12 (a) Wire frame with movable slide AB used to measure the surface tension of a film in the frame $ABCD$. (b) Shows the thickness of the film.

13-12. We can pick up a film on this frame by dipping it in a soap solution. This film will have two rectangular surfaces. The film will tend to contract, and, since AB is movable, the film will pull this wire toward CD with some force F.

To keep the wire AB in equilibrium, a force F to the right has to be applied to it. This force can be used to measure the surface tension. The force exerted on the wire depends on the length of the wire. *We define the surface tension* S *as the force exerted by a single surface on a section of unit length.* In this example the force exerted by each surface is $F/2$. If l is the length of the wire, the surface tension is therefore

$$S = \frac{F}{2l}\cdot$$

If the surface area is increased by moving the wire AB through a

distance x, the work done is

$$\mathscr{W} = Fx,$$

and, since
$$F = 2lS,$$

therefore
$$\mathscr{W} = S \times 2lx.$$

Now $2lx$ is the increase in surface area of the film; setting $2lx = A$, we find

$$S = \frac{\mathscr{W}}{A}. \tag{13-9}$$

The surface tension thus represents the work done per unit area in increasing the area of a film. From mechanics we know that the most stable configuration of a mechanical system is the position of lowest potential energy. Thus a soap bubble or a water droplet assumes a spherical shape, for the spherical surface contains the greatest volume per unit of surface area of any three-dimensional figure and is therefore the surface of lowest energy.

The surface tension of a liquid depends on the nature of the liquid and the nature of the substance outside the liquid surface, that is, whether it is air or the vapor of the liquid itself. The values of the surface tensions of liquids are given in Table 13-4. The surface tension also depends on the temperature of the system, decreasing as the temperature rises.

TABLE 13-4 SURFACE TENSION

Liquid in Contact with Air	Temperature in °C	Surface Tension in dynes/cm
Ethyl alcohol	20	22.3
Water	0	75.6
	20	72.8
	60	66.2
	100	58.9
Mercury	25	473
Olive oil	20	32
Glycerin	20	63.1
Soap solution		26

Soap films provide an analogue solution to many engineering problems. The soap film has the property that the surface tension at every point in the surface is a constant. In an ideal engineering structure every element might well be stressed to the same tensile or compressive stress, to gain the maximum strength per unit weight of the total structure. Suppose one wished to design a Plexiglas dome for a pressurized vessel in which the

opening was to be of irregular shape. One way to find the best shape for the dome is to build a model vessel with the desired opening, cover the opening with a soap film, and blow a bubble. The soap bubble automatically will come to the best design shape for the desired pressure differential. If proper scaling factors are known, the design of the desired dome may be determined from this experiment.

13-10 Pressure and Curved Surfaces

Let us consider the forces holding a spherical surface together. Suppose a sphere of radius R is imagined to be parted by a plane diaphragm, as shown in Figure 13-13. Let us imagine that the upper (dotted) part of the sphere

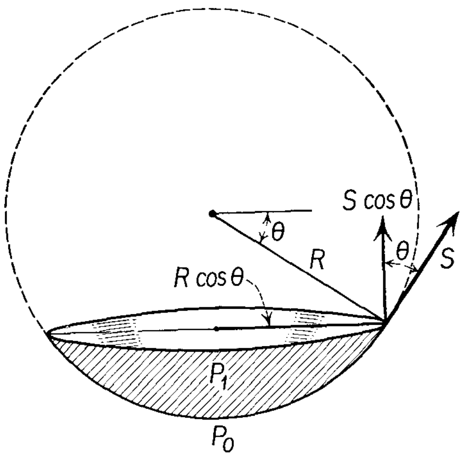


Fig. 13-13

has been removed. The interior of the cap contains fluid at pressure P_1 which is greater than the pressure P_0 outside the volume bounded by the cap and diaphragm. The force which tends to blow the diaphragm off the spherical cap is given by the difference in pressure $P_1 - P_0$ multiplied by the area of the diaphragm $\pi(R \cos \theta)^2$. The only force which tends to hold the spherical cap to the diaphragm is the surface tension S, whose component perpendicular to the diaphragm is $S \cos \theta$. The total force exerted by the surface tension is the product of the perpendicular component by the perimeter of the circle which bounds the diaphragm and the spherical cap, for, by definition, the surface tension is a force per unit length of surface. The spherical cap will be in equilibrium on the diaphragm when these two forces are equal. Thus

$$(P_1 - P_0) \times (\pi R^2 \cos^2 \theta) = (S \cos \theta) \times (2\pi R \cos \theta),$$

or

$$P_1 - P_0 = \frac{2S}{R}. \tag{13-10}$$

Equation (13-10) has been deduced by replacing the upper (dotted) portion of the sphere by an imaginary diaphragm. We may now imagine the dotted portion of the sphere to be replaced over the lower spherical cap, and the diaphragm removed. The equation now describes the relationship between the gauge pressure within the sphere $P_1 - P_0$, the surface tension S, and the radius R of the sphere.

Thus the difference in pressure on the two sides of a spherical surface due to the surface tension depends inversely on the radius of the sphere. P_0 is the pressure on the convex side of the surface and P_1 is the pressure on the concave side. The above result is not dependent on the angle θ and is therefore independent of the size of the spherical cap analyzed. It applies

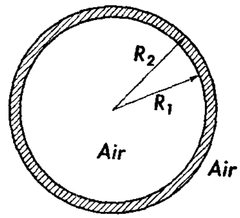


Fig. 13-14 A soap bubble has two spherical surfaces whose radii R_1 and R_2 differ very slightly.

to a whole sphere or any part of it. Thus the pressure inside a water droplet is greater than the atmospheric pressure, and the pressure inside a small droplet is greater than the pressure inside a large droplet. When two water droplets are placed in contact, the water tends to move from a region of high pressure into a region of low pressure, and the large droplet swallows the small one. From yet another point of view, Equation (13-10) shows why water droplets at rest are spherical rather than, say, pear-shaped, for the water within the smaller end would be at higher pressure than the water at the large end of a pear-shaped droplet. Equation (13-10) is applicable to the determination of the stress in any spherical shell. In a metallic shell of a given thickness, the stress must be multiplied by the thickness of the shell to give a quantity whose dimensions are force per unit length. In this way we can determine the pressure differential a spherical shell can withstand, or the breaking strength of a spherical diaphragm.

In the case of a soap bubble blown in air, the pressure difference is practically twice as great as given by Equation (13-10), that is, $4S/R$, since a soap bubble has two spherical surfaces of radius R_1 and R_2, as in Figure 13-14. Since the thickness of the soap film is very small, the difference between the two radii may be neglected, and in the equation for the pressure differential between the inside and the outside of the bubble we simply write the average radius R for the radius of the bubble.

Illustrative Example. Determine the gauge pressure inside a small raindrop 3 mm in diameter.

The gauge pressure is $P_1 - P_0$, for the pressure P_1 is the pressure inside the spherical surface, and the outside pressure P_0 is the atmospheric pressure. Substituting $S = 73$ dynes/cm for the surface tension of water, and $R = 0.15$ cm in Equation (13-10), we find

$$P_1 - P_0 = \frac{2 \times 73 \text{ dynes/cm}}{0.15 \text{ cm}},$$

$$P_1 - P_0 = 973 \text{ dynes/cm}^2.$$

13-11 Capillarity

We have already seen that, if a capillary tube is inserted into a liquid, the levels inside and outside the tube will differ by an amount h. In some cases the liquid will be higher in the capillary tube; in other cases it will be lower, depending upon the relative values of the forces of adhesion and cohesion. Furthermore, the surface of the liquid in the capillary tube, sometimes called the *meniscus*, may be approximated as spherical, if the bore is cylindrical and of sufficiently small diameter.

The concept of surface tension, and the pressure differences associated with curved surfaces discussed in the preceding paragraphs, enables us to obtain a simple relationship between the difference in levels h inside and outside a capillary tube and the radius r of this tube. Let us suppose that the angle of contact between the liquid and the material of the capillary tube is θ. The meniscus is of spherical shape of radius R such that

$$R \cos \theta = r.$$

If, as in Figure 13-15(a), we complete the spherical surface generated by the meniscus in dotted lines, we see that the atmosphere is inside the sphere and that the liquid in the capillary tube is outside the sphere. Thus the pressure in the liquid just outside the meniscus must be below atmospheric pressure by an amount given by Equation (13-10).

$$P_1 - P_0 = \frac{2S}{R},$$

$$P_{\text{atm}} - P_{\text{liquid}} = \frac{2S}{r/\cos \theta},$$

$$P_{\text{atm}} - P_{\text{liquid}} = \frac{2S \cos \theta}{r}.$$

The pressure at the free surface of the liquid outside the capillary tube is the atmospheric pressure, and at the same level inside the capillary tube the pressure must also be atmospheric. But the pressure beneath the meniscus has been reduced by an amount given by the formula above, so that the

meniscus must rise until the hydrostatic pressure generated by the column of liquid is equal to the reduction in pressure generated by the curvature of the meniscus, and we have

$$h\rho g = \frac{2S\cos\theta}{r},$$

$$h = \frac{2S\cos\theta}{r\rho g}, \tag{13-11}$$

where ρ is the density of the liquid.

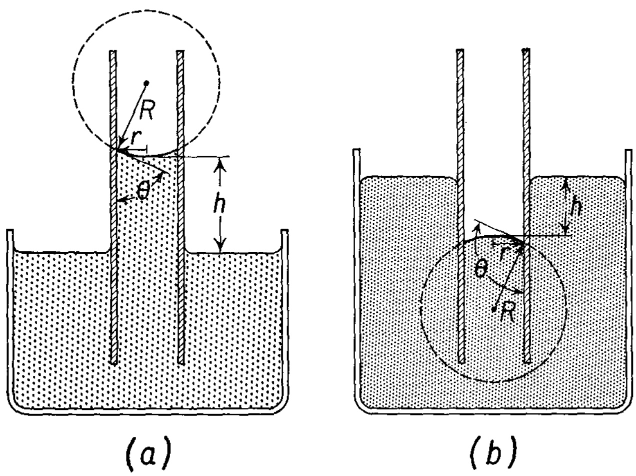


Fig. 13-15 Surface-tension forces in capillary tubes (a) when contact angle θ is less than 90°; (b) when contact angle θ is greater than 90°.

The same analysis will hold if the surface in the capillary tube is depressed by an amount h, as in Figure 13-15(b). Here the liquid may be thought of as being inside the sphere and at a higher pressure than the atmospheric pressure. In Equation (13-11), when the angle of contact is greater than 90°, the value of its cosine is a negative number, hence h will be negative, indicating that the level is depressed in the capillary tube.

From Equation (13-11) we see that a liquid will rise higher in a capillary tube of small bore than in a tube of large diameter. Capillary action is the basis of operation of mops, of sponges, of lampwicks, and of many other devices, where the fine spaces between the threads act as capillary tubes. In reading a mercury barometer, correction must be made for the effect of surface tension upon the height of the mercury column in order to obtain accurate measurements of the atmospheric pressure, for the height of the column is depressed from its true height by capillary action, which depends upon the diameter of the tube.

Illustrative Example. Two glass capillary tubes, each 1 mm in radius, are put into two different liquids, one in water and the other in mercury. Compare the liquid levels in the two tubes.

Let us take the level of the liquid outside each capillary tube as the zero reference level. From Equation (13-11) the level of the liquid inside the capillary will differ from that outside by an amount

$$h = \frac{2S\cos\theta}{r\rho g}. \tag{13-11}$$

For the case of water, $\theta = 0°$, hence $\cos\theta = 1$; $S = 73$ dynes/cm, $r = 0.1$ cm, $\rho = 1$ gm/cm^3, and $g = 980$ cm/sec^2. Letting $h = h_1$ for water, we get

$$h_1 = \frac{2 \times 73 \text{ dynes/cm}}{0.1 \text{ cm} \times 1 \text{ gm/cm}^3 \times 980 \text{ cm/sec}^2},$$

from which $$h_1 = \frac{146}{98} \text{ cm} = 1.5 \text{ cm}.$$

For the case of mercury, $\theta = 139°$, hence $\cos\theta = \cos 139° = -\cos 41° = -0.755$; $S = 473$ dynes/cm, $r = 0.1$ cm, $\rho = 13.6$ gm/cm^3, and $g = 980$ cm/sec^2. Letting $h = h_2$ for mercury, we get

$$h_2 = -\frac{2 \times 473 \text{ dynes/cm} \times 0.755}{0.1 \text{ cm} \times 13.6 \text{ gm/cm}^3 \times 980 \text{ cm/sec}^2},$$

from which $$h_2 = -0.536 \text{ cm}.$$

13-12 Phenomena Associated with Surface Tension

There are many phenomena associated with surface tension, a few of which will be described here. For example, if a drop of oil is allowed to fall on a large clean surface of water, the oil will spread out over this surface until it is 1 molecule thick; it forms a monomolecular layer. If the volume of the original oil drop is known, and if the area of the monomolecular layer is measured, its thickness can be computed. This thickness will then give us one dimension of the oil molecule. Studies of the structure of oil molecules by means of x-rays show that they are not spherical but have one long dimension and two shorter dimensions. In a monomolecular layer, the molecules stand on end, so that a measure of the thickness of the oil film will yield the longest dimension.

If two small wooden matchsticks are floated near each other on the surface of water, they will be pulled toward each other. The liquid between the matchsticks is raised to a level higher than the rest of the surface by the surface tension; the pressure in the liquid between the sticks is thus decreased to a value less than atmospheric pressure. If a drop of alcohol is placed between them, the two matchsticks will be pulled apart. The effect of the alcohol is to decrease the surface tension; that is, the surface tension

of the solution of alcohol in water is less than that of pure water. The difference in the surface tensions of the two sides of each stick supplies the forces which pull them apart.

When small pieces of camphor are dropped onto a clean surface of water, these pieces will perform very erratic motions. Although camphor is soluble in water only to a very slight extent, wherever it is dissolved, the surface tension is reduced. Each little piece of camphor will experience forces caused by the different surface tensions around it, and these forces set the particle in motion. The motion will cease when the surface tension becomes uniform and equal to that of the solution of camphor in water.

Problems

13-1. A copper wire 80 cm long and 0.25 cm in diameter is suspended from a rigid framework. A body whose mass is 5 kg is hung at the end of the wire. Determine (a) the stress in the wire and (b) the strain produced.

13-2. A steel wire 1.5 m long and 0.04 cm in diameter supports a cylinder whose mass is 4.0 kg. Determine (a) the stress in the wire, (b) the strain produced, and (c) the elongation of the wire.

13-3. A brass wire 3.0 ft long and 0.04 in. in diameter supports a body whose weight is 3 lb. Determine (a) the stress in the wire, (b) the strain produced, and (c) the increase in length of the wire.

13-4. A brass wire 4.0 m long and 2.0 mm in diameter is suspended from a hook in a beam in the ceiling. A cylinder whose mass is 6.0 kg is hung from the other end. Determine the increase in length of the brass wire.

13-5. An aluminum wire 200 cm long and 0.5 mm in diameter has a series of cylinders hung from it in succession. Each cylinder has a mass of 10 gm. The measured changes in length expressed in centimeters, as determined by a telescope and scale method, are: 0.0014, 0.0029, 0.0042, 0.0056, and 0.0070. Plot a graph with the stress as ordinate and the strain as abscissa; from the slope of this graph, determine Young's modulus for this aluminum wire.

13-6. A steel rod 6.0 in. long and 0.5 in. in diameter is to be used as a piston in a cylinder to produce a pressure of 2,000 lb/in.^{2}. Determine the decrease in length of the rod produced by this stress.

13-7. Glycerin is subjected to a pressure of 850 atm. Determine the percentage change in its volume.

13-8. Determine the bulk modulus of an oil if a volume of 1,000 cm^3 shows a decrease in volume of 0.3 cm^3 when subjected to a pressure of 12 atm.

13-9. A cube of copper 5 cm on an edge is subjected to two oppositely directed shearing forces along two of its faces. Each force is 900 nt. Determine (a) the shearing stress, (b) the shearing strain, and (c) the angle, in degrees, through which the cube has been sheared.

13-10. The maximum permissible design stress for an elevator cable is 10,000 lb/in.^{2}. What diameter cable should be used for an elevator weighing 2 tons when fully loaded, if the acceleration is to be 10 ft/sec^2?

13-11. A brass rod and a copper rod, each 2 ft long and 1 in. in diameter, are

joined at one end, and the combined rod is subjected to a compressive force of 10,000 lb. Find (a) the stress in each rod and (b) the strain in each rod.

13-12. What is the density of water at a point where the pressure is 100 atm if the density at sea level is 1 gm/cm^3?

13-13. Suppose that a square specimen of cross-sectional area A is placed in tension by equal and opposite forces F applied to its ends. Calculate the tensile stress (stress component perpendicular to the face) and the shear stress (stress component parallel to the face) across a plane face inclined at an angle θ to the normal cross section of the specimen. (a) At what angle is the tensile stress a maximum? (b) At what angle is the shear stress a maximum?

13-14. Three capillary tubes of diameters 0.5 mm, 1.0 mm, and 1.5 mm, respectively, are supported in a jar of water. Determine the height to which the water will rise in each of these tubes.

13-15. Three holes of diameters 1.0 mm, 1.5 mm, and 2.0 mm, respectively, are bored in a block of paraffin. The paraffin is partly immersed in water. Determine the level of the water in each hole.

13-16. A capillary tube 1.0 mm in diameter is placed in a soap solution of density 1 gm/cm^3. The liquid in the tube rises to a height of 0.45 cm above the level of the rest of the surface. Determine the surface tension of this solution, assuming the contact angle to be zero.

13-17. A soap film is formed on a rectangular frame 2 cm by 8 cm, as in Figure 13-12. (a) Determine the force that the film exerts on the shorter wire. (b) If this wire is moved through a distance of 5 cm, determine the amount of work done. Assume that the temperature remains constant in this process.

13-18. Calculate the gauge pressure inside a raindrop which is 4 mm in diameter. Assume the temperature to be 20°C.

13-19. Calculate the gauge pressure inside a drop of mercury whose temperature is 25°C and whose diameter is 4 mm.

13-20. Determine the gauge pressure inside a soap bubble which is 5 cm in diameter. Assume the temperature to be 20°C.

13-21. Two rectangular glass plates are spaced 1 mm apart. They are partly immersed in a dish of water at 20°C, with the plates placed so that the air space between them is in a vertical plane. Determine how high the water will rise in this air space above the level of the water in the dish. [NOTE: Consider the forces acting on a surface film 1 cm wide in contact with each plate. Balance these forces against the weight of water lifted through a height h.]

13-22. A hollow glass tube has a soap bubble of 5 cm diameter formed on one end and another soap bubble of 2 cm diameter formed on the other end. Determine the pressure difference at the ends of the tube. Explain what will happen as a result of this pressure difference.

13-23. Two glass plates, each having a large surface, are clamped together along one edge and separated by spacers a few millimeters thick along the opposite edge to form a wedge-shaped air film. These plates are then placed vertically in a dish of colored liquid. Calling x the horizontal distance measured from the edge where the thickness of the air film is zero, show that the vertical distance y through which the liquid rises in the air space varies inversely as x. [NOTE: The thickness of the air film increases as the distance x increases.]

Part Two

HEAT

14

Temperature

14-1 Concept of Temperature

Temperature is one of the fundamental concepts of physics. We are all able to recognize that some bodies are hotter than others, but our temperature sense is qualitative rather than quantitative and is capable of only a limited range. The sense of touch can frequently be used to distinguish between hotter and colder objects, provided that these lie in a temperature range consistent with the stability of human tissue. Even within this range the sense of touch is often unreliable as a measure of temperature. The metal bracket holding a wooden rail may feel much colder to the touch than the railing itself, even though both are at the same temperature.

It is a matter of common observation that some of the physical properties of many substances are altered when the temperature is changed. The volume or the pressure of a gas increases when the temperature is raised. The length of a copper rod changes with changing temperature. Some of the electric and magnetic properties of substances vary with changes in temperature. The changes that take place in these physical properties can be used to measure the changes in temperature which produced them.

In order to make a measurement of the temperature of a body, it is important to be able to decide when two bodies are at the same temperature. Suppose that the length of a copper rod is measured while it is in the laboratory and exposed to the air in the room; if the rod is then put into a mixture of ice and water, its length will first decrease and then reach a new value which will remain constant as long as it is in the ice-water mixture. We then say that the temperature of the copper rod is the same as that of the ice-and-water mixture. The two systems, the copper rod and the ice-water mixture, are in thermal contact and have reached *thermal equilibrium.* Thus *two systems are said to be at the same temperature when they are in*

thermal equilibrium. We may thus consider temperature to be a property of a system which determines whether or not it is in thermal equilibrium with any other system that is placed in thermal contact with it.

14-2 Thermometers

A physical device which measures temperature is called a *thermometer.* In order to construct a thermometer, it is necessary to choose some thermometric property of a system whose value depends upon its temperature. Although the length of a solid rod is seldom used as this thermometric property, the length of a liquid column in a glass tube is used very frequently. In this liquid-in-glass thermometer, the difference in volume expansion between the liquid and the glass container is visible as a change in length of the liquid column when it is allowed to expand into a very fine capillary tube attached to the glass bulb containing most of the liquid, as shown in Figure 14-1.

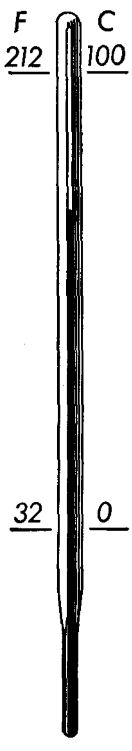


Fig. 14-1 A mercury-in-glass thermometer with the ice point and the steam point marked in both the Fahrenheit and Celsius scales.

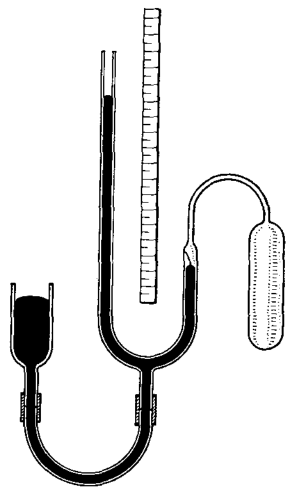


Fig. 14-2 Constant-volume gas thermometer.

The *constant-volume gas thermometer,* in which the pressure of the gas is the thermometric property, provides an extremely accurate measurement of temperature. This thermometer, shown in Figure 14-2, utilizes the

change in pressure of a gas with temperature, when the gas is maintained at constant volume. Gases customarily used are air, hydrogen, or helium.

The color and intensity of the light emitted from a furnace vary with the temperature. An *optical pyrometer* is a device in which the temperature of a furnace is measured by comparing the light emitted from the furnace with the light emitted by an electrically heated filament which has been previously calibrated. The character of the light emitted by an incandescent body is a thermometric property suitable for use in the determination of temperature.

The electrical resistance of a wire changes with temperature. As we shall see in a subsequent chapter, the electrical resistance of a metal is an easily measured property. This thermometric property provides the basis for the *resistance thermometer.*

If the ends of two wires made of different metals or alloys are connected together and their junctions maintained at different temperatures, it is found that an electric current flows in the wire loop. The current is produced by an electromotive force whose value depends upon the difference in temperature of the junctions. This thermometric property provides the basis of the *thermocouple* type of thermometer. When a number of small thermocouples are connected together, the effect is enhanced, and the result is called a *thermopile.*

14-3 Temperature Scales

In order to be able to assign a number to the temperature of a body, it is necessary to agree upon a method for setting up a temperature scale. We must first choose some thermometric property of a system, for example, the length of a mercury column in a glass capillary tube, or the electrical resistance of a platinum wire, or the pressure of a gas kept at a constant volume. Let us call the value of the chosen thermometric property X and let T be the temperature of a system that surrounds this thermometer and is in thermal contact with it. When thermal equilibrium is reached, the thermometer and the surrounding system are at the same temperature. If the temperature of the system is changed, the value of the thermometric property of the thermometer is also changed. The temperature, as indicated by the thermometer, will be some function of X; let us call this function $T(X)$. Let us assume that $T(X)$ is a linear function of X, thus

$$T(X) = aX,$$

where a is a constant of proportionality.

Thus, if at one temperature the value of the thermometric property is X_1, and if at a higher temperature the value of this thermometric property

is X_2, then we can write

$$\frac{T(X_1)}{T(X_2)} = \frac{X_1}{X_2}. \tag{14-1}$$

Before 1954 two *fixed points* were used to assign numbers in calibrating a thermometer. Since 1954, by international agreement, only one fixed point is being used. Since the older scales are still in common use, we shall describe the earlier method first.

A *fixed point* on a temperature scale is a number that is assigned to the temperature of an easily reproducible state of a system. Before 1954 the two fixed points were (a) the temperature of a system consisting of a mixture of ice in equilibrium with water open to the air at standard atmospheric pressure with the water saturated with air (called the *ice point*) and (b) the temperature of steam in equilibrium with pure water at standard atmospheric pressure (called the *steam point*).

To measure the temperature of a system, the thermometer is put in thermal contact with it; the value X of the thermometric property at this temperature $T(X)$ is measured. Calling X_i the value of this property at the ice point, we have

$$\frac{T(X_i)}{T(X)} = \frac{X_i}{X}. \tag{14-2}$$

Similarly, if X_s is the value of the thermometric property at the steam point, then

$$\frac{T(X_s)}{T(X)} = \frac{X_s}{X}. \tag{14-3}$$

From the above equations we get

$$\frac{T(X_s) - T(X_i)}{T(X)} = \frac{X_s - X_i}{X}, \tag{14-4}$$

so that

$$T(X) = \frac{T(X_s) - T(X_i)}{X_s - X_i} X. \tag{14-5}$$

On the Celsius scale of temperature, also called the centigrade scale of temperature, first devised by A. Celsius (1701–1744), the temperature interval $T(X_s) - T(X_i)$ is assigned the value of 100°C, the ice point is called 0°C, and the steam point 100°C, the interval being divided into 100 equal divisions called degrees Celsius.

On the Fahrenheit scale, first devised by G. Fahrenheit (1686–1736), the interval between the ice point and the steam point is assigned the value 180°F, the ice point is called 32°F, and the steam point 212°F, the interval being divided into 180 equal divisions called degrees Fahrenheit.

As the accuracy of temperature measurements increased, discrepancies arose principally because of the difficulty of reproducing the ice point. It will be recalled that the ice point is the temperature of an equilibrium mixture of ice and air-saturated water at atmospheric pressure. The discrepancies between measurements made at standardizing laboratories amounted to as much as 0.04°C, whereas the accuracy of measurement in this temperature region was good to about 0.001°C. In 1954 it was decided to change the method of calibrating thermometers and to use only *one fixed point;* this fixed point is the temperature of an equilibrium mixture of ice, water, and water vapor; this state is called the *triple point of water* and exists at only one definite temperature and pressure (see Chapter 17). The number chosen for this fixed point is 273.16 degrees Kelvin, written as 273.16°K. (This is in honor of Lord Kelvin who developed the thermodynamic scale of temperature. See Chapter 19.) Thus if X_3 is the value of the thermometric property of the thermometer at the triple point, then from Equation (14-1)

$$\frac{T(X)}{T(X_3)} = \frac{X}{X_3}, \tag{14-6}$$

and since

$$T(X_3) = 273.16\text{°K}, \tag{14-7}$$

we have

$$T(X) = 273.16\text{°K}\,\frac{X}{X_3}\,. \tag{14-8}$$

The triple-point cell used by the National Bureau of Standards is illustrated in Figure 14-3. Very pure, air-free water is introduced into the cell which is then sealed. The cell is then cooled in a thermos flask by an ice bath until ice, water, and water vapor are present simultaneously in the cell, indicating that the triple point of water has been attained.

If the constant-volume gas thermometer is used for measuring temperature, the thermometric property that is measured is the pressure P of the gas. For this case Equation (14-8) becomes

$$T(P) = 273.16\text{°K}\,\frac{P}{P_3}, \tag{14-9}$$

where P_3 is the pressure of the gas at 273.16°K or the triple-point temperature. As the result of many careful experiments, we find that the value of the temperature of a particular system depends upon the nature of the gas that is used in the gas thermometer. However, as smaller and smaller amounts of gas are used in the thermometers, so that the pressure of the gas gets smaller and the density gets smaller, all gas thermometers give the same result for the temperature T of the system. Thus

$$T = \underset{(\text{density}\to 0)}{273.16\text{°K}} \left(\frac{P}{P_s}\right). \tag{14-10}$$

We shall consider some of the interesting properties of gases in this and later chapters.

The relationship between the Celsius scale and the Kelvin scale can now be stated in terms of the single fixed point. On the Celsius scale the triple point is defined as 0.01°C; any temperature t on the Celsius scale can be defined in terms of the temperature T on the Kelvin scale as

$$t = T - 273.15. \qquad \textbf{(14-11)}$$

The relationship between the temperature of a body expressed on the Celsius scale t_c and its temperature expressed on the Fahrenheit scale t_f can be found by plotting the Fahrenheit temperature as ordinate and the Celsius temperature as abscissa on rectangular coordinate paper, as shown in Figure 14-4. The curve relating these two temperatures is a straight line of slope $\frac{180}{100}$, which intercepts the Fahrenheit axis at 32°F. Applying the slope-intercept form of the equation of a straight line

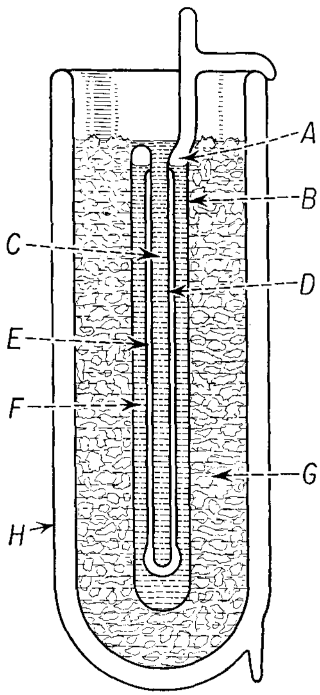


Fig. 14-3 Diagram of the National Bureau of Standards triple-point cell, BD, in use in an ice bath G, within a thermos flask H. A, water vapor; C, thermometer well; E, ice mantle; F, liquid water.

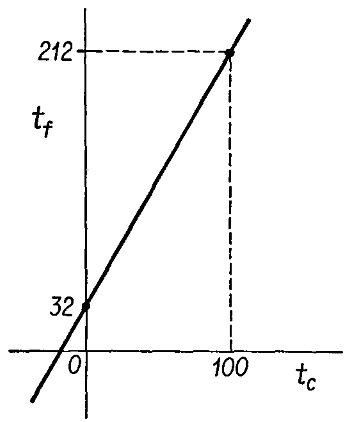


Fig. 14-4 Graph of Fahrenheit temperature versus Celsius temperature.

$$y = mx + b,$$

where m is the slope and b is the y intercept, we find

$$t_f = \tfrac{9}{5}t_c + 32. \tag{14-12}$$

This equation may be solved for t_c to yield

$$t_c = \tfrac{5}{9}(t_f - 32). \tag{14-13}$$

Illustrative Example. Find the centigrade temperature at which a centigrade thermometer will read the same number as a Fahrenheit thermometer.

The relationship between the reading of a centigrade thermometer and a Fahrenheit thermometer is always given by Equation (14-12) as

$$t_f = \tfrac{9}{5}t_c + 32.$$

The additional condition imposed by the problem may be expressed analytically as

$$t_f = t_c.$$

Solving these two equations simultaneously, we find

$$t_c = -40.$$

14-4 Thermal Expansion of Solids

A change in the temperature of a substance is nearly always accompanied by a change in its physical dimensions. The expansion which takes place in a perfect crystal when its temperature is increased depends upon the direction in which the expansion is measured relative to the crystal axis. Most crystalline solids are made up of grains within which the crystals are oriented in one direction. The crystal directions are randomly oriented from grain to grain, so that we find that the expansion of most solids is often the same in every direction relative to the crystal axis.

The change in length of a solid rod is a smooth function of temperature and may be represented by an infinite series of the form

$$L_t = L_0[1 + \alpha\,(\Delta t) + \alpha'\,(\Delta t)^2 + \cdots], \tag{14-14}$$

where L_t is the length of the rod at temperature t, L_0 is the length of the rod at some reference temperature t_0, and Δt is the temperature difference $t - t_0$. The coefficients α, α', and so on, are constants which must be evaluated by experiment for each different material. It is common practice to approximate the expansion of a solid by making use of a number of straight lines tangent to the curve which describes the length of a rod as a function of temperature, as shown in Figure 14-5. In this case only the first two terms of the right-hand side of Equation (14-14) are required to represent any one of these tangent lines, and the equation becomes

$$L_t = L_0[1 + \alpha(\Delta t)], \tag{14-15}$$

where α is called the *coefficient of linear expansion* at the temperature t_0. Since the slope of each of the tangent lines in Figure 14-5 depends upon the reference temperature at which it is constructed, the coefficient of linear expansion depends upon the reference temperature at which it is evaluated.

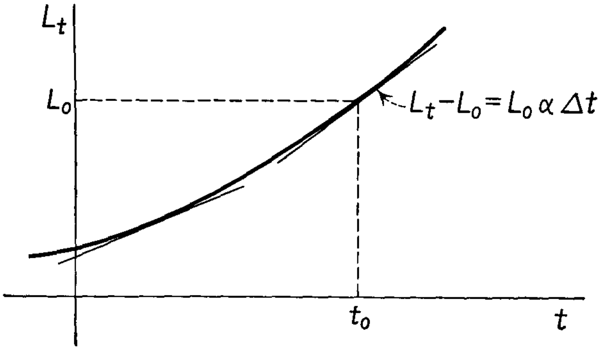


Fig. 14-5

The slope of the tangent line at temperature t_0, at which the length of the rod is L_0, is given by $L_0\alpha$. If we write the change in length of the rod as

$$\Delta L = L_t - L_0,$$

we may express the coefficient of linear expansion as

$$\alpha = \frac{\Delta L/L_0}{\Delta t}\,. \tag{14-16}$$

In other words, the coefficient of linear expansion is the fractional change in length of the rod per degree of temperature change. The units of α are therefore (degrees)$^{-1}$. Since a Fahrenheit degree represents a smaller temperature interval than a centigrade degree, the coefficient of linear expansion per degree Fahrenheit is five ninths of the coefficient of linear expansion per degree centigrade. The linear-expansion coefficients of several solids are given in Table 14-1.

If two metals, say brass and steel, are welded or riveted together so that they form a single straight piece at room temperature, then, when the temperature is raised, the strip will bend in the form of an arc with the brass on the outside, as shown in Figure 14-6. This is due to the fact that brass has a greater coefficient of expansion than steel. A bimetallic strip of this kind is frequently used in thermostats. The strip is fixed at one end, and the bending of the free end may be used to actuate a switch at some predetermined temperature.

A solid in which the expansion is the same in every direction is said to be *isotropic* with regard to this property. When the temperature of an

TABLE 14-1 COEFFICIENTS OF LINEAR EXPANSION AND VOLUME EXPANSION

Substance		α
Aluminum	(20–300°C)	25.7×10^{-6} per °C
Brass	(0–100°C)	19.3
Copper	(25–300°C)	17.8
Pyrex glass	(21–470°C)	3.6
	(550–570°C)	15.1
Invar	(20°C)	0.9
Lead	(18–100°C)	29.40
Platinum	(40°C)	8.99
Steel	(40°C)	13.2
Tungsten	(0–100°C)	4.3
	(1000–2000°C)	6.1
Substance		β
Mercury	(0–100°C)	181.8 per °C

isotropic solid is changed, the length of each linear dimension is changed. The area of each element of area is changed, and the volume of each volume element is changed. We may represent the expansion of area and the volume expansion by mathematical series analogous to Equation (14-14) and may approximate the expansion by use of the first two terms of the series, as before. To represent the area expansion, we write

$$A_t = A_0[1 + \sigma(\Delta t)], \tag{14-17}$$

where σ (sigma) is the coefficient of area expansion at the temperature t_0. Similarly, to represent the volume expansion we write

$$V_t = V_0[1 + \beta(\Delta t)], \tag{14-18}$$

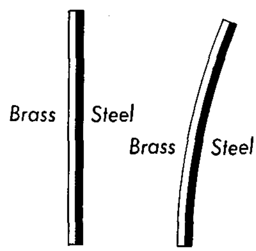


Fig. 14-6 Bending of a bimetallic strip when heated.

where β (beta) is the coefficient of volume expansion at the temperature t_0. The coefficient of area expansion and the coefficient of volume expansion may be expressed by equations analogous to Equation (14-16) as

$$\sigma = \frac{\Delta A/A_0}{\Delta t}, \tag{14-19}$$

and

$$\beta = \frac{\Delta V/V_0}{\Delta t}. \tag{14-20}$$

We may relate the coefficient of linear expansion to the coefficient of area expansion by considering the expansion of a square plate of dimensions

$L_0 \times L_0$ at temperature t_0. The area of this plate at temperature t_0 is L_0^2. When the temperature of the plate is t, the length of a side is L_t, given by Equation 14-15, and the new area of the plate is

$$\begin{aligned} A_t &= L_t^2 = \{L_0[1 + \alpha(\Delta t)]\}^2 \\ &= L_0^2[1 + 2\alpha\,(\Delta t) + \alpha^2\,(\Delta t)^2], \\ A_t &= A_0[1 + 2\alpha\,(\Delta t) + \alpha^2\,(\Delta t)^2]. \end{aligned}$$

In general, the coefficient of linear expansion is a small number. Over the limited temperature interval for which the coefficient of linear expansion represents the true expansion of the body, we may neglect the term in α^2 as being small in comparison with the other terms on the right-hand side of the equation. Thus we have

$$A_t = A_0[1 + 2\alpha\,(\Delta t)].$$

Comparing the above result with Equation (14-17), we see that the coefficient of area expansion σ is numerically equal to twice the coefficient of linear expansion α. In the form of an equation we have

$$\sigma = 2\alpha. \tag{14-21}$$

Similarly, we may find that the coefficient of volume expansion β is numerically equal to three times the coefficient of linear expansion, or

$$\beta = 3\alpha. \tag{14-22}$$

Coefficients of areal expansion and of volume expansion of solids are not usually tabulated in compilations of technical data, for the reason that they may be readily obtained from the tabulated coefficients of linear expansion.

Illustrative Example. A hollow copper sphere has an inner radius of 4 cm and an outer radius of 5 cm at a temperature of 20°C. Find the change in volume of the spherical cavity when the temperature is raised to 420°C.

The change in volume of a copper sphere is the same whether it is solid or hollow. We may think of a solid sphere as consisting of a solid central core and a hollow spherical shell whose inside diameter is the same as the diameter of the core. The two parts fit perfectly at all temperatures. Thus the volume of a spherical cavity at any temperature is the same as the volume of a solid sphere of copper at that temperature. To find the change in volume of a spherical cavity in a hollow copper sphere, we find the change in volume of a solid sphere of copper subjected to the same temperature change.

The coefficient of linear expansion of copper valid in the range 0–625°C has been measured as $16.07 \times 10^{-6}/°C$. The change in volume may be obtained from Equation (14-20). The numerical values are

$$\begin{aligned} \beta &= 3\alpha = 48.21 \times 10^{-6}(°C)^{-1}, \\ V_0 &= \tfrac{4}{3}\pi r^3 = 268 \text{ cm}^3, \\ \Delta t &= 400°C. \end{aligned}$$

Thus
$$\Delta V = \beta V_0 \Delta t,$$
$$\Delta V = 5.17 \text{ cm}^3.$$

Thus the spherical cavity expands from a volume of 268 cm^3 to a volume of 273.2 cm^3.

14-5 Thermal Expansion of Liquids

As a general rule, a liquid expands when its temperature is raised. The notable exception to this is water, which contracts when its temperature

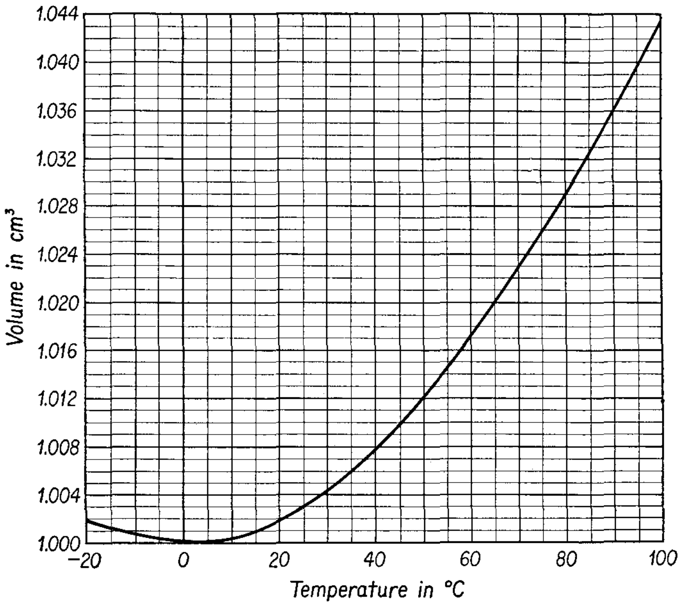


Fig. 14-7 Curve showing the volume of water as its temperature is raised from −20°C to 100°C with its minimum volume or maximum density at 4°C.

is raised in the limited region from 0°C to 4°C. Above 4°C water expands with an increase in temperature. The behavior of water at atmospheric pressure in the range from −20°C to 100°C is shown in Figure 14-7. (NOTE: Water is a supercooled liquid from −20°C to 0°C.)

One method for determining the coefficient of volume expansion of a liquid is to put the liquid into a container of known volume provided with a narrow tube at one end, as shown in Figure 14-8. A glass container is usually employed. The level of the liquid in the tube is observed at the initial temperature t_0 and at the final temperature t. The surface of the liquid is exposed to the atmosphere so that the pressure of the liquid remains

constant. Since liquids generally have greater coefficients of expansion than glass, the level of the liquid will rise as the temperature is raised. Only the relative expansion of the liquid with respect to the container can be directly determined by this method. If the coefficient of volume expansion of the container is known, the coefficient of volume expansion of the liquid

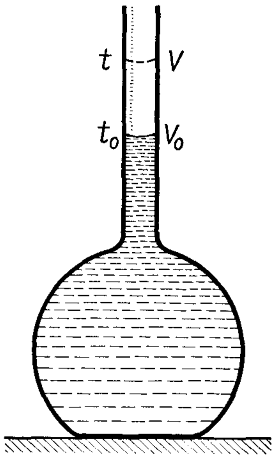


Fig. 14-8 Expansion of a liquid in a container when heated.

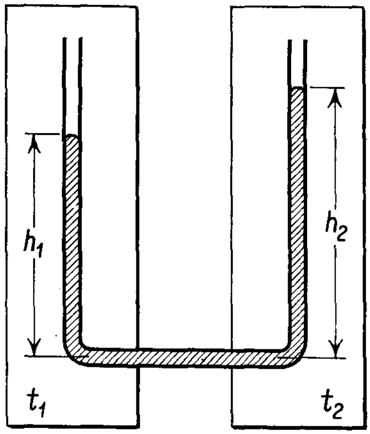


Fig. 14-9 Dulong and Petit apparatus.

can be determined. The apparent change in volume of the liquid is the difference between the change in volume of the liquid and the change in volume of the container.

The expansion of a liquid may be determined by a method, introduced by Dulong and Petit in 1817, in such a way that the measurement does not depend upon the expansion of the container. In its most elementary form the apparatus consists of two vertical tubes maintained at different temperatures. The two vertical tubes are connected by a horizontal tube, as shown in Figure 14-9. At the level of the horizontal tube, the pressure must be the same in both columns. The observed difference in the height of the liquid in the two columns is due to the difference in density of the liquid at the two temperatures.

14-6 Expansion of Gases

In studies of the expansion of solids or liquids, the substance studied is generally maintained in an open system, exposed to the atmosphere, so that the expansion coefficient is determined at constant pressure—the pressure

of the atmosphere. Since a gas must be studied in a closed container, a change in temperature may involve a change in both the volume and the pressure of the gas. It is customary to study the behavior of the gas at constant pressure, where the volume changes with temperature, or at constant volume, where the pressure changes with tempeature.

An apparatus suitable for studying the change in pressure of a gas at constant volume is shown in Figure 14-10. The gas under investigation is kept in the bulb A and in the narrow tube leading to the manometer. The mercury in tube C of the manometer is always kept at the same level by being raised to the index point I by raising or lowering the tube B. The U-shaped section of the tube R is frequently a flexible rubber hose.

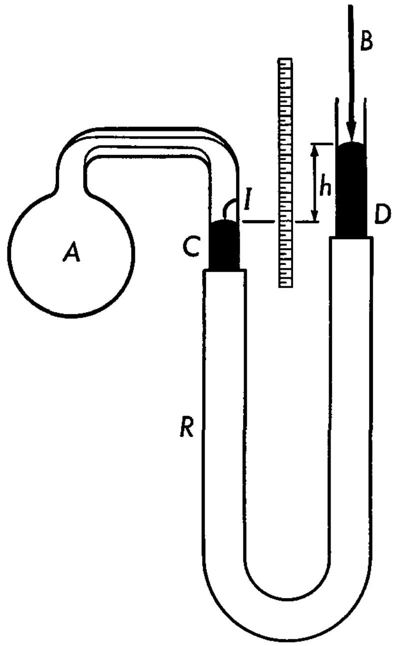


Fig. 14-10

Let us define the *coefficient of pressure change at constant volume* β' as

$$\beta' = \frac{\Delta P/P}{\Delta t}. \qquad \text{(14-23)}$$

When the coefficient of pressure change at constant volume of hydrogen is measured at 0°C, this coefficient has the value of approximately 0.00366 per degree centigrade. A value very close to this is obtained with other gases, provided that the pressure is not too high and the temperature is significantly above the temperature at which the gas liquefies. The interpretation of this numerical value is that the pressure of hydrogen at constant volume will change by 0.00366 or 1/273.2 of its pressure at 0°C for each centigrade degree change of temperature. The use of a constant-volume gas thermometer has already been discussed in Sections 14-2 and 14-3.

By a slight modification of the apparatus pictured in Figure 14-10, the pressure of the gas may be kept constant, and its volume may be measured as a function of the temperature. The *coefficient of volume change at constant pressure* β may be defined as

$$\beta = \frac{\Delta V/V}{\Delta t}. \qquad \text{(14-24)}$$

When the coefficient β is measured for hydrogen at 0°C, its value is again

found to be 0.00366 per degree centigrade. The coefficient of volume change at constant pressure of other gases is close to this value, provided that the pressure is not too high and the gas is far from the temperature at which it becomes liquid.

14-7 Absolute Scale of Temperature

The fact that the constants β' and β have the same numerical value for hydrogen, and that these constants are approximately the same for other gases, has led to the introduction of a scale of temperature known as the *absolute gas scale of temperature.* From Equations (14-23) and (14-24) we see that, if hydrogen is maintained at constant volume, its pressure should be equal to zero at a temperature of $-273.2°C$. Similarly, if hydrogen is kept at constant pressure, its volume should be zero at a temperature of $-273.2°C$, provided that it remains a gas.

Let us rewrite Equation (14-23) in terms of the initial pressure P_i, the final pressure P_f, the initial temperature t_i, and the final temperature t_f. We find

$$\beta' = \frac{(P_f - P_i)}{P_i(t_f - t_i)}.$$

If we take the initial temperature as

$$t_i = 0°C,$$

and set

$$\beta' = 1/273.2 \text{ per degree centigrade,}$$

we find, after some algebraic manipulation,

$$\frac{P_f}{P_i} = \frac{t_f + 273.2}{273.2}.$$

Let us define the absolute centigrade gas temperature T as the centigrade temperature plus 273.2, or

$$T = t + 273.2.$$

For the case chosen the initial absolute temperature is $T_i = 273.2$. Thus we have

$$\frac{P_f}{P_i} = \frac{T_f}{T_i},$$

or

$$P = KT, \tag{14-25}$$

where K is some constant for a particular quantity of gas maintained at a constant volume. From Equation (14-24) we may find, by a similar derivation, that

$$V = K'T. \tag{14-26}$$

Equation (14-26) is called *Gay-Lussac's law.*

From the basic conception of the absolute gas scale of temperature, we see that negative numbers for the absolute temperature would lead to such absurd results as the existence of negative pressures and negative volumes for a gas. Hence the temperature of a body expressed on the absolute scale must always be a positive number. It is not possible at this point to say whether there is a physically achievable absolute zero of temperature. All substances become liquids at temperatures above the absolute zero. The lowest temperature measured with a gas thermometer is about $-272°C$, or $1°$ abs, using helium at low pressure.

It is apparent from the preceding discussion that what is needed is a temperature scale which is independent of the properties of a particular substance. There is such a scale, known as the *absolute thermodynamic scale*, or *Kelvin scale of temperature*, which will be discussed in Chapter 19. There we shall see that temperatures on the thermodynamic scale are in agreement with temperatures on the absolute gas scale for a perfect gas, and are very close to temperatures on the absolute gas scale achieved with a constant-volume hydrogen thermometer.

Although the temperature $273.2°$ abs is sufficiently accurate for our purposes, it should be noted that the average value of the best experimental determinations of the ice point is

$$T = 273.165° \text{ abs}.$$

The lowest temperature which has been obtained experimentally is $18 \times 10^{-6}\,°$ abs. This has been done by a process involving the magnetization and demagnetization of certain magnetic materials which were previously cooled to a temperature of about $1°$ abs.

The absolute temperature can also be expressed in terms of the Fahrenheit degree by means of the equation

$$T_F = \tfrac{9}{5}T, \tag{14-27}$$

where T_F is the absolute temperature in Fahrenheit degrees, and T is the absolute temperature in centigrade degrees. When absolute temperature is measured in Fahrenheit degrees, the temperature scale is called the *Rankine* scale of temperature. This scale is used in engineering in the United States and in Great Britain.

14-8 Thermal Stresses

If a rod is heated while its ends are confined, the rod is thereby put into a condition of internal compressive stress. If the rod is cooled when its ends are confined, it assumes a condition of internal tensile stress. To calculate the stress within the rod, we assume that the rod has acquired its final configuration in a virtual two-step process, in which the first step is the

change in temperature of the free rod, while the second step is the application of forces to the ends of the rod to return it to its initial length.

When the temperature of a rod of length L, made of a material whose coefficient of linear expansion is α, is altered by an amount Δt, the change in the length of the rod is given by Equation (14-16) as

$$\Delta L = \alpha L \, \Delta t.$$

The fractional change in length of the rod is given by

$$\frac{\Delta L}{L} = \alpha \, \Delta t.$$

Let us suppose, for definiteness, that the temperature of the rod has increased, and that the rod has increased in length. We may return the rod to its original dimensions by applying a compressive stress sufficient to produce a decrease in the length of the rod equal to ΔL, or to produce a strain in the rod equal to $\Delta L/L$. From the definition of Young's modulus of elasticity, we have

$$Y = \frac{\text{stress}}{\text{strain}} = \frac{\text{stress}}{\alpha \, \Delta t} \cdot$$

Thus the stress produced in a rod whose modulus of elasticity is Y will be given by

$$\text{Stress} = \alpha Y \, \Delta t, \tag{14-28}$$

when the rod is confined so that it cannot expand or contract when it is subjected to the temperature change Δt. The state of thermal stress in a rod is determined by its coefficient of linear expansion, Young's modulus, and the temperature change to which the rod has been subjected. The force which must be applied to the ends of the rod to produce this stress is determined from the definition of stress as the force per unit area of the rod, as given in Equation (13-1).

Thermal stresses are widely encountered in practical engineering problems. Thus the steel rails of a railroad are commonly provided with expansion joints, but it is possible to lay continuous steel rail to any convenient length, provided that the rails are securely anchored to a roadbed with spikes and crossties capable of exerting a force sufficient to produce the stress given by Equation (14-28). It is not adequate to exert the force at the ends of the rails, for the rails would fail by buckling. The continuous rail must be spiked to the roadbed at suitable intervals. In laying concrete road or curbing, it is common practice to provide expansion joints because of the greater expense and difficulty of securing the road or curb to the ground. Similarly, it is difficult to provide large, long buildings with sufficient restraint, and expansion joints must be provided to keep cracks from

forming in brick walls. Steel tires are fastened to the cast-iron wheels of a railroad car by a process of heating the steel rim and allowing it to cool and contract onto the wheel. The stress in the tire exerts forces on a short element of length of the tire which have a normal component, holding the tire to the wheel by frictional forces between the tire and the wheel, as shown in Figure 14-11. Such a tight fit between two members is called a *shrink fit*.

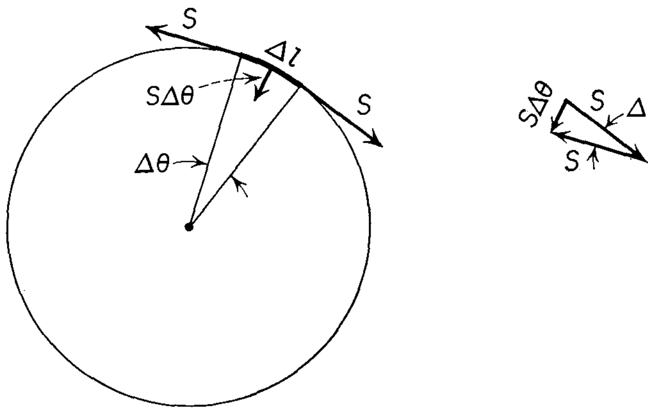


Fig. 14-11 The radial force exerted by a circular hoop under tension S upon a section of length Δl is given by $S\Delta\theta$, where $\Delta\theta$ is the angle subtended by the section Δl at the center of the circle.

When a solid is heated and subsequently cooled, as in heat treatment and quenching, or in a welded member, internal thermal stresses are often generated through the irregular cooling of the member. When hot, a solid body may easily be deformed without development of internal strains, but when the solid has cooled sufficiently, any further deformation is accompanied by the generation of internal stresses. Thus an irregularly shaped forging being quenched after heat treatment is cooled first where it is thinnest. As the thicker sections subsequently cool and contract, they are restrained by the cooler sections of the forging, and a state of internal stress is produced which may be greater than the rupture strength of the material of the forging, so that cracks develop. A drinking glass into which boiling water is poured will often crack from the thermal stresses developed at the boundary between the hot and cold portions of the glass.

14-9 The Meaning of Temperature

In the discussion of the concept of temperature in Sections 14-1 to 14-3, we have defined the meaning of temperature through the operation of

measuring temperature with the aid of thermometric properties of macroscopic systems. Thus the measurement of temperature requires that matter be present in bulk. The concept of temperature is a macroscopic one. It is meaningless to ask the question, "What is the temperature of a proton moving at a speed of 100 mi/hr?"

The definition of temperature was also based upon the concept of equilibrium; that is, it is assumed that any changes in the temperature of the system are taking place sufficiently slowly so that the state of the system could be considered constant during the time of measurement. If the state of the system changes rapidly, so that it does not have sufficient time to come to equilibrium, the meaning of temperature is somewhat ambiguous. This is the case in a flame, or in the exhaust gases of an engine, where the processes of combustion take place with great rapidity, and the gases do not have time to come to equilibrium. In such cases it is often found that different methods of measuring temperature yield different results, for the measured temperature is a function of the method of measurement as well as the condition of the object whose temperature we wish to determine, and the number assigned as the temperature has meaning only in comparative terms.

Problems

14-1. At what centigrade temperature will the reading of a Fahrenheit thermometer be numerically equal to twice the reading of a centigrade thermometer?

14-2. How long must a steel rod be in order that its length will increase by 0.02 in. as a result of a temperature change of 10°C?

14-3. The distance between two markers is measured with a steel tape at 25°C. The reading of the tape is 80 ft. If the calibration of the tape is correct at 0°C, determine the distance between the markers.

14-4. A copper ring has an inside diameter of 4.98 cm at 20°C. To what temperature must it be heated so that it will just fit on a shaft 5.00 cm in diameter?

14-5. What will be the stress in the ring of Problem 14-4 after it has shrunk in place onto the shaft? Assume that the dimensions of the shaft remain constant.

14-6. A glass flask having a volume of 1 liter at 20°C is filled with mercury at that temperature. What volume of mercury will run over the lip of the flask when the temperature is raised to 100°C?

14-7. The pressure of a volume of gas at 27°C is 546 mm of mercury. What will be its pressure, expressed in millimeters of mercury, if the temperature is increased to 28°C, the volume being kept constant?

14-8. A constant-pressure air thermometer contains a mass of air whose volume is 300 cm^3 at 0°C. What will be its volume at 50°C?

14-9. The following data were taken in an experiment with a constant-volume air thermometer, such as that shown in Figure 14-10. Barometric pres-

sure 754.0 mm; height of column C 48.4 cm and height of column D 44.7 cm at the ice point; height of column C 48.4 cm and height of column D 71.0 cm at the steam point; height of column C 48.4 cm and height of column D 62.3 cm when the bulb is surrounded with warm water. Determine (a) the coefficient of pressure change of the air and (b) the temperature of the warm water.

14-10. A clock regulated by a seconds pendulum made of brass has a correct period of 2 sec when the temperature is 70°F. Determine the gain or loss, in seconds per day, when the temperature rises to 97°F. Assume it to be a simple pendulum.

14-11. Derive the equation

$$\beta = 3\alpha$$

relating the coefficient of volume expansion β of an isotropic solid to its coefficient of linear expansion α.

14-12. A steel rod 1 m long and 0.5 cm in diameter is clamped between two fixed supports at its ends. The temperature of the rod is raised 30°C. Determine (a) the stress in the rod and (b) the force exerted by each support. Young's modulus $= 20 \times 10^{11}$ dynes/cm^2.

14-13. A steel bomb is filled with water at 10°C. If the system is heated to 100°C, determine the increase in pressure of the water (a) neglecting the thermal expansion of the steel and (b) taking into account the thermal expansion of the steel. Neglect the change in dimensions of the steel bomb due to the tension it experiences. Take the bulk modulus of water as 2×10^4 atm^{-1}.

14-14. Derive a formula for correcting the reading of a mercury barometer calibrated at 20°C when the barometer is read at a temperature of t°C. Neglect the expansion of the scale attached to the barometer. Express your result in terms of the correction Δh, the reading of the barometer h, the coefficient of volume expansion β of mercury, and the temperature. Does the expansion of the glass container affect your result?

14-15. A block of aluminum is immersed in water at a temperature of 20°C, and the buoyant force on the aluminum is observed to be 10 lb. What will be the buoyant force on the block of aluminum at a temperature of 4°C?

14-16. A steel tire 2 in. in width whose inner diameter is 0.999 ft and whose outer diameter is 1.05 ft is to be heated so that it may be placed on a cast-iron wheel 1.000 ft in diameter. (a) Assuming that the cast-iron wheel is perfectly rigid, find the tensile stress in the tire when it has cooled. (b) If the coefficient of friction between the tire and the wheel is 0.2, find the force which must be applied to the tire to pry it from the wheel. Take Young's modulus for steel as 29×10^6 lb/in.^{2}.

14-17. A steel rod 1 ft long is welded at one end to a copper rod of the same diameter and length, and the two bars are mounted between rigid supports. Find the stress in each bar when their temperature is increased by 50°C. Young's modulus for steel is 30×10^6 lb/in.^{2}; for copper: 18×10^6 lb/in.^{2}.

14-18. A steel cable 0.5 in.^{2} in cross-sectional area is tightened to a tension of 20,000 lb/in.^{2} when the cable is at 0°C. What is the tension in the cable when the temperature is increased to 20°C? Young's modulus is 30×10^6 lb/in.^{2}.

14-19. Find the coefficient of volume expansion of water at (a) 40°C and (b) 70°C.

15

Heat and Work

15-1 The Nature of Heat

Until about 1750 the concepts of heat and temperature were not clearly distinguished. The two concepts were thought to be equivalent in the sense that bodies at equal temperatures were thought to "contain" equal amounts of heat. Joseph Black (1728–1799) was the first to make a clear distinction between heat and temperature. Black believed that heat was a form of matter, which subsequently came to be called *caloric*, and that the change in temperature of a body when caloric was added to it was associated with a property of the body which he called the capacity. Later investigators endowed caloric with additional properties. The caloric fluid was thought to embody a kind of universal repulsive force. When added to a body, the repulsive force of the caloric fluid caused the body to expand. To explain the liberation of heat when a block of metal was filed, it was postulated that small filings were less able to retain caloric, by virtue of their large surface area, than a block of solid metal. Attempts were made to measure the weight of caloric by trying to observe a change in the weight of a body when its temperature was raised, but these experiments were contradictory. Among others, Count Rumford (1753–1814), an American born Benjamin Thompson, who gained his title in the service of the Elector of Bavaria, found that the weight of a block of gold was unaltered by as much as 1 part in 1,000,000 when raised from the freezing point of water to bright-red heat.

As director of the arsenal at Munich, Rumford observed that the heat generated in boring cannon did not depend upon the number of chips evolved. He performed experiments which showed that the heat capacity of brass chips was the same as that of solid brass, contradicting the theory advocated by the adherents of the caloric theory of heat. He concluded that the evolution of heat during the operation of boring a cannon could not be attributed to the liberation of caloric, and he set up an experiment in which the heat evolved in rubbing a blunt horse-driven drill against a

brass casting was used to heat a known quantity of water. Rumford was able to trace the heat evolved to the effort of the horse rather than to the ability of brass filings to retain caloric fluid.

It was not until the middle of the nineteenth century that the presently accepted view of the nature of heat evolved. In 1842 J. R. Mayer laid the foundation of the principle of the conservation of energy, proposing that heat was a form of energy. He utilized some well-known data on the heat necessary to maintain the temperature of expanding air, which normally cools on expansion, to compute a numerical relationship between units of heat and units of mechanical energy which is called the *mechanical equivalent of heat.* The suggestion of Mayer was investigated in great detail by James Prescott Joule (1818–1889), who measured the mechanical equivalent of heat by every experimental means at his disposal. Joule utilized mechanical energy to churn water and mercury. He generated heat frictionally and electrically. The experiments of Joule firmly established the concept that heat was a form of energy and that an equivalent quantity of heat was always obtained when a given quantity of mechanical energy was expended.

We have already shown that temperature is a property of a system which determines whether or not it is in equilibrium with any other system placed in thermal contact with it. If two systems which are at different temperatures are placed in thermal contact, energy will flow from the one at higher temperature to the one at lower temperature. We shall use the term *heat* to designate the energy in transit owing to a difference in temperature between two systems. The details of the transformations which occur when this energy enters or leaves a system form an important part of physics, and we shall consider many of these transformations in this book. Some of these will be considered from the macroscopic or large-scale point of view, while other transformations will be considered from the microscopic point of view, involving some ideas concerning the structure of matter.

We have already outlined briefly (Section 13-1) our present view concerning the structure of matter with particular reference to the forces existing between molecules. Another approach to the subject of the structure of matter is in terms of the energy of the constituent particles, not only of the molecules of the substance but also of their atoms and of the electrons, protons, and neutrons which constitute these atoms. Each of these constituent particles undoubtedly has energy, probably both kinetic and potential. A complete theory of the structure of matter would enable us to evaluate the energy of each particle under all possible conditions, and from this we would then be able to evaluate the energy of the substance under any given set of physical conditions. This energy constitutes the *internal energy* of the substance under these conditions. However, there is

at present no satisfactory theory which will enable us to evaluate the internal energy of a substance. There is, however, another approach by which we can evaluate *changes in the internal energy* of a substance. For example, if a known quantity of heat is added to a substance and no external work is done on or by this substance, then this quantity of heat will be converted into additional internal energy of the substance. The complete theory of the structure of matter would be able to trace the changes in the energies of the individual particles of the substance corresponding to the quantity of heat added. However, the theory of the structure of matter in its present state is unable to provide us with a detailed analysis of how the internal energy of a substance is distributed among its constituent particles. But we can say that if U_i is the internal energy of the substance before a quantity of heat Q is added to it, and if U_f is the internal energy of the substance after the addition of this amount of heat, then

$$Q = U_f - U_i, \tag{15-1}$$

providing no external work was done on or by the substance.

One might inquire as to what evidence we have that the internal energy has been changed by the addition of heat to a substance. The evidence is of an indirect kind. We may note that the temperature of the substance has been changed, or that its physical state has been changed—say from a liquid to a gas—or that its pressure has been changed. A complete theory would relate the energies of the particles of the substance to its temperature, pressure, and physical state. To date, we have made only small progress in this direction. Probably the greatest progress has been made in the study of the behavior of substances which approximate closely the behavior of *perfect* or *ideal gases.* This will be studied in greater detail in the next chapter. We may anticipate this by stating that if the assumption is made that the absolute temperature of a perfect gas is proportional to the average kinetic energy of the gas molecules, the results are consistent with the known properties of a gas. While this direct proportionality between temperature and the kinetic energy of molecules holds for an ideal gas, it does not necessarily follow that it holds for a real substance in the solid, liquid, or gaseous phase. For real substances, the relationship between temperature and the kinetic energy of the molecules is a more complicated one.

When a body is translated with uniform speed there is an ordered motion of the body and it has kinetic energy, but there is no alteration of the internal energy of the body that is associated with the random motions of the fundamental particles composing the body. Consequently the temperature of a body is unaltered by a uniform translational motion.

15-2 Units of Heat; Heat Capacity

Since heat is a form of energy, any of the energy units such as the erg, joule, or foot pound may be used to measure a quantity of heat. However, it is usually more convenient to use some other unit of heat defined in terms of one of the effects produced by the addition of heat to a body, and then to determine the relationship between this heat unit and the more common energy units.

By tradition, heat is measured in terms of the temperature changes produced in water. Several units of heat are in common use. The *calorie* (abbreviated cal) is defined as *the quantity of heat which will raise the temperature of one gram of water from* 14.5°C *to* 15.5°C. This unit of heat is called *the* 15-*degree calorie.* In biology, in dietetics, and in some branches of engineering, a unit of 1,000 calories is frequently used. This is called the *large calorie*, or *kilogram calorie* (abbreviated kilo cal), for it is *the heat required to raise one kilogram of water through the above temperature interval.* The *British thermal unit* (abbreviated Btu) is defined as *the quantity of heat which will raise the temperature of a pound mass of water from* 63°F *to* 64°F.

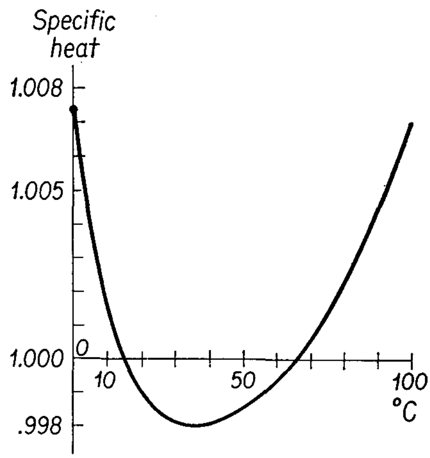


Fig. 15-1 Specific heat of air-free water as a function of temperature at 1 atm pressure.

The quantity of heat that is required to raise the temperature of a body by one degree is called its *heat capacity.* The heat capacity of a unit mass of a substance is called its *specific heat,* and is usually designated by the symbol c. The specific heat is generally stated in units of calories per gram per degree centigrade, or in units of Btu per pound per degree Fahrenheit, and has the same numerical value in both systems of units.

As shown in Figure 15-1, the specific heat of water in the interval 0°C to 100°C is not constant. The variation is of the order of $\frac{1}{2}$ per cent and will be neglected in the calculations performed in this chapter. Under these circumstances we may abridge the definitions of the calorie and the Btu, so that for practical purposes these are, respectively, the heat required to raise the temperature of a gram of water one degree centigrade, and the heat required to raise the temperature of a pound of water one degree Fahrenheit. Since there are approximately 454 gm in a pound mass, a Btu of heat will

raise the temperature of 454 gm of water by $\frac{5}{9}$°C, and is therefore approximately equal to 252 cal.

From the definitions of heat capacity and specific heat, it is evident that the heat capacity of a body may be expressed as the product of the mass of the body by the specific heat of the substance from which it is made. It is also common practice to refer to the *water equivalent* of a body as the mass of water numerically equal to the heat capacity of the body.

When a body of mass m is heated through a temperature interval Δt, the heat Q required is given by

$$\boxed{Q = mc\,\Delta t,} \tag{15-2}$$

where c is the specific heat of the body. The specific heats of some common substances are given in Table 15-1. In general, the specific heat varies

TABLE 15-1 SPECIFIC HEATS

Substance		Specific Heat $\frac{\text{cal}}{\text{gm°C}}$ or $\frac{\text{Btu}}{\text{lb mass °F}}$
Aluminum	(−250°C)	0.0039
	(−100°C)	0.1676
	(20°C)	0.214
	(100°C)	0.225
Brass	(20–100°C)	0.092
Copper	(15–100°C)	0.093
Crown glass	(10–50°C)	0.161
Ice	(−20°C)	0.48
	(−10°C)	0.53
Steam, 1 atm	(110°C)	0.481
	(120°C)	0.477
Iron	(20°C)	0.107
Lead	(20°C)	0.031
Mercury	(20°C)	0.033
Silver	(20°C)	0.056
Tin	(20°C)	0.054
Zinc	(20°C)	0.092
Ammonia (liquid)	(20°C)	1.12
Ethyl alcohol	(25°C)	0.456

with temperature; the heat required to raise a body from some initial temperature t_i to a final temperature t_f can be found by integration if c is

known as a function of temperature. Thus

$$Q = \int_{ti}^{t_f} mc\, dt.$$

For small temperature intervals, and for the calculations of this chapter, we shall neglect the variations of the specific heat with temperature.

The heat capacity of a mole of a substance is often more interesting theoretically than the heat capacity of a unit mass. The heat capacity of a mole is referred to as the *molar heat capacity* of the substance. This is given by the product of the specific heat by the molecular weight. It is interesting that, over a century ago, Dulong and Petit observed that the product of the specific heat by the atomic weight of many crystalline solids, composed of elements rather than compounds, was constant. The average atomic heat capacity of some 63 elements is about 6.15 cal/gm atomic wt, while the individual values of these substances ranged between 5.4 and 7.0 cal. Modern research has shown that these results depend on the temperature, for the specific heat of a solid approaches zero as the temperature approaches the absolute zero. There are similar generalizations which can be made regarding the molar heat capacity of gases. We shall discuss some of the implications of these generalizations in a subsequent chapter.

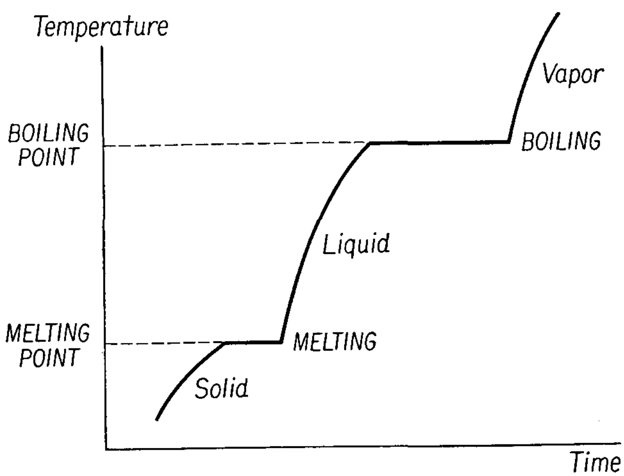


Fig. 15-2

15-3 Latent Heat

When a solid substance is heated at a uniform rate, its temperature rises steadily until its melting temperature is reached, as shown in Figure 15-2, and the substance becomes liquid. During the interval of time that the

temperature of the substance remains constant, heat is being continually absorbed. The substance is melting, and it coexists in the solid and liquid phases. In general, when the phase of a substance is altered, as from solid to liquid or from liquid to vapor, or when the crystalline form of the solid is altered, the internal energies of the two phases are different, and heat is absorbed by the substance making the phase change. An equal amount of heat is liberated when the phase change is made in the reverse direction. This heat is called the *latent heat.* The heat required to change a unit mass of substance from the solid to liquid phase at constant temperature is called the *latent heat of fusion.* If we continue to heat the substance shown in Figure 15-2, the temperature of the liquid continues to rise steadily until the boiling point is reached. Once again the rise in temperature is arrested until the liquid is completely converted to a vapor. The heat required to change a unit mass of liquid from the liquid to the vapor phase at constant temperature is called the *latent heat of vaporization.* As we will see in a subsequent chapter, the temperature at which the phase change occurs is a function of the pressure, and the latent heat associated with a particular phase change is also dependent upon the pressure and temperature at which it occurs.

For the water substance at standard atmospheric pressure, the latent heat of fusion is approximately 80 cal/gm. The latent heat of vaporization at standard atmospheric pressure is approximately 540 cal/gm. The latent heat, generally represented by the symbol L, is associated in part with the differences in internal energy of the two phases and in part with the mechanical work which must be done by the substance in pushing back the atmosphere.

15-4 Calorimetry

One of the methods used in heat measurements, or *calorimetry,* is called the *method of mixtures.* In this method a hot body at some temperature t_1 is placed in thermal contact with a second body at temperature t_2. There is a heat exchange between the two bodies, and, when equilibrium is reached at some intermediate temperature t, *the heat gained by the cooler body must be equal to the heat lost by the warmer body,* assuming that there have been no losses of heat to the outside world. One way to minimize heat losses is to arrange the experiment so that the final equilibrium temperature is approximately equal to the ambient temperature of the laboratory. Another way is to surround the bodies with an enclosure which is maintained at the average temperature of the system with external sources of heat and thermostatic controls.

Illustrative Example. Twenty grams of ice, originally at $-10°C$, are put into a copper can of mass 60 gm containing 180 gm of water at 80°C. The copper

can is called a *calorimeter.* Find the final equilibrium temperature of the system. Assume that we may neglect the loss of heat to the outside world.

Let us assume that in the final equilibrium condition all the ice has melted, and that equilibrium temperature of the system is t. To reach this state the temperature of the ice must first be raised to 0°C, and in this process the ice must gain an amount of heat given by

$$Q_1 = m_i c_i \,\Delta t_i = 20\text{ gm} \times 0.5\,\frac{\text{cal}}{\text{gm}^\circ\text{C}} \times 10^\circ\text{C} = 100\text{ cal.}$$

At 0°C the ice must acquire its latent heat of fusion to melt to water, and the heat gained in this process is given by

$$Q_2 = m_i L = 20\text{ gm} \times 80\,\frac{\text{cal}}{\text{gm}} = 1{,}600\text{ cal.}$$

Finally the ice water must be raised to a final temperature t, gaining heat in amount given by

$$Q_3 = m_i c_w \,\Delta t = 20\text{ gm} \times 1\,\frac{\text{cal}}{\text{gm}^\circ\text{C}} \times (t - 0^\circ\text{C}) = 20\,t\,\frac{\text{cal}}{^\circ\text{C}}\cdot$$

The total heat gained by the ice in reaching its final equilibrium state as water at temperature t is the sum of Q_1, Q_2, and Q_3, and is given by

$$Q_1 + Q_2 + Q_3 = 1{,}700\text{ cal} + 20\,t\,\frac{\text{cal}}{^\circ\text{C}}\cdot$$

At the same time the calorimeter can and the water must lose an equal amount of heat in being cooled from 80°C to t. The heat lost by the copper calorimeter can is given by

$$Q_4 = m_{\text{cu}} c_{\text{cu}} \,\Delta t = 60\text{ gm} \times 0.093\,\frac{\text{cal}}{\text{gm}^\circ\text{C}} \times (80^\circ\text{C} - t),$$

or

$$Q_4 = 447\text{ cal} - 5.6\,t\,\frac{\text{cal}}{^\circ\text{C}}\cdot$$

The heat lost by the water in cooling from 80°C to t is

$$Q_5 = m_w c_w \,\Delta t = 180\text{ gm} \times 1\,\frac{\text{cal}}{\text{gm}^\circ\text{C}} \times (80^\circ\text{C} - t),$$

or

$$Q_5 = 14{,}400\text{ cal} - 180\,t\,\frac{\text{cal}}{^\circ\text{C}}\cdot$$

The total heat lost by the water and the calorimeter can is

$$Q_4 + Q_5 = 14{,}847\text{ cal} - 185.6\,t\,\frac{\text{cal}}{^\circ\text{C}}\cdot$$

Finally we solve the problem by equating the heat gained to the heat lost,

$$1{,}700\text{ cal} + 20\,t\,\frac{\text{cal}}{^\circ\text{C}} = 14{,}847\text{ cal} - 185.6\,t\,\frac{\text{cal}}{^\circ\text{C}},$$

$$205.6\, t \frac{\text{cal}}{°\text{C}} = 16{,}547 \text{ cal},$$

$$t = 8.1°\text{C}.$$

Thus the final temperature of the mixture is 8.1°C.

A second calorimetric procedure, based upon the melting of ice at 0°C, is called the *ice calorimeter*. A block of ice at 0°C is carefully dried, and provided with a cavity and an ice cover. A heated specimen at known initial temperature t is placed in the cavity. When the specimen has cooled to the ice point, the water in the cavity is collected and weighed. Since the water is at a temperature of 0°C, the only heat gained by the ice which has melted is associated with the latent heat of fusion. We may write

$$m_w L = mc(t - 0),$$

where m_w is the mass of the water in the cavity, and m, c, and t are the mass, specific heat, and initial temperature of the specimen. The ice calorimeter was used by Bunsen to determine the specific heat of many metals. There is little difficulty with loss of heat from the interior cavity of the ice calorimeter, for heat exchanges with the laboratory affect the outside surface of the block rather than the interior cavity. The greatest difficulty lies in completely collecting the water which has been melted.

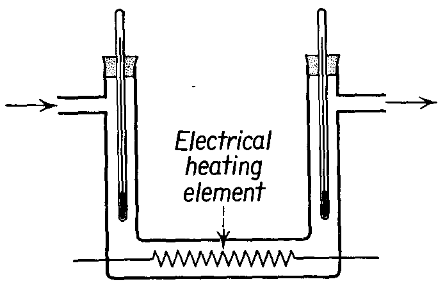


Fig. 15-3 Schematic diagram of a continuous-flow calorimeter.

The *heat of combustion* of a solid fuel, or of a food substance, may be determined in a *bomb calorimeter*. A known mass of the substance is placed in a strong container, or bomb, which has been provided with a supply of oxygen, often in chemical form, and a means of igniting the mixture. The bomb is placed in water, and the heat liberated in the combustion may be calculated by the principles of the method of mixtures. Similar procedures may be used to measure the heat of formation and the heat of solution of chemical compounds.

In the *continuous-flow calorimeter*, shown in Figure 15-3, a steady source of heat is supplied to a fluid flowing through the calorimeter. The rate of flow and the temperature of the fluid as it passes into and out of the calorim-

eter are measured. The heat of combustion of a fuel may be measured by burning the fuel in the calorimeter. If it is desired to measure the specific heat of a liquid, it is heated by an electrical heater which evolves heat at a known rate; measurement of the rate of flow of the liquid and the intake and outlet temperatures, provide the data required to calculate the specific heat.

It should be noted that the specific heat measured and used in most calorimetric experiments is an average specific heat in which the average is taken over the temperature interval used in the experiment. It is much more difficult to measure the specific heat at a particular temperature, and the methods for this measurement will not be elaborated here.

In all heat measurements it is necessary to make rather elaborate corrections for the interchange of heat with the laboratory, and to make suitable provision for good thermal contact. One cannot properly measure the temperature of a block of metal simply by placing a thermometer upon it. It is necessary to pay attention to the uniformity of temperature of the specimen, to the thermal contact, and to thermal losses from the measuring device.

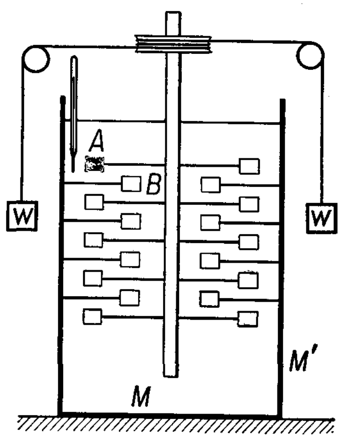


Fig. 15-4 Joule's apparatus for determining the mechanical equivalent of heat.

15-5 Work and Heat

One form of the apparatus used by Joule in his experiments to determine the mechanical equivalent of heat is illustrated in Figure 15-4. The system on which work was done was a quantity of water of mass M contained in a well-insulated glass jar of mass M'. A set of paddles A was attached to a shaft which was set in rotation by a set of weights W, connected to the paddles through a system of strings and pulleys. The motion of the paddles A past the stationary vanes B churned the water, thus increasing its temperature through a temperature interval Δt. If the weights started from rest and acquired a velocity v after falling through a height h, the mechanical energy $\mathscr{W}$ lost by the weights and delivered to the water was

$$\mathscr{W} = mgh - \tfrac{1}{2}mv^2, \tag{15-3}$$

where m is the total mass of the falling weights. The amount of heat that would have to be added to the water (and the jar) to produce the same

change in temperature is

$$Q = Mc_w\,\Delta t + M'c_g\,\Delta t. \tag{15-4}$$

Equation (15-3) gives the work done expressed in mechanical energy units, while Equation (15-4) gives the equivalent amount of heat in heat units. From this and similar experiments it is found that

$$4.185 \times 10^7 \text{ ergs} = 1 \text{ cal},$$

$$4.185 \text{ joules} = 1 \text{ cal},$$

$$778 \text{ ft lb} = 1 \text{ Btu}.$$

If we desire to equate the values of $\mathscr{W}$ and Q in an equation, it is necessary to introduce a conversion factor J representing the mechanical equivalent of heat. Thus

$$\mathscr{W} = JQ, \tag{15-5}$$

where $\mathscr{W}$ is the energy expressed in mechanical units and Q is the energy expressed in heat units. The value of J may then be expressed in appropriate units as

$$J = 4.185 \text{ joules/cal},$$

or

$$J = 778 \text{ ft lb/Btu}.$$

15-6 The First Law of Thermodynamics

We are now in a position to extend the principle of conservation of energy to include heat energy and internal energy as well as mechanical energy. When so formulated, the principle of conservation of energy is called the *first law of thermodynamics.* Let us consider that the substance or device under investigation is within an imaginary closed shell. We shall call everything within that shell the *system.* The shell may simply enclose a block of metal, or it may be so large as to enclose the entire solar system.

To develop our ideas about a suitable sign convention, let us pattern our thinking after the operation of an engine in which it is desired that the engine absorb heat from some source and do mechanical work. A quantity of heat ΔQ delivered to the system will be called positive. If heat is expelled by the system, we shall call that heat negative. A quantity of work $\Delta\mathscr{W}$ done by the system on the outside world will be positive work. Mechanical work done by the outside world upon the system will be negative. In general, the result of adding a quantity of heat ΔQ to a system is that an amount of work $\Delta\mathscr{W}$ will be done by the system on some outside agency, and the internal energy of the system will be changed by an amount ΔU.

The first law of thermodynamics can be written in the form of an equation as

$$\boxed{\Delta Q = \Delta U + \Delta \mathscr{W},} \tag{15-6}$$

where all quantities must be expressed in the same units. Unlike the usage of the suffix *dynamics* in the word "hydrodynamics," the word "thermodynamics" does not refer to heat in motion. The subject of thermodynamics deals with the relationship between heat and mechanical energy, as typified in Equation (15-6).

If no work is done on or by the system, the change in its internal energy is equal to the heat added to or removed from the system. The addition of heat to a system increases its internal energy. Evidence for the increase in internal energy is often the increase in the temperature of the system.

A physical process in which no heat enters or leaves the system is called an *adiabatic* process. In an adiabatic process $\Delta Q = 0$. Many of the processes of nature and of engineering are adiabatic or nearly so. A process which takes place inside a chamber with perfectly insulating walls is adiabatic. Many processes which take place quickly, so that there is little time for heat to be transferred, are nearly adiabatic. Thus, when air is compressed in a bicycle pump, the compression is practically adiabatic.

An interesting aspect of the first law of thermodynamics is that if we grant the impossibility of constructing a perfectly frictionless machine, Equation (15-6) asserts the impossibility of constructing a machine which will run forever; that is, a special type of perpetual-motion machine. If no mechanical work is done on or by the machine, the quantity $\Delta \mathscr{W}$ is equal to zero. Since heat is generated by friction in the bearings and is expelled from the system, the quantity ΔQ is negative. Thus the internal energy of the system must decrease, and the machine must ultimately run down.

Problems

15-1. An aluminum calorimeter whose mass is 100 gm contains 200 gm of water at 24°C. A mass of 500 gm of tin at a temperature of 95°C is placed in the water. Determine the final temperature of the mixture.

15-2. In determining the specific heat of a metal, a 150-gm sample is heated to 99.5°C and then put into 225 gm of water at 18°C which is contained in a copper calorimeter whose mass is 160 gm. The final temperature of the mixture is 22.4°C. Determine the specific heat of the sample.

15-3. A metal calorimeter whose heat capacity is 20 cal/°C contains 300 gm of water at 10°C. One hundred grams of copper at 150°C and 250 gm of zinc at 115°C are placed in the water. What is the final temperature of the mixture?

15-4. A lead ball weighing 1 lb is dropped from a height of 80 ft to the pave-

ment. Assuming that all of this energy is converted into internal energy of the lead ball, determine the rise in temperature of the ball.

15-5. The water at Niagara Falls drops 160 ft. Assuming that all of the energy of the water is converted into internal energy, what is the increase in temperature of the water after it strikes the bottom of the falls?

15-6. The powder which is used to fire an 8-gm bullet produces 800 cal of heat when burning. In the firing of the bullet, only 25 per cent of this energy is converted into kinetic energy of the bullet. Determine the muzzle velocity of the bullet.

15-7. A large glass jar whose mass in 250 gm contains 2,000 gm of water at 18°C. An electric motor delivers $\frac{1}{8}$ hp to a stirrer in the water for 10 min. Determine the rise in temperature of the water if all this energy goes into the water and the jar.

15-8. A copper cylinder whose mass is 50 gm contains 200 gm of water. The cylinder is dragged over a rough floor by a force of 100,000 dynes, and the acceleration produced is observed to be 200 cm/sec^2. (a) Determine the force of friction between the cylinder and the floor. (b) If the cylinder is moved through a distance of 100 m, determine the rise in temperature of the water and the cylinder, assuming that all of the work against friction is converted into internal energy in the cylinder.

15-9. A block of 500 gm of copper at 750°C is dropped into an ice calorimeter. How many grams of ice are melted?

15-10. A container, of water equivalent 5 gm, contains 45 gm of ice at -10°C. What will be the final condition and temperature of the material in the container if 100 gm of water at 20°C are poured into the container?

15-11. An ice crusher containing 1000 gm of ice at 0°C is driven for 15 min by a 100-watt motor. What is the final temperature of the ice, assuming that all of the energy of the motor is converted to internal energy of the ice? How much of the ice will be melted?

15-12. A container of crushed ice is at -15°C. The water equivalent of the container is 20 gm, and the ice has a mass of 200 gm. How much steam at 100°C has to be passed into the container in order that the ice be melted and the entire system, consisting of ice water, condensed steam, and container, reach a final temperature of 20°C?

15-13. A tank of water containing 100 lb of water is to be heated from 50°F to 65°F in 15 min by an electric heater. If the container is of negligible heat capacity, what is the minimum rating, in watts, of the heater which should be purchased to do this job?

15-14. A liter of water per minute flows through a tank which is heated by an electric heating element. The water flows into the tank at a temperature of 55°F and is to flow out of the tank at a temperature of 155°F. What should be the minimum rating, in watts, of the heater required to do this job?

15-15. An electric mixer is supplied with electrical energy at the rate of $\frac{1}{8}$ hp and delivers mechanical energy to a bowl of batter at the rate of 5,000 joules/min. (a) What is the efficiency of the mixer? (b) What is the increase in the internal energy of the motor in 10 min? (c) If the water equivalent of the motor is 1,000 gm, what is the increase in the temperature of the motor in 10 min?

16

Kinetic Theory of Gases

16-1 General Gas Law

The behavior of a gas under various conditions of temperature and pressure has already been studied in some detail. When the pressure of a constant mass of gas is not too great, say less than about 2 atm, we find that a gas obeys the following relationships:

$$\text{at constant temperature} \quad PV = \text{constant}; \tag{16-1}$$

$$\text{at constant volume} \quad P = KT; \tag{16-2}$$

$$\text{at constant pressure} \quad V = K'T. \tag{16-3}$$

These three equations are special cases of a single experimental equation which gives the relationship between the pressure P, the volume V, and the

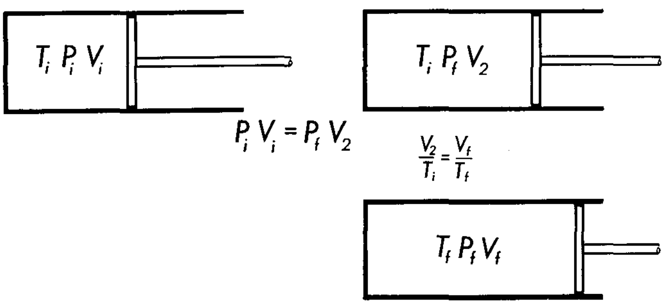


Fig. 16-1 Two steps in the derivation of the general gas law; an isothermal process followed by a constant-pressure process.

absolute temperature T of a constant mass of gas. We may derive the general form of the gas law from the above equations.

Let us consider a gas contained in a cylinder with a closely fitting piston, as shown in Figure 16-1. The initial condition of the gas may be

described in terms of its initial pressure P_i, its initial volume V_i, and its initial temperature T_i. The gas is allowed to expand at constant temperature, say by keeping the cylinder immersed in a bath of melting ice, until its new pressure is P_f and its new volume is V_2. Since the expansion was at constant temperature, we find from Equation (16-1) that

$$P_iV_i = P_fV_2.$$

Now suppose that the gas is heated to a higher temperature T_f, the volume being allowed to expand to a new value V_f, but the pressure on the piston being maintained at the same value P_f throughout this process. Then, from Equation (16-3) we may write

$$\frac{V_2}{T_i} = \frac{V_f}{T_f} = K',$$

or

$$V_2 = V_f\frac{T_i}{T_f}.$$

Substituting for V_2 into the first of the above equations, we find

$$\frac{P_iV_i}{T_i} = \frac{P_fV_f}{T_f}. \tag{16-4}$$

Equation (16-4) is one form of the general gas law. Since the initial state, described by the subscript i, and the final state, described by the subscript f, are entirely arbitrary, the only way in which the quantities on the right- and left-hand sides of the equation can be equal is for each quantity to be separately equal to the same constant. Thus we may rewrite the gas law as

$$\boxed{\frac{PV}{T} = c,} \tag{16-5}$$

where c is a constant whose value depends upon the mass of the enclosed gas. Any convenient units may be used for the pressure and volume, but the temperature T must always be the absolute temperature.

Illustrative Example. A given mass of air occupies a volume of 2,000 cm^3 at 27°C when its pressure corresponds to the pressure at the base of a column of mercury 75 cm high. The air is compressed until its volume is 1,200 cm^3, and its pressure corresponds to 225 cm of mercury. Determine the temperature of the gas after it has been compressed.

From Equation (16-4) we have

$$\frac{P_iV_i}{T_i} = \frac{P_fV_f}{T_f}.$$

The pressure at the base of a column of mercury h cm high is given by $P = h\rho g$, where ρ is the density of the mercury. Substituting numerical values, we have

$$\frac{75 \text{ cm} \times \rho g \times 2{,}000 \text{ cm}^3}{300.2^\circ \text{ abs}} = \frac{225 \text{ cm} \times \rho g \times 1{,}200 \text{ cm}^3}{T_f},$$

from which

$$T_f = 540.4^\circ \text{ abs.}$$

16-2 The Universal Gas Constant *R*

The constant c appearing in Equation (16-5) can be evaluated for any given mass of a gas. Let us designate the value of this constant for a *gram molecular weight*, or *mole*, of a gas by the symbol R. *A gram molecular weight of any substance is an amount of that substance whose mass, expressed in grams, is numerically equal to the molecular weight of the substance.* In the limit of low pressures, the value of R is independent of the chemical nature of the gas, so that R is known as the *universal* gas constant. In the event that n moles of gas are present in a container, Equation (16-5) may be rewritten as

$$PV = nRT. \tag{16-6}$$

The numerical value of the gas constant R can be determined by noting that 1 mole of any gas occupies a volume of 22.4 liters at a pressure of 76 cm of mercury at 0°C; putting these values into Equation (16-6), we get

$$R = 8.31 \times 10^7 \frac{\text{erg}}{\text{mole }^\circ\text{K}} = 1.99 \frac{\text{cal}}{\text{mole }^\circ\text{K}} = 8.21 \times 10^{-2} \frac{\text{liter atm}}{\text{mole }^\circ\text{K}}.$$

16-3 Kinetic Theory of Gases

From the preceding discussion we have seen that all gases exhibit similar thermal and mechanical properties, regardless of their chemical composition, as long as their pressure is sufficiently small. This behavior is quite unlike that of the same substances in liquid or solid form, where these substances exhibit widely different thermal and elastic properties. We are led to infer that the molecules of a gas are sufficiently far apart so that they rarely interact with each other. The pressure of a gas then results from the collisions of the molecules of the gas with the walls of the container. The moving molecules of the gas completely fill every container in which the gas is placed.

We may construct a theory of an ideal gas which is in good agreement with the experimental results described in the preceding sections on the basis of a few simple assumptions. We shall assume that *a gas is composed of molecules that are so small that, to a first approximation, they may be considered as point masses.* We assume further that the molecules *do not*

exert forces on each other except during collisions. We shall further assume that *the molecules of the gas are perfectly elastic,* and that the container is made of perfectly elastic, rigid walls. This implies that mechanical energy is conserved in collisions between molecules. If this were not the case, we would expect to observe that the pressure of a tank of gas would diminish with time, as the molecules lost mechanical energy in inelastic collisions. For the sake of simplicity we shall assume that the gas is in a cubical container of edge d and of volume $V = d^3$.

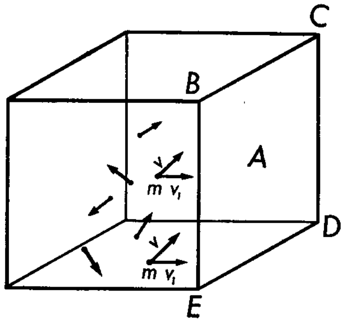


Fig. 16-2 Molecules with equal velocity components v_1 near face *BCDE* of the cube.

The pressure exerted by the gas on the walls of the container is due to the impact of the molecules on the walls, and, when in equilibrium, is equal to the pressure throughout the gas. To calculate this pressure let us assume that the impact of a molecule with a wall is an elastic impact; that is, if a molecule is approaching the wall with a velocity v and momentum mv, then it will leave the wall with a velocity $-v$ and a momentum $-mv$. The change in momentum of the molecule produced by this impact will thus be $-2mv$. To determine the pressure on the walls of the container, let us first calculate the force exerted by the molecules on one of the six faces of the cube, say the face *BCDE* of Figure 16-2, and then divide by its area.

Let us consider those molecules which at some instant are very close to this face. Only those molecules whose velocities have components perpendicular to this face, and directed toward it, will strike it and rebound. Suppose we consider a small number of molecules which have the same value v_1 for this velocity component. The number of these molecules which will strike this face during a small time interval Δt will be one half of the number contained in a small volume $A\ \Delta l$, where A is equal to the area of the face of the cube and $\Delta l = v_1\ \Delta t$; the other half having a velocity component of magnitude v_1 are moving away from the wall. If n_1 represents the number of molecules per unit volume which have a velocity component of magnitude v_1, then the number striking this face of the cube in time Δt will be

$$\frac{n_1}{2} A v_1\ \Delta t.$$

Since each such molecule will have its momentum changed by $-2mv_1$ as a result of this impact, the impulse imparted to the wall will be equal and opposite to it, or $+2mv_1$. The impulse $F_1\,\Delta t$ on the wall produced by these collisions in time Δt will then be

$$F_1\,\Delta t = \frac{n_1}{2} A v_1\,\Delta t \times 2mv_1,$$

from which

$$F_1 = An_1mv_1^2.$$

The pressure on the wall produced by the impact of these molecules is

$$p_1 = \frac{F_1}{A} = n_1mv_1^2.$$

We can now consider another group of molecules, n_2 per unit volume, which have a slightly different velocity component v_2 in this direction; they will produce an additional pressure p_2 given by

$$p_2 = n_2mv_2^2.$$

In this way, we can break up the gas into different groups of molecules, each group contributing a similar term to the pressure on this face of the cube. The total pressure P due to all the different groups of molecules will therefore be of the form

$$P = n_1mv_1^2 + n_2mv_2^2 + n_3mv_3^2 + \cdots.$$

This equation can be simplified by introducing a new term called *the average of the squares of the components of the velocities of all the molecules moving perpendicular to face A* and defined by the equation

$$\overline{v_A^2} = \frac{n_1v_1^2 + n_2v_2^2 + n_3v_3^2 + \cdots}{n},$$

in which n represents the total number of molecules per unit volume. Substituting this value of $\overline{v_A^2}$ in the equation for the pressure, we get

$$P = nm\overline{v_A^2}. \tag{16-7}$$

There will be a similar expression for the pressure on each of the six faces of the cube, except that the factor $\overline{v_A^2}$ will be replaced by the appropriate average of the squares of the components of the velocities of the molecules for that particular face.

The velocity v of any one molecule may be in any direction; it can be resolved into three mutually perpendicular components v_x, v_y, v_z. The magnitude of v in terms of the magnitudes of these components is given by

$$v^2 = v_x^2 + v_y^2 + v_z^2.$$

There will be a similar equation for the square of the velocity of each molecule of the gas in terms of the squares of its three mutually perpendicular components. If we add the squares of the component velocities in the x direction and divide this sum by the total number of molecules, we will get the average value of the square of this velocity component; it will be represented by $\overline{v_x^2}$. Similarly, $\overline{v_y^2}$ and $\overline{v_z^2}$ will represent the average squares of the velocities in the y and z directions, respectively. By adding these average squares of the three velocity components, we get

$$\overline{v^2} = \overline{v_x^2} + \overline{v_y^2} + \overline{v_z^2},$$

where $\overline{v^2}$ is the average of the squares of the velocities of all the molecules. Since the velocities of the molecules have all possible directions, the average value of the squares of the velocity in any one direction should be the same as in any other direction, or

$$\overline{v_x^2} = \overline{v_y^2} = \overline{v_z^2},$$

so that

$$\overline{v^2} = 3\overline{v_x^2}.$$

If we take the x direction as perpendicular to the face A, we can write

$$\overline{v^2} = 3\overline{v_A^2},$$

so that Equation (16-7) becomes

$$\boxed{P = \tfrac{1}{3}nm\overline{v^2}.} \tag{16-8}$$

Recalling that the kinetic energy of a moving molecule is equal to $\frac{1}{2}mv^2$, Equation (16-8) may be written as

$$P = \tfrac{2}{3}n(\tfrac{1}{2}m\overline{v^2}).$$

Since n is the number of molecules per unit volume, we see that the pressure is numerically equal to two thirds the kinetic energy of the molecules in a unit volume of gas.

Let us suppose that N_0 is the total number of molecules in a mole of gas, called *Avogadro's number*, which is contained in a volume V. Then the number of molecules per unit volume n is given by the expression

$$n = \frac{N_0}{V}.$$

Substituting for n into Equation (16-8), we find

$$PV = \tfrac{2}{3}N_0 \times \tfrac{1}{2}m\overline{v^2}. \tag{16-9}$$

Equation (16-9) is a theoretical result obtained from our hypotheses about an ideal gas, relating the pressure and volume of 1 mole of an ideal gas.

If we compare this result to the experimental equation given in Equation (16-6), which for 1 mole of gas becomes

$$PV = RT,$$

we find the theoretical and experimental results to be in agreement if

$$RT = \tfrac{2}{3}N_0\,\tfrac{1}{2}m\overline{v^2},$$

or if

$$\tfrac{1}{2}m\overline{v^2} = \frac{3}{2}\frac{R}{N_0}\,T.$$

It is customary to define a new constant k, called *Boltzmann's constant*, such that

$$k = \frac{R}{N_0}\cdot \tag{16-10}$$

Since R is the gas constant per mole, and N_0 is the number of molecules in a mole of gas, the constant k may be described as the gas constant per molecule. In terms of k the preceding equation becomes

$$\tfrac{1}{2}m\overline{v^2} = \tfrac{3}{2}kT; \tag{16-11}$$

that is, the mean kinetic energy of translation of a molecule of gas is given by $\frac{3}{2}$ the product of Boltzmann's constant by the absolute temperature. This equation gives us some physical meaning of temperature for an ideal gas. For such a gas the temperature is associated with the kinetic energy of the random translational motion of the molecules of the gas. According to Equation (16-11) the average energy of each molecule, and therefore the total internal energy of an ideal gas, is associated with its temperature. Thus the *internal energy of an ideal gas is a function of its temperature only*, and not of its pressure or volume.

In our derivation of the gas law in the form of Equation (16-9), we used the word "molecules" to describe the particles with which we were dealing. These molecules were described by the condition that they were small, relatively far apart, and perfectly elastic. Thus this equation might be used to describe the behavior of any aggregate of particles whose physical dimensions were small compared to their average separation, provided that these particles were elastic and rarely interacted with each other. The neutrons in a nuclear reactor satisfy these conditions. If the neutrons are in equilibrium with the material of the reactor at a temperature T, we speak of them as *thermal neutrons*. The mean velocity $\bar{v}$ of these thermal neutrons may be obtained from Equation (16-11). The particles of a *colloidal suspension* may also be thought of as though they were molecules of an ideal gas, and it is found that these also obey the gas laws.

Equation (16-9) incorporates another result called Avogadro's hypothesis, first stated by Avogadro in 1811, that *all gases occupying equal*

volumes at the same temperature and pressure contain equal numbers of molecules. The accepted value for the number of molecules in a mole of gas N_0 is

$$N_0 = 6.023 \times 10^{23} \text{ molecules/gm molecular wt.}$$

As we have already seen, 1 mole of gas occupies a volume of 22.4 liters at 0°C and at a pressure of 76 cm of mercury. If we perform the calculation indicated in Equation (16-10) to find the numerical value of Boltzmann's constant, we obtain

$$k = 1.38 \times 10^{-16} \text{ erg/}^\circ\text{K}.$$

16-4 Work Done by a Gas

Whenever a gas expands against some external force, it does work on the external agency; conversely, whenever a gas is compressed by the action of some outside force, work is done on the gas. To calculate the work done

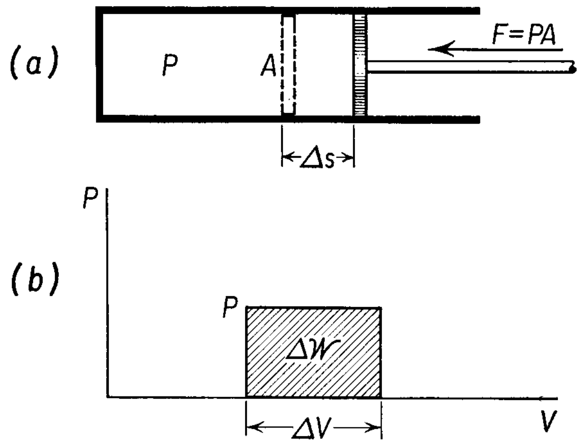


Fig. 16-3 (a) Expansion of a gas at constant pressure. (b) Graphical representation of the work done $\Delta\mathscr{W}$ as an area on the PV diagram.

by a gas, consider a gas enclosed in a cylinder with a tight fitting piston. The piston may be connected to a mechanical device on which it exerts some force. The force F acting on the piston owing to the pressure P of the gas is given by

$$F = PA,$$

in which A is the cross-sectional area of the piston, as shown in Figure 16-3. Suppose that the piston is pushed out a small distance Δs, while the pressure of the gas remains essentially constant. The work $\Delta\mathscr{W}$ done by the gas in

moving the piston is given by

$$\Delta\mathcal{W} = F\,\Delta s = PA\,\Delta s,$$

or

$$\Delta\mathcal{W} = P\,\Delta V. \tag{16-12}$$

Thus the work done by an expanding gas at constant pressure is equal to the product of the pressure by the change in volume. No mechanical work is done by a gas unless there is a change in the volume of the gas. The first law of thermodynamics,

$$\Delta Q = \Delta U + \Delta\mathcal{W}, \tag{15-6}$$

may be rewritten for processes involving gases as

$$\Delta Q = \Delta U + P\,\Delta V. \tag{16-13}$$

Thus, in any process in which the volume of the gas remains constant, called an *isovolumic process*, any heat delivered to the gas must appear as internal energy and is therefore exhibited as a change in the temperature of the gas.

Let us calculate the work done by an ideal gas which expands *isothermally*, that is, at constant temperature, from an initial volume V_1 to a final volume V_2. From the gas law the relationship between the variables of pressure, volume, and temperature for one mole of gas may be stated as

$$PV = RT.$$

The work done may be represented as an integral, from Equation (16-12), as

$$\mathcal{W} = \int d\mathcal{W} = \int_{V}^{V_2} P\,dV,$$

and, substituting for P its value from the gas law,

$$P = \frac{RT}{V},$$

we find

$$\mathcal{W} = RT\int_{V_1}^{V_2}\frac{dV}{V}$$

at constant temperature. Recalling that

$$\int\frac{dx}{x} = \ln x + C,$$

where C is a constant of integration, we find that

$$\mathcal{W} = RT\ln\frac{V_2}{V_1} \tag{16-14}$$

for the work done by one mole of an ideal gas in an isothermal expansion

at temperature T from an initial volume V_1 to a final volume V_2. From Equation (16-14) we see that when the gas expands, that is, when V_2 is greater than V_1, the work done is positive, as is consistent with the sign convention developed in Section 15-6. The work done is shown as the area under the isotherm in Figure 16-4.

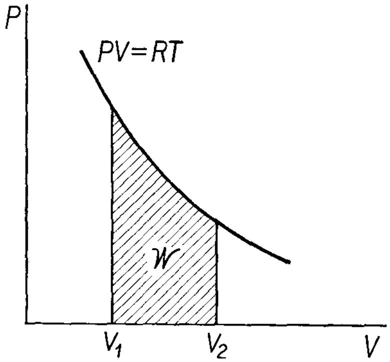


Fig. 16-4 Work done by a mole of a gas in an isothermal expansion at temperature T from volume V_1 to volume V_2.

16-5 Molar Heat Capacity of a Gas

When a quantity of heat ΔQ is delivered to a gas, it may change the internal energy of the gas by an amount ΔU and may also result in the performance of an amount of external work $\Delta \mathcal{W}$ by the gas upon the outside world, in accordance with the first law of thermodynamics. If the volume of the gas is kept constant, all the heat is converted into internal energy. Since the internal energy U of a mole of gas is a function of temperature only, we may define the molar heat capacity at constant volume of a mole of gas C_V as

$$C_V = \frac{\Delta Q}{\Delta T} = \frac{\Delta U}{\Delta T} \quad \text{(at constant volume)}, \tag{16-15}$$

so that the change in internal energy ΔU may be expressed as

$$\Delta U = C_V \, \Delta T. \tag{16-16}$$

From Equation (16-13) the first law of thermodynamics as applied to an ideal gas may be rewritten as

$$\Delta Q = C_V \, \Delta T + P \, \Delta V. \tag{16-17}$$

Let us consider the change in temperature of a mole of gas when a quantity of heat ΔQ is delivered to the gas while the pressure is held constant but the volume is permitted to change. The thermal energy delivered to the gas must now be used to do external work as well as to change the internal energy of the gas. The rise in temperature of the gas will therefore be less than in the case where the volume of the gas is kept constant. The molar heat capacity at constant pressure C_P might be expressed in such units as calories per mole per degree, and may be defined through the equation

$$C_P = \frac{\Delta Q}{\Delta T} \quad \text{(at constant pressure)}, \tag{16-18}$$

and from Equation (16-17) we find

$$C_P = \frac{C_V \,\Delta T + P\,\Delta V}{\Delta T}\cdot$$

From the gas law for a mole of gas we have

$$PV = RT,$$

or

$$V = \frac{R}{P}\,T.$$

At constant pressure both R and P are constant, so that a change of volume ΔV is related to a change in temperature ΔT through the equation

$$\Delta V = \frac{R}{P}\,\Delta T.$$

Substituting this result into the preceding equation for C_P, we find that

$$C_P = C_V + R. \tag{16-19}$$

The molar heat capacity of a gas at constant pressure is always greater than the molar heat capacity at constant volume by the gas constant R.

The internal energy of a monatomic gas, such as helium, is entirely in the form of kinetic energy of translation of the random motions of the atoms of the gas. From Section 16-3 this internal energy may be stated as

$$U = N_0 \times \tfrac{1}{2}m\overline{v^2} = \tfrac{3}{2}RT. \tag{16-20}$$

Since R is a constant, we find that the change in internal energy ΔU associated with a change in temperature ΔT is given by

$$\Delta U = \tfrac{3}{2}R\,\Delta T.$$

Thus the molar heat capacity of a monatomic gas at constant volume may be found by substituting the preceding result into Equation (16-15), to find

$$C_V = \tfrac{3}{2}R \quad \text{(monatomic gas)}. \tag{16-21a}$$

Substituting this result into Equation (16-19), we obtain

$$C_P = \tfrac{5}{2}R \quad \text{(monatomic gas)}. \tag{16-21b}$$

It is customary to designate the ratio of the specific heat at constant pressure to the specific heat at constant volume by the letter γ (gamma). Thus

$$\gamma = \frac{C_P}{C_V}\cdot \tag{16-22}$$

Substituting from Equations (16-21) into (16-22), we find the value of γ for a monatomic gas to be

$$\gamma = \tfrac{5}{3} \quad \text{(monatomic gas)}.$$

In our development of the kinetic theory of an ideal gas we assumed that a molecule could be considered as a point mass and showed that the average kinetic energy of translation per molecule is $\frac{3}{2}kT$ (Equation 16-11). There are three independent directions of motion of translation, say the x, y, and z directions. We say that the molecule has three *degrees of freedom*; that is, three coordinates are necessary to specify the position of the molecule at any instant, one for each degree of freedom. Since there is no reason for preferring one direction rather than another, we postulate the principle of *equipartition of energy*, that each degree of freedom should have the same amount of energy. Referring to Equation 16-11, the amount of energy to be associated with each degree of freedom per molecule is $\frac{1}{2}kT$. The internal energy of a mole of a monatomic gas will then be $\frac{3}{2}RT$, as given by Equation 16-20.

The idea of degrees of freedom can be extended to diatomic and polytomic gases. It can be shown that the value of γ can be expressed as

$$\gamma = \frac{f+2}{f},$$

where f is the number of degrees of freedom per molecule. For a monatomic gas we find that $\gamma = \dfrac{3+2}{3} = \dfrac{5}{3} = 1.67$, in agreement with measured values as shown in Table 16-1.

We may extend these ideas to a diatomic molecule which we may imagine to be two point masses a fixed distance apart; the line joining the two atoms is the axis of the molecule. If the diatomic molecule is considered as a rigid body, then it will have three degrees of freedom owing to the translational motion of the entire molecule, plus a certain number of degrees of freedom owing to the rotational motion of the molecule. A glance at Table 16-1 shows that $\gamma = 1.4$ for diatomic gas, indicating that $f = 5$; thus there must be two additional degrees of freedom of rotation. These would correspond to rotations about two mutually perpendicular axes in a plane at right angles to the line joining the two atoms.

Thus if at ordinary temperatures the molecules of a diatomic gas may be thought of as rigid, and possessing no vibrational energy, the mean energy of each molecule must be $\frac{5}{2}kT$. The total internal energy of a mole of such a gas is given by

$$U = N_0 \times \tfrac{5}{2}kT = \tfrac{5}{2}RT \quad \text{(diatomic gas)}. \tag{16-23}$$

From this expression we find the molar heat capacities of a diatomic gas to be

$$C_V = \tfrac{5}{2}R, \qquad C_P = \tfrac{7}{2}R \quad \text{(diatomic gas)}, \tag{16-24}$$

and the ratio of the specific heats γ is

$$\gamma = \tfrac{7}{5} \quad \text{(diatomic gas).} \tag{16-25}$$

These results are in rather remarkable agreement with experiment, as shown in Table 16-1.

TABLE 16-1 THE MOLAR HEAT CAPACITY AT CONSTANT VOLUME, AND THE RATIO OF SPECIFIC HEATS FOR SEVERAL GASES

Gas	Atoms per Molecule	C_V*		γ	
		Theory	Experiment	Theory	Experiment
Argon	1	2.98	2.98	1.67	1.67
Helium	1	2.98	2.98	1.67	1.66
Oxygen	2	4.97	5.04	1.40	1.40
Nitrogen	2	4.97	4.93	1.40	1.40
Carbon monoxide	2	4.97	4.94	1.40	1.40

* The units of C_V are given as calories per mole per degree centigrade.

At high temperatures the classical theory of the specific heats of gases is no longer adequate to describe the experimental determinations. Additional rotational and vibrational modes are excited at high temperatures in a manner which is best described by the quantum theory of specific heats; this is beyond the scope of this book.

16-6 Adiabatic Processes

An adiabatic process is one in which no heat enters or leaves the system. In the form of an equation

$$\Delta Q = 0.$$

If we apply the first law of thermodynamics, in the form of Equation (16-17), to adiabatic processes, we find

$$\Delta Q = 0 = C_V\,\Delta T + P\,\Delta V.$$

Thus

$$\Delta T = -P\frac{\Delta V}{C_V}\cdot$$

The general gas law describes the behavior of a gas under all circumstances. Thus for a mole of gas we have

$$PV = RT.$$

If P, V, and T are permitted small variations, we find, on taking differentials, that

$$P\,\Delta V + V\,\Delta P = R\,\Delta T.$$

If we replace the gas constant R by

$$R = C_P - C_V$$

from Equation (16-19), and substitute the value of ΔT into the above equation, we find

$$P\,\Delta V + V\,\Delta P = -(C_P - C_V)\,\frac{P\,\Delta V}{C_V} = -(\gamma - 1)\,P\,\Delta V,$$

for, from Equation (16-22), $\gamma = C_P/C_V$. On transposing, and dividing the above equation by the product PV, we find

$$\frac{\Delta P}{P} + \gamma\frac{\Delta V}{V} = 0.$$

In the limit of small increments we may replace Δ by d, and integrate to find

$$\ln P + \gamma \ln V = \text{constant},$$

or

$$PV^{\gamma} = \text{constant}. \tag{16-26}$$

The value of the constant is determined by the quantity of gas present, so that Equation (16-26) may be used to describe the relationship between the pressure and the volume of any quantity of gas undergoing an adiabatic change. If a gas originally at pressure P_1, volume V_1, and temperature T_1 is compressed adiabatically, as in an insulated cylinder, to a new pressure P_2, the new volume of the gas V_2 may be found from Equation (16-26) by writing

$$P_1V_1^{\gamma} = P_2V_2^{\gamma}. \tag{16-27}$$

The final temperature of the gas T_2 may then be obtained from the gas law

$$\frac{P_1V_1}{T_1} = \frac{P_2V_2}{T_2},$$

in which all quantities except T_2 are now known.

Illustrative Example. A mass of gas occupies a volume of 8 liters at a pressure of 1 atm and a temperature of 300° abs. It is compressed adiabatically to a volume of 2 liters. Determine (a) the final pressure and (b) the final temperature, assuming it to be an ideal gas whose value of $\gamma = 1.5$.

(a) The final pressure of the gas can be determined with the aid of Equation (16-27) thus,

$$P_2 = P_1\left(\frac{V_1}{V_2}\right)^{\gamma},$$

so that

$$P_2 = 1\text{ atm}\left(\frac{8}{2}\right)^{1.5},$$

from which

$$P_2 = 8\text{ atm}.$$

(b) The final temperature can be found with the aid of the general gas law thus,

$$T_2 = T_1 \frac{P_2 V_2}{P_1 V_1},$$

so that

$$T_2 = 300° \text{ abs} \frac{8 \text{ atm} \times 2\text{l}}{1 \text{ atm} \times 8\text{l}},$$

or

$$T_2 = 600° \text{ abs}.$$

16-7 The Maxwell Distribution Function

In Section 16-3 we showed how the properties of a gas could be accounted for on the basis of a very simple set of hypotheses about the nature of a gas. We assumed that the gas was made up of many molecules in rapid motion, and that the molecules were sufficiently far apart so that the forces

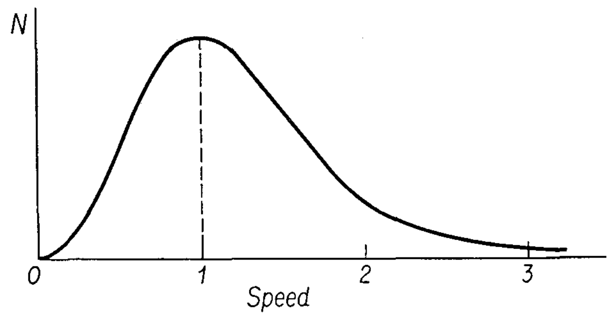


Fig. 16-5 The Maxwellian distribution of molecular speeds. Relative numbers of molecules having speeds in a unit speed interval at various speeds are shown as ordinate, while speeds in units of the most probable speed are shown as abscissa.

that one molecule exerted on another were of minor importance and could be neglected. We assumed that the molecules were perfectly elastic so that there was no loss in mechanical energy in collisions between molecules of the gas and the walls of the container. On the basis of such arguments, we could account for the gas law, and we were able to show that the temperature of a gas was directly related to the average value of the kinetic energy of its molecules.

When a gas is in equilibrium at an absolute temperature T, the distribution of velocities of the molecules of the gas is given by Figure 16-5, called the Maxwellian distribution, according to the theory first developed by Maxwell (1831–1879). This theory has now been well verified by experiment as actually describing the behavior of gas molecules. In fact, the Maxwellian distribution may be taken as the meaning of the temperature

of a gas, for, if a collection of gas molecules has a velocity distribution which differs from Figure 16-5, then we may say that the gas has not yet reached thermal equilibrium and therefore does not have a well-defined temperature.

The average of the velocity components of the molecules of a gas in a particular direction must be zero. If this were not so, the gas and its container would be in translational motion. However, the average value of the squares of the molecular velocities is not zero, and is given by Equation (16-11). From Figure 16-5 we see that molecules whose speeds are more than three times the most probable speed are extremely rare. Nevertheless, there are some molecules in the gas which have very large speeds, for the distribution curve approaches the horizontal axis asymptotically. We must also note that, at a given temperature, the molecules of a gas of low molecular weight are in more rapid motion than the molecules of a gas of high molecular weight. This has the interesting consequence that hydrogen and helium are steadily diffusing out of the earth's atmosphere, for, at the temperature of the outer air, some of these lighter molecules are moving sufficiently rapidly to attain the escape velocity of 11 km/sec necessary for a projectile to escape the gravitational pull of the earth.

If we call the energy required to disrupt a chemical molecule its *binding energy*, we see that, as the temperature of a gas is raised, a greater proportion of gas molecules may have kinetic energies greater than the binding energy, so that a molecule may be decomposed as a result of energy transfer during a collision. Thus molecules which are stable at ordinary temperatures must have binding energies which are large compared to the mean kinetic energy of a molecule at room temperature, as given by Equation (16-11).

It is interesting that modern theories of the structure of atoms and molecules have provided a justification of a basic assumption of the kinetic theory of gases. According to the *quantum theory*, molecules exist only in certain *quantum states*, each having a fixed amount of energy. These are sometimes called *energy levels*. The molecule normally exists in its state of lowest energy, called its *ground state*, and can only absorb energy in a collision with another molecule in exactly the right amount to raise it to a state of higher energy, called an *excited state*, or to disrupt it completely. In general, it is very unlikely that the colliding molecules will have just the right amount of energy for excitation, so that the collisions between the molecules of a gas result in no absorption of energy by the molecules. The kinetic energy is conserved in the collision rather than being transferred to internal excitation energy of one molecule. The collisions are therefore perfectly elastic.

Problems

16-1. A closed vessel contains dry air at 25°C and 76 cm of mercury pressure. Its temperature is raised to 100°C. Determine the pressure of the air, neglecting the change in volume of the container.

16-2. A mass of oxygen occupies a volume of 1 liter at a pressure of 76 cm of mercury when its temperature is 40°C. The gas is allowed to expand until its volume is 1.5 liters and its pressure is 80 cm of mercury. (a) Determine its final temperature. (b) Determine the number of moles of oxygen in the system.

16-3. Derive the general gas law from Equations (16-1) to (16-3) by considering that the gas is taken from $T_iP_iV_i$ by a constant-volume process to $T_2P_fV_i$, and thence by a constant-pressure process to $T_fP_fV_f$.

16-4. A certain gas has a density of 0.001 gm/cm^3 when its temperature is 50°C and its pressure is 4 atm. What pressure will be needed to change the density of the gas to 0.002 gm/cm^3 when its temperature is 100°C?

16-5. An automobile tire has a volume of 1,000 in.^{3} and contains air at a gauge pressure of 24 lb/in.^{2} when the temperature is 0°C. What will be the gauge pressure of the air in the tires when its temperature rises to 27°C and its volume increases to 1,020 in.^{3}?

16-6. Determine the pressure of 4.032 gm of hydrogen which occupies a volume of 16.8 liters at a temperature of 0°C. The molecular weight of hydrogen is 2.016.

16-7. Determine the average value of the kinetic energy of the molecules of a gas (a) at 0°C and (b) at 100°C.

16-8. (a) What is the mass of a hydrogen molecule? (b) Determine the average velocity of a molecule of hydrogen at 27°C.

16-9. The molecules of a certain gas have a mass of 5×10^{-24} gm. What is the number of molecules per cubic centimeter of this gas when its pressure is 10^6 dynes/cm^2 and its temperature is 27°C?

16-10. Calculate the work done in compressing one mole of oxygen from a volume of 22.4 liters at 0°C and 1 atm pressure to 16.8 liters at the same temperature.

16-11. A cylinder contains a mole of hydrogen at 0°C and 76 cm of mercury pressure. Calculate the amount of heat required to raise the temperature of this hydrogen to 50°C (a) keeping the pressure constant, and (b) keeping the volume constant. (c) What is the volume of the hydrogen when at 0°C?

16-12. A cylinder contains 32 gm of oxygen at 0°C and 76 cm of mercury pressure. Calculate the amount of heat required to raise the temperature of this mass of oxygen to 80°C (a) keeping the pressure constant and (b) keeping the volume constant. (c) How much mechanical work is done by the oxygen in each case?

16-13. A mass of a monatomic gas occupies a volume of 400 cm^3 at a temperature of 17°C and a pressure of 76 cm of mercury. The gas is compressed adiabatically until its pressure is 90 cm of mercury. Determine (a) the final volume of the gas and (b) the final temperature of the gas.

16-14. A mass of a diatomic gas occupies a volume of 6 liters at a temperature

of 27°C and 75 cm of mercury pressure. The gas expands adiabatically until its volume is 8 liters. What is the final temperature of the gas?

16-15. A mole of gas at atmospheric pressure and 0°C is compressed isothermally until its pressure is 2 atm. How much mechanical work is done on the gas during this operation?

16-16. Ten grams of oxygen are heated at constant atmospheric pressure from 27°C to 127°C. (a) How much heat is delivered to the oxygen? (b) What fraction of the heat is used to raise the internal energy of the oxygen?

16-17. An air bubble of volume 20 cm^3 is at the bottom of a lake 40 m deep where the temperature is 4°C. The bubble rises to the surface where the temperature is 20°C. Assuming that the temperature of the bubble is the same as that of the surrounding water, what is its volume just as it reaches the surface?

16-18. An ideal gas for which $\gamma = 1.5$ is enclosed in a cylinder of volume 1 m^3 under a pressure of 3 atm. The gas is expanded adiabatically to a pressure of 1 atm. Find (a) the final volume and (b) the final temperature of the gas if its initial temperature was 20°C.

16-19. A mass of 1.3 kg of oxygen of molecular weight 32 is enclosed in a cylinder of volume 1 m^3 at a pressure of 10^5 nt/m^2 and a temperature of 20°C. From these data find the universal gas constant R assuming oxygen to be an ideal gas.

16-20. A gas of mass m and molecular weight M undergoes an isothermal expansion from an initial pressure P_1 and volume V_1 to a final pressure P_2 and volume V_2 while at temperature T. Find (a) the work done by the gas in this expansion, (b) the heat flow to the gas, and (c) the change in internal energy of the gas in terms of these symbols.

16-21. A piece of putty is placed in a vise with insulating jaws. A constant force of 100 nt is applied through a distance of 2 cm. The putty is found not to have its volume changed in this process. What is the change in the internal energy of the putty?

16-22. Prove that $TV^{\gamma-1}$ = constant for an adiabatic process.

16-23. Show that the work done by a gas in an adiabatic expansion from initial conditions P_i, V_i to final conditions P_f, V_f is given by

$$\mathscr{W} = \frac{P_i V_i^\gamma}{1-\gamma} (V_f^{1-\gamma} - V_i^{1-\gamma}).$$

17

The Phases of Matter

17-1 Phases of a Substance

A substance which has a definite chemical composition can exist in one or more *phases*, such as the vapor phase, the liquid phase, or the solid phase. When two or more such phases are in equilibrium at any given temperature and pressure, there are always surfaces of separation between the two phases.

In the solid phase a pure substance generally exhibits a well-defined crystal structure in which the atoms or molecules of the substance are arranged in a repetitive lattice. Many substances are known to exist in several different solid phases at different conditions of temperature and pressure. These solid phases differ in their crystal structure. Thus ice is known to have six different solid phases, while sulphur has four different solid phases.

A substance in the liquid phase is fluid; that is, it flows under the application of a shearing stress. Many substances normally classified as solids are really highly viscous fluids. Such substances (glass or pitch are examples), exhibit no well-defined melting point but soften and flow more readily as the viscosity decreases with increasing temperature. The common observation that fluids cannot withstand a shearing stress depends upon the fact that the viscosity of most fluids is small. A highly viscous fluid such as glass is capable of withstanding considerable shearing stress without apparent deformation. When the applied shearing stress is made to oscillate rapidly, even normal liquids display a shear modulus.

Based on information largely derived from x-ray scattering experiments similar to the experiments used to determine the crystal structure of solids (see Chapter 44), we know that the molecules of a liquid are not entirely independent of each other but are coordinated. The central molecule of a group of coordinated molecules is surrounded by a group of from about 3 to 11 other molecules in the first coordination shell, at a distance of about 1 molecular diameter from the central molecule. A smaller number of

molecules forms the second coordination shell, at a distance of about 2 molecular diameters from the central molecule. Thus there is considerable organization of the molecules of a liquid. The degree of organization of the liquid tends to increase as the temperature of the liquid approaches the freezing point, and, conversely, the degree of organization tends to decrease as the liquid approaches the boiling point.

Most liquids exhibit a single liquid phase. The most abundant isotope of helium, of atomic weight 4, is a notable exception. At atmospheric pressure helium liquefies at a temperature of 4.2° abs. Upon further cooling to 2.2° abs, the original liquid, known as helium I, changes to helium II which displays an abnormally low viscosity and a high thermal conductivity. Helium II flows through capillary tubes with great rapidity, and passes through tightly packed plugs of porous material as though they were sieves. The superfluidity of helium II is thought to be associated with the distribution of the energy levels of liquid helium near the absolute zero. It is interesting in this connection that the isotope of helium of atomic weight 3 displays no superfluidity.

All substances exhibit only one vapor phase. A vapor at low pressure is often termed a gas. As we have seen in the preceding chapter, the molecules of a gas are isolated from each other and interact with each other only rarely. As the pressure is increased, each gas molecule collides with other molecules more frequently, with a tendency for the molecules to form in clusters. The ideal gas law is correct only as long as no clusters are formed, for each cluster behaves like a complex molecule and tends to reduce the effective number of molecules per unit volume. The condensation of droplets from a vapor may be interpreted as the formation of giant clusters.

We have already considered two types of phase changes, solid to liquid and liquid to vapor (see Section 15-3). A third type of phase change, that from solid to vapor, is called *sublimation.* Heat must be added to produce this change of phase; the quantity of heat required to change a unit mass of a substance from solid to vapor at constant temperature is called the *heat of sublimation.* Conversely, when a substance is condensed directly from the vapor to the solid phase, heat must be removed from the substance.

In this chapter we shall consider some important phenomena associated with changes of phase.

17-2 Vaporization. Vapor Pressure

One method for studying the process of vaporization and the properties of a vapor is illustrated in Figure 17-1. A tube A about 1 m long is first filled with mercury and then inverted and put into a long reservoir R containing mercury. The level of the mercury in A will be at the barometric height

above the level in R. The space above the mercury in A contains mercury vapor at a very low pressure, which we shall neglect for the purpose of the present discussion. If a small quantity of liquid ether is put into the open end of tube A, the ether will rise to the top of the mercury column and vaporize into the space above it. Let us suppose that there is no trace of liquid ether above the mercury column; that is, the ether has completely evaporated. We shall observe that the level of the mercury in column A is much lower than before. If we push the tube A slowly into the reservoir R, we shall find that at some stage in this process a small layer of liquid ether will appear on top of the mercury column in A. As the tube is pushed down still farther, the thickness of the ether layer increases.

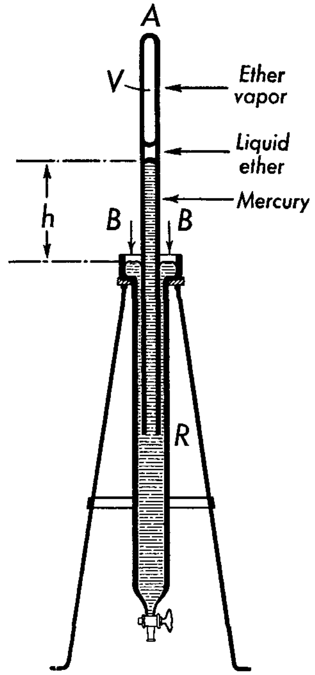


Fig. 17-1 Method of measuring the saturated vapor pressure of ether.

If we determine the pressure due to the ether vapor, that is, the difference between the barometric height and the height h of the column of mercury in tube A, we note that, as long as there is a film of liquid ether at the top of the mercury column, the pressure of ether vapor remains constant. A vapor which is in contact with its parent liquid is said to be a *saturated vapor*. The *pressure of a saturated vapor depends only upon its temperature*. If the apparatus is modified so that we can vary the temperature of the liquid ether and of its saturated vapor, it will be found that the pressure of the saturated vapor increases rapidly with increasing temperature. Under the conditions of the experiment, the liquid phase is in equilibrium with its saturated vapor. There is a continuous interchange of molecules between the liquid and the vapor phases such that the mass of the liquid phase and the mass of the vapor phase remain constant at a particular pressure, volume, and temperature.

When water, placed in an open dish, is heated by a flame, the water will show an increase in temperature until the boiling point is reached. While the water is boiling, the temperature remains constant. The heat supplied to the dish is used to convert the liquid water to the vapor phase. Since the dish is open to the atmosphere, the pressure of the water vapor just above the liquid surface must be equal to the pressure of the atmosphere. The *boiling point* is that temperature at which the pressure of the

saturated vapor is equal to the atmospheric pressure. When the pressure on the liquid is 76 cm of mercury, the boiling point of water is, by definition, 100°C or 212°F. The boiling point of a liquid at this pressure is called its *normal boiling point*. The normal boiling points of several substances are shown in Table 17-1.

TABLE 17-1 NORMAL BOILING POINTS

Substance	Temperature in °C
Sulphur*	444.6
Mercury	356.7
Water*	100.0
Alcohol (ethyl)	78.3
Ether	34.6
Oxygen*	−182.97
Nitrogen	−195.8
Hydrogen	−252.8
Helium	−269

* Fixed points of the International Temperature Scale of 1948.

The relationship between the temperature of water and the pressure of its saturated vapor can be investigated by placing the water in a boiler which has a thermometer and a pressure gauge fitted into it. The water is first boiled vigorously to drive off the air above it, and the boiler valve is then closed. In such a boiler the water is in equilibrium with its saturated vapor, and the relationship between vapor pressure and temperature may be plotted, as in Figure 17-2. This curve is called the *vaporization curve of water*. From the preceding discussion we see that a point P along the vaporization curve OA of Figure 17-2 represents the temperature at which water boils at a particular pressure. If the temperature is greater than the boiling point at a particular pressure, the water must all be in the vapor phase. If the temperature is less than the boiling point at that pressure, the water must all be in the liquid phase.

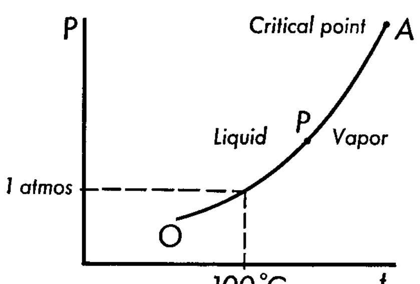


Fig. 17-2 Vaporization curve of water.

When a mass m of liquid is transformed from the liquid to the vapor phase at a temperature t, a quantity of heat mL must be supplied to the liquid, where L is the latent heat of vaporization at this temperature. This heat must supply the energy required to change the internal energy

of the substance and to perform the mechanical work associated with the change in volume of the substance. If U_v is the internal energy of a mass m of vapor at the temperature t, and U_l is the corresponding internal energy of the liquid, we have, from the first law of thermodynamics,

$$mL = U_v - U_l + P(V_v - V_l), \tag{17-1}$$

where P is the pressure at which the phase change takes place and V_v and V_l are the volumes of the vapor and liquid, respectively. Although Equation (17-1) has been written to describe the phase change from liquid to vapor, an analogous equation can obviously be written to describe any other phase change. In transformations involving the vapor phase, the $P\,\Delta V$ term may represent a significant part of the heat required to vaporize a substance, because of the large changes in volume associated with such transformations. It must be observed that the measurement of the latent heat of vaporization at a particular temperature is made under conditions of constant pressure, unless specifically noted to the contrary.

17-3 The Critical Point

The vaporization curve is not indefinite in extent; it has both an upper and a lower limit. The upper limit is known as the *critical point,* shown as point A in Figure 17-2. The temperature and pressure at the critical point are known as the *critical temperature* and *critical pressure.* If a glass vial containing a liquid and its saturated vapor is heated to the critical temperature, the meniscus dividing the liquid from the vapor phase disappears. At temperatures above the critical temperature the substance cannot exist as a liquid; that is, no matter how great the pressure, it cannot be put into the liquid phase. At the critical temperature the densities of the liquid and the vapor are equal, and the heat of vaporization is zero. This suggests that the molecules of a substance have sufficient kinetic energy to overcome the bonds of coordination through which molecules become coordinated in the liquid phase when the temperature is greater than the critical temperature. A distinction is sometimes made between the vapor states above and below the critical temperature: above the critical temperature a substance is usually called a gas; below the critical temperature a substance is usually called a vapor. The critical temperature of carbon dioxide is 31.1°C or 88°F, and its critical pressure is 73 atm. Tanks of carbon dioxide, used for carbonating soft drinks, are usually labeled "liquid carbonic gas," which is clearly a misnomer on a hot summer day.

The substances which are known as gases at ordinary temperatures have very low critical temperatures. They must first be cooled to low temperatures before they can be liquefied. In terms of our conception of a vapor as clusters of very few molecules, and a liquid droplet as a giant

cluster, we may infer that a substance which is chemically active, and therefore must be surrounded by a relatively intense attractive force field, should be more easily formed into clusters and liquefied than a substance which is chemically inactive. A second factor which must influence the formation of clusters of molecules in a gas is the length of time they spend in each other's neighborhood. According to the kinetic theory of gases, the average speed of the molecules of a gas at a given temperature depends only upon the molecular mass, with the more massive molecules having the lower speeds. The substance which should be hardest to liquefy is the lightest of the noble gases, that is, helium. This is indeed the case, for the critical temperature of helium is −268°C, or 5° abs, the lowest of all known substances. The critical temperature and pressure of some common substances are shown in Table 17-2.

TABLE 17-2 CRITICAL CONSTANTS

Substance	Critical Temperature in °C	Critical Pressure in Atmospheres
Ammonia	132	112
Carbon dioxide	31.1	73
Ether	194	35.5
Helium	−268	2.3
Hydrogen	−240	13
Nitrogen	−147	34
Oxygen	−119	50
Water	374	218

17-4 Evaporation

A liquid such as water or alcohol, when left open to the atmosphere, will evaporate; that is, the liquid will change to a vapor and go into the atmosphere. If the atmosphere above the liquid surface is set into motion, the rate of evaporation will be increased. The heat that is required to vaporize a liquid must come either from the external surroundings or from the remaining liquid. If the liquid is placed in a fairly well insulated container, or if the process of evaporization is so rapid that the liquid cannot get sufficient heat from the surrounding bodies, the temperature of the liquid will be lowered.

If a large jar is placed over the liquid container, some of the liquid will evaporate until equilibrium is established between the liquid and its vapor, that is, until the partial pressure of the vapor in the air is equal to the saturated vapor pressure at the temperature of the air. The air is then said to be *saturated* with vapor. If the jar over the liquid container is

removed and a fresh supply of unsaturated air is blown over the liquid, evaporation will start again and will continue as long as the vapor pressure in the air is less than the saturated vapor pressure. Substances such as ether, which have a large vapor pressure at atmospheric temperature, evaporate rapidly when exposed to a moving air stream. Substances with very low vapor pressure, such as octoil *S*, a vacuum-pump oil with a vapor pressure of about 10^{-7} mm Hg at room temperature, evaporate at an extremely slow rate.

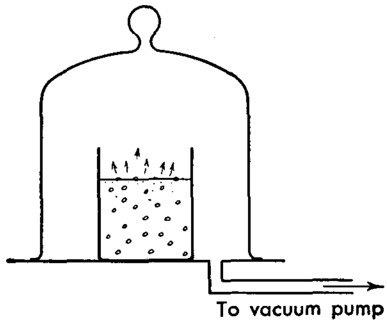


Fig. 17-3 Rapid evaporation of water when the pressure above it is reduced causes it to cool very rapidly. If the evaporation is sufficiently rapid, the remaining water may freeze.

From the viewpoint of the kinetic theory of matter, we may conceive of the process of evaporation as one in which only the more rapidly moving molecules of the liquid have sufficient energy to overcome the forces of attraction at the liquid surface. The escape of the rapidly moving molecules from the liquid decreases the average energy of the molecules of the liquid and therefore lowers the internal energy and the temperature of the remaining liquid. If a small beaker of water is placed under a bell jar which is continuously evacuated by a vacuum pump, as shown in Figure 17-3, the water vapor is rapidly removed from the surface of the liquid. The selective removal of the most rapidly moving molecules from the liquid lowers the temperature of the liquid, and if the pumping action is sufficiently fast, the water may freeze. Thus it is possible to freeze water without first providing a low-temperature reservoir, that is, a body whose temperature is at or below the ice point.

17-5 Throttling Process

When a fluid goes from a region of high pressure to one of lower pressure, it undergoes a *throttling process.* Let us suppose that a gas is confined in a cylinder at a pressure P_1, and that, by motion of a piston, the gas in the cylinder is forced through a narrow aperture in the base of the cylinder into a second cylinder, as shown in Figure 17-4. By means of a second psiton in the second cylinder, the gas in that cylinder is maintained at a

lower pressure P_2. Assume that both cylinders have insulating walls, so that the heat ΔQ passing into or out of the system is zero. We may calculate the work done in the throttling process by imagining that the piston at the left in Figure 17-4 is moved to the right by a force F_1 such that

$$F_1 = P_1 A,$$

where A is the area of the piston. The piston moves a distance l_1 in sweeping the gas out of the left-hand cylinder, so that the work done on the gas is

$$F_1 l_1 = P_1 A l_1 = P_1 V_1,$$

where V_1 is the original volume of the gas in the left-hand cylinder. In accordance with the sign convention adopted for the first law of thermo-

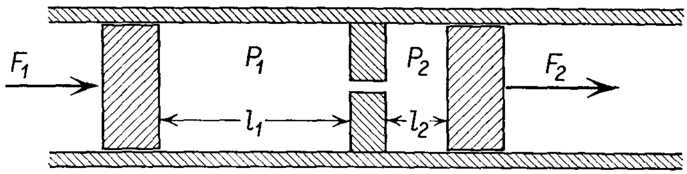


Fig. 17-4 Throttling process.

dynamics, as in Section 15-6, this is work done on the gas and is therefore negative work. Similarly, the work done by the gas in expanding into the second cylinder is $+P_2V_2$. Substituting these results into the first law of thermodynamics, as expressed in Equation (15-6), we find

$$0 = U_2 - U_1 - P_1V_1 + P_2V_2,$$

where U_1 and U_2 represent the internal energy of the gas in cylinder 1 and cylinder 2, respectively. On transposing, we find

$$U_1 + P_1V_1 = U_2 + P_2V_2,$$

or

$$U + PV = \text{constant}. \tag{17-2}$$

The quantity $U + PV$ is called the *enthalpy*. Thus the enthalpy of a fluid remains unchanged in a throttling process. The throttling process is the basis for mechanical refrigeration systems and for the liquefaction of gases.

In a refrigeration system a refrigerant is compressed to a high pressure by a pump, is cooled to room temperature by circulating air or water, and is permitted to expand through an expansion valve to a lower pressure. This process is a throttling process, in which the enthalpy of the refrigerant remains constant, according to Equation (17-2). The enthalpy is tabulated for many refrigerants as a function of the temperature, so that a knowledge of the initial and final pressures and the initial temperature in a throttling process enables one to find the final temperature by consulting an appropri-

ate enthalpy table. Liquids which evaporate as a result of a throttling process always drop in temperature and are therefore always used in refrigerating systems (see Section 19-6).

If a perfect gas should undergo a throttling process, the final temperature should be the same as the initial temperature, while for real gases the temperature change in a throttling process may be positive or negative, depending on the state of the gas. From the microscopic point of view, if the forces that molecules exert on each other are predominantly attractive, work will be done in the expansion, and the temperature will be lowered; if the forces between molecules are repulsive, the molecules will have greater kinetic energy when farther apart, and the temperature will be raised in the expansion. If the molecules exert no force on each other, as in a perfect gas, no work is done in the expansion, and there will be no change in kinetic energy and hence no change in temperature.

In one type of liquefaction process, the gas to be liquefied undergoes a throttling process. The initial state of the gas—that is, its pressure, volume, and temperature—and the pressure of the final state are so chosen, that the temperature of the gas will be lowered during this process. The cooler gas then circulates around the incoming gas so that its temperature is lowered still further. This recirculation process continues until some of the emergent fluid is partially liquefied; the liquid is collected in a thermos flask, and the cold vapor is recirculated.

17-6 Fusion

The temperature at which a solid melts depends upon the pressure. For most substances the temperature of the melting point increases with increasing pressure. In a few exceptional cases, such as ordinary ice, cast iron, and type metal, there is a decrease in volume on melting. This unusual behavior of type metal is of practical value in the casting of type; clear sharp type is produced as the solidifying metal expands against the mold. The expansion which takes place when water freezes has many important consequences. The specific gravity of ice is 0.92, so that the ice formed in a lake floats on top of the water. Since the ice is a poor conductor of heat, the water in the deeper portions of a lake may remain unfrozen throughout the winter so that aquatic life can continue beneath the surface. The curve *OB* in Figure 17-5 shows the variation of the melting point of ice with pressure. Each point on the curve represents a definite temperature and pressure at which ice and water are in equilibrium.

The fact that an increase in pressure results in a lowering of the melting point of ice leads to a series of interesting results. We make snowballs by compressing loose snowflakes. The ice beneath an ice skate is at higher pressure than the ice around it. This increase in pressure causes the ice

or snow to melt, even though the temperature is below 0°C. Upon release of the pressure, the water refreezes. This process of melting at a temperature below 0°C because of the increased pressure, and then refreezing when the pressure is removed, is known as *regelation.* If a wire is looped over a block of ice and a heavy weight is hung from the wire, the wire will pass completely through the block of ice through the process of regelation. After the wire has passed through it, the block is still in one solid piece.

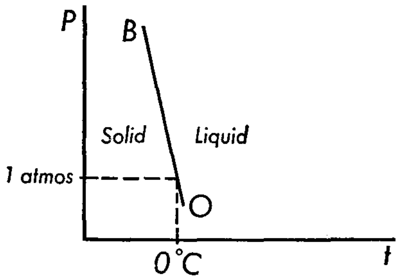


Fig. 17-5 Fusion curve of water.

It can be shown on the basis of thermodynamic principles that the latent heat L, the temperature T, the change in volume of a unit mass of substance ΔV, and the pressure P associated with a phase change are related through the equation

$$\frac{dP}{dT} = \frac{L}{T\,\Delta V}. \tag{17-3}$$

This equation is known as *Clapeyron's equation.* The quantity dP/dT is the slope of curve on the PT diagram which describes the boundary between the two phases, such as the vaporization curve OA of Figure 17-2 or the fusion curve OB of Figure 17-5. In melting, a positive quantity of heat L is always delivered to a unit mass of the substance. The slope of the curve will depend upon the volumes of a unit mass of substance in the two phases. In the case of the transition from ice to water, the quantity $\Delta V = V_w - V_i$; that is, the change in volume is the volume of a unit mass of substance in the final state V_w minus the specific volume in the initial state V_i. Since here ΔV is a negative quantity, the slope of the fusion curve is negative. In most phase changes the final state is less dense than the initial state, and the slope of the curve is positive, as in Figure 17-2.

The melting points of several substances are shown in Table 17-3.

TABLE 17-3 MELTING POINTS AT ATMOSPHERIC PRESSURE

Substance	Temperature in °C	Substance	Temperature in °C
Aluminum	660	Platinum	1773.5
Copper	1083	Silver*	960.8
Gold*	1063.0	Sodium	97.5
Lead	327.4	Tungsten	3410
Mercury	−38.87	Water*	0
Nickel	1455	Zinc	419.5

* Fixed points of the International Temperature Scale of 1948.

17-7 Superheating and Supercooling

If a liquid is heated very slowly, it may remain in the liquid phase at temperatures above the boiling point. If a vapor is cooled very slowly, it may be cooled below the boiling point without the formation of droplets of liquid. Similarly, a liquid may be cooled below its freezing point without freezing setting in. These phenomena are called *superheating* and *supercooling*. When a substance is superheated or supercooled, mechanical agitation, or some other disturbance, may serve to initiate the phase transition. Under certain conditions dry-ice crystals or crystals of sodium iodide may be introduced into a cloud to start the formation of water droplets and cause rain. It is customary to introduce ceramic beads in boiling flasks in chemistry laboratories to prevent superheating and the consequent eruptive boiling, or bumping, which takes place under these circumstances.

Both of these phenomena have been exploited in nuclear physics as a means of visualizing the passage of charged particles through matter in two very important devices called the *cloud chamber* and the *bubble chamber*. When rapidly moving charged particles pass through matter, they leave behind an invisible trail of electrically charged molecules called ions. In a supercooled vapor these ions act as nuclei for droplet formation, and in a superheated liquid they act as nuclei for vapor formation. In a cloud chamber a trail of water droplets is used to mark the passage of a charged particle, and in a bubble chamber the trail is marked by vapor bubbles in an otherwise clear liquid. Photographs of tracks in a cloud chamber and in a bubble chamber are shown in Chapters 45 and 46.

At the present time there is no satisfactory theory which quantitatively accounts for phase transitions and for the phenomena of superheating and supercooling from the molecular viewpoint. This study is part of a branch of physics known as *statistical mechanics*.

17-8 Sublimation

The change from the solid directly into the vapor phase, although very common, is not usually observed directly because the more common vapors are usually colorless. A piece of solid carbon dioxide, which is white, goes directly into the vapor phase at atmospheric pressure. This substance, known as "dry ice," does not melt, because the liquid phase does not exist at ordinary temperatures at atmospheric pressure. Another common example is the sublimation of tungsten in an electric light bulb. When the filament is hot, some of the tungsten passes into the vapor phase and is subsequently deposited as a solid onto the cooler glass bulb, thereby blackening the bulb.

To focus our attention on the process of sublimation, let us consider a flask containing some ice, and let us suppose that the air has been completely removed from the flask. If we keep the ice at a temperature of $-10°C$, the ice will sublime, forming water vapor in the flask. This process will continue until the vapor pressure reaches a value of 1.97 mm of mercury. Thereafter the vapor pressure will remain the same as long as the temperature of the system is $-10°C$. In this state the ice and water vapor are in equilibrium. If the temperature is lowered, some of the vapor will condense, and the equilibrium vapor pressure will be lowered; if the temperature is raised, some of the ice will sublime, and the equilibrium vapor pressure will be raised. The curve *OC* in Figure 17-6 shows the relationship between the vapor pressure and the temperature of sublimation for ice. The equilibrium vapor pressure drops rapidly with decreasing temperature. In accordance with Equation (17-3), the sublimation curve has a positive slope.

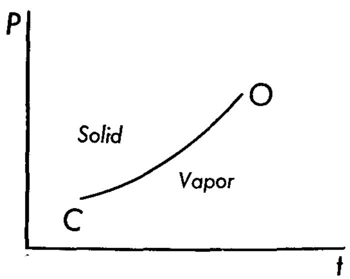


Fig. 17-6 Sublimation curve of water.

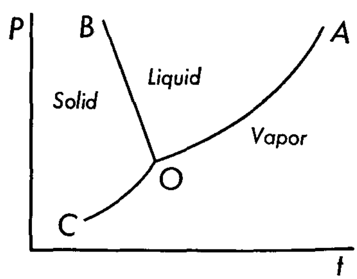


Fig. 17-7 Triple point of water.

17-9 Triple Point

The three curves for water—the vaporization curve, the fusion curve, and the sublimation curve—are plotted on a single graph in Figure 17-7. A point on a curve represents a state of equilibrium between two phases at a definite temperature and pressure. All three curves intersect at one point *O*, which is known as the *triple point*. At this point all three phases of water are in equilibrium. The temperature of the triple point is 0.01°C, and the pressure is 4.59 mm of mercury. As long as the temperature and pressure are maintained at these values, ice, water, and water vapor will coexist in the same flask and remain in equilibrium. If the temperature should be raised slightly, the ice will melt, and the state of the system will be represented by a point on the curve *OA*. If the conditions of temperature or pressure are changed in any way, one of the phases will disappear.

The temperatures at which specified phase changes occur under a standard pressure of 1,013,250 dynes/cm^2, that is, a standard atmosphere,

have provided the means of achieving a practical scale of temperature. The fundamental fixed points (the ice point and the steam point) and the primary fixed points, chosen by the Tenth International Conference of Weights and Measures in 1948, are shown in Table 17-4. In 1954 a pro-

TABLE 17-4 FUNDAMENTAL AND PRIMARY FIXED POINTS OF THE INTERNATIONAL TEMPERATURE SCALE UNDER THE STANDARD PRESSURE OF 1,013,250 DYNES/CM²

Fixed Point	Temperature
Boiling point of oxygen	−182.970°C
Melting point of ice*	0°C
Boiling point of water*	100°C
Boiling point of sulphur	444.600°C
Freezing point of silver	960.8°C
Freezing point of gold	1063.0°C

* Fundamental fixed points.

posal made earlier by Giauque was adopted, to the effect that the temperature scale be redefined on the basis of the absolute zero of temperature and the triple point of water. As shown in Section 14-3, the temperature assigned the triple point of water was 273.16° abs. At the present time it is impossible to distinguish between the temperature scale based upon the triple point of water and the temperature scale based upon the ice point and the steam point.

17-10 The *P-V-T* Surface

As we have seen, the phase of a substance is determined by a knowledge of its pressure, volume, and temperature. If the pressure, volume, and temperature of a unit mass of a substance are plotted on three perpendicular axes, a surface is generated which summarizes all the information which might be given in PT diagrams, in PV diagrams, and in tables of density. Such surfaces for a substance which expands on freezing and for a substance which contracts on freezing are shown in Figure 17-8. The condition of a unit mass of the substance is therefore determined by any two of the three variables P, V, and T, when the substance is in equilibrium. For example, when the conditions of pressure and temperature are such that the substance is in one or another of its phases, the pressure and temperature determine the volume of a unit mass of the substance and, therefore, its density. The triple point is a point only on the PT diagram. On the PVT surface the triple point is a line perpendicular to the PT plane. Thus the volume of a constant mass of a substance at the triple point may have a wide range of values. In fact, if this were not so it would be difficult, if not impossible,

to achieve a mixture of solid, liquid, and vapor in equilibrium within a closed container, for it would be necessary to introduce into the container a mass of substance precisely appropriate to the volume of the container.

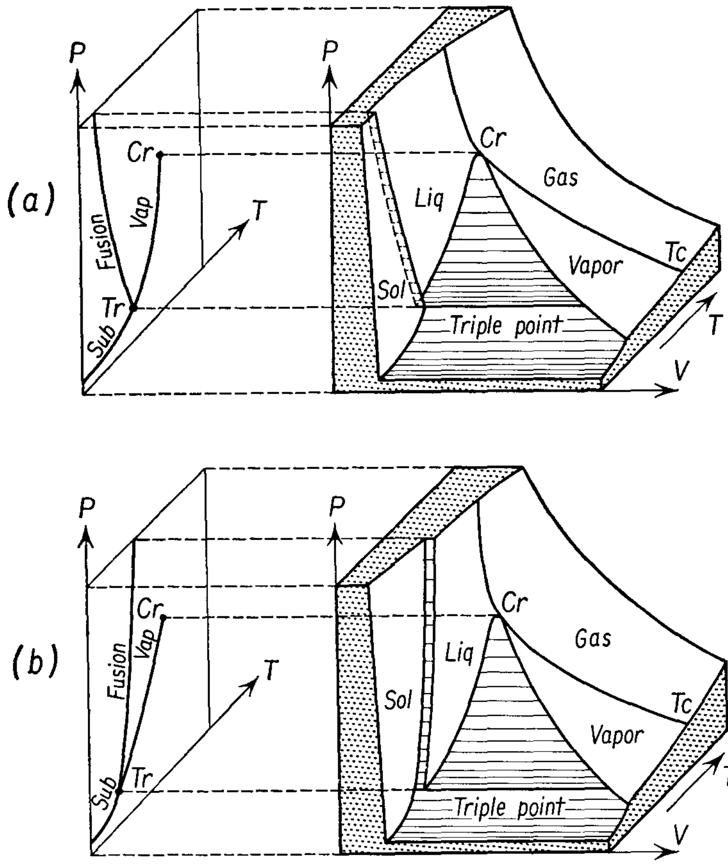


Fig. 17-8 (a) *P-V-T* surface for a substance which expands on freezing. (b) *P-V-T* surface for a substance which contracts on freezing. (By permission, from *Heat and Thermodynamics*, 4th ed., by Mark W. Zemansky. Copyright 1957. McGraw Hill Book Company, Inc.)

The *PVT* surface may be considered as a graph of an empirical *equation of state* of a particular substance, for it represents the functional relationship between the thermodynamic variables of pressure, volume, and temperature of the substance; in addition, the surface provides information about the state, or phase, of the substance. From the *PVT* surface we may note that if the pressure of a liquid is reduced, at constant temperature, the liquid passes into the vapor phase. In gasoline engines this

produces the phenomenon called "vaporlock," for the pumping of liquids in pipes depends upon the fact that a liquid is essentially an incompressible fluid, while a vapor is very compressible. Similar problems arise in connection with ships' propellers, where the reduced pressure generated by the rotation of the propeller causes *cavitation,* in which the generation and collapse of vapor bubbles not only reduces the efficiency of the propeller but causes considerable wear of the propeller blades.

17-11 Humidity of the Atmosphere

One of the most important constituents of the atmosphere is water vapor. The amount of water vapor in the air is a variable quantity. As we have seen, a mass of air is saturated when the pressure of the water vapor in that mass of air is equal to the saturated vapor pressure at the temperature of the air. Figure 17-9 is a curve showing the saturated vapor pressure as a function of the temperature of the air, while Table 17-5 gives the saturated vapor pressure at various temperatures. Usually, however, the actual vapor pressure is less than the saturated vapor pressure. The term *relative humidity* is defined as the ratio of the actual vapor pressure to the saturated vapor pressure at the temperature of the air. Thus if r is the relative humidity, p the actual vapor pressure, and P the saturated vapor pressure at the temperature of the air, then

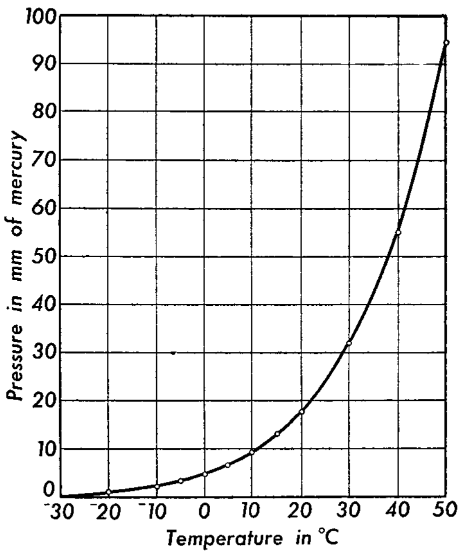


Fig. 17-9 Saturation vapor pressure curve.

$$r = \frac{p}{P} \tag{17-4}$$

and is usually expressed in percentages. For example, suppose that the actual vapor pressure is 3.0 mm of mercury when the temperature of the air is 50°F. Since the saturated vapor pressure at this temperature is 9.2 mm of mercury, the relative humidity is

$$r = \frac{3.0}{9.2} = 0.325 = 32.5 \text{ per cent.}$$

At any given temperature, the mass of water vapor in the air is proportional to the pressure of the water vapor. Hence the relative humidity can also be defined as the ratio of the mass of water vapor in a given volume of air to the mass of water vapor required to saturate it. Table 17-5 gives the mass of water vapor in a cubic meter of saturated air at various temperatures.

TABLE 17-5 PROPERTIES OF SATURATED WATER VAPOR

Temperature in °C	Temperature in °F	Pressure in Millimeters of Mercury	Grams of Water Vapor in a Cubic Meter of Air
−10	14	2.15	2.16
−5	23	3.16	3.26
0	32	4.58	4.85
5	41	6.54	6.80
10	50	9.21	9.40
15	59	12.79	12.83
20	68	17.54	17.30
30	86	31.82	30.37
40	104	55.32	51.12
50	122	92.51	
60	140	149.41	
70	158	233.7	
80	176	355.1	
90	194	525.8	
100	212	760.0	
120	248	1,489.1	
140	284	2,710.9	
160	320	4,636	
180	356	7,520	
200	392	11,659	

A simple instrument for measuring the relative humidity is the hair *hygrometer*. Human hair is used in its construction. The length of the hair increases as its cells absorb moisture from the air. The hair hygrometer consists of a bundle of human hairs exposed to the atmosphere; one end of the bundle is attached to an adjustable screw and the other end to a lever which moves a pointer over a scale. The apparatus is sketched in Figure 17-10. The scale is calibrated in terms of the relative humidity.

Another type of hygrometer consists of a combination of a wet-bulb thermometer and a dry-bulb thermometer. The two thermometers are identical except that a piece of wet muslin is wrapped around the bulb of one thermometer and kept moist by a wick dipping into a trough of water.

If the air is not saturated with water vapor, water will evaporate from the wet-bulb thermometer, producing a drop in temperature. The relative humidity is then determined from the difference in the readings obtained from the two thermometers and the tables or graphs of humidity.

17-12 Dew-Point Temperature

It is a common experience to observe moisture condensing on the outside surface of vessels containing cold beverages. This moisture is produced by the condensation of the water vapor from the air onto the cold surface. When the relative humidity is less than 100 per cent, the vapor pressure is less than the saturation pressure. The temperature to which the air must be lowered in order to become saturated with the mass of water vapor in it remaining constant is called the *dew-point temperature.* When the dew-point temperature is known, its location on the saturation pressure curve of Figure 17-9 will also give the actual vapor pressure in the air. Since the saturation vapor pressure at the temperature of the air is also known from this curve, the relative humidity is easily determined. Human comfort depends on relative humidity as well as upon temperature, and this fact must be taken into account in air-conditioning equipment. Unless the temperature and relative humidity fall within certain limits, sometimes called the comfort zone, most of us experience discomfort.

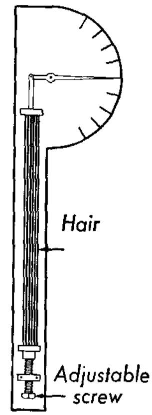


Fig. 17-10 Hair hygrometer.

Illustrative Example. When the temperature of the air is 86°F, a dew-point determination shows that the dew-point temperature is 50°F. Determine the relative humidity of the atmosphere.

From Table 17-5, the saturated pressure P at 86°F is 31.8 mm. At the dew-point temperature, 50°F, the saturated vapor pressure p is 9.2 mm. The relative humidity r is therefore

$$r = \frac{p}{P} = \frac{9.2}{31.8} = 0.29 = 29 \text{ per cent.}$$

Problems

17-1. When 1 gm of water boils at 1 atm pressure, its volume changes from 1 cm^3 in the liquid phase to 1,671 cm^3 in the vapor phase. Apply the first law of thermodynamics to this process and calculate (a) the work done by the fluid in expanding against the external pressure, (b) the change in internal energy in this process, and (c) the fraction of the latent heat of vaporization which goes into internal energy.

17-2. A 1-ton air conditioner is rated as being capable of manufacturing 1 ton of ice at 0°C in a 24-hr day, from water at 0°C. What is the rate, in watts, at which a 1-ton air conditioner can remove heat from a room?

17-3. The latent heat of vaporization of water under standard conditions is 539.6 cal/gm. (a) What is the vapor pressure of water under these conditions? (b) Using the data of problem 17-1 and Clapeyron's equation, find the vapor pressure of water at a temperature of 101°C.

17-4. Water boils at a temperature of 417°F at a pressure of 300 lb/in^2. Under these conditions 1 lb of water has a volume of 0.0189 ft^3 and 1 lb of steam has a volume of 1.5414 ft^3. The heat of vaporization of water at this temperature is 740 Btu/lb. (a) Determine the work done by 1 lb of water in changing to steam. (b) Determine the change in internal energy of the system in this process.

17-5. Prove that the enthalpy of a perfect gas is a function of the temperature only.

17-6. (a) Calculate the slope of the fusion curve of water at 0°C if the specific gravity of ice is 0.92. (b) Determine the pressure necessary to lower the melting point to -1°C.

17-7. The air admitted to a room on a winter day when the outside temperature is 32°F has a dew point of 14°F. The air is heated electrically until its temperature is 68°F. (a) What is the relative humidity of the outside air? (b) What is the relative humidity of the heated air in the room?

18

Transfer of Heat

18-1 Methods of Transmitting Heat

The methods by which heat is transmitted can be classified into three distinct types known as *convection*, *conduction*, and *radiation*. In any actual case of heat transmission, a combination of these methods may be operating simultaneously, and the principal problem is to determine the rate at which heat flows from the source at higher temperature to the source at lower temperature.

We may distinguish between the three processes of heat transfer by considering whether a medium is required for the transfer of heat, and whether that medium is at rest or in motion. In the process of conduction, thermal energy is transmitted by a medium which is at rest. The process of convection requires a moving medium. In general, a fluid transports the energy. In processes of natural convection, the density differences between the heated fluid and its cooler neighborhood generate buoyant forces which cause the heated fluid to move. Heat energy is delivered to a fluid in one region of the container and becomes internal energy of the fluid. The fluid is set in motion, and the internal energy is liberated as heat in some other portion of the container. Radiation requires no medium. The energy reaching us from the sun and stars comes in the form of radiation through the vacuous space between the sun and the earth and between the stars and the earth. Radiation may be transmitted through a substance, as in the transmission of sunlight through the air or through a windowpane, and, in so doing, changes may take place in the substance and in the character of the radiation.

Thus we see that the transfer of heat by conduction and convection requires the presence of aggregates of matter. Heat energy cannot be transferred by conduction or convection from an isolated atom or molecule, but even completely isolated atoms or molecules may gain or lose energy in the process of radiation.

18-2 Conduction

The method of transferring heat by conduction can be illustrated by means of a long cylindrical copper rod which has one end placed in a gas flame while the other end is placed in a mixture of ice and water, as shown in Figure 18-1. The amount of heat which is conducted through the copper rod in any time interval, assuming that the loss of heat to the surrounding

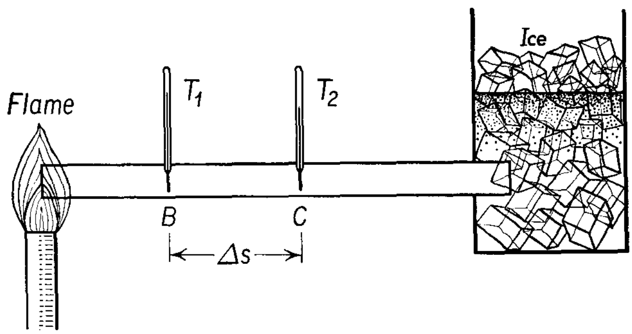


Fig. 18-1 Method of measuring the temperature gradient along a conductor of heat to determine its conductivity.

atmosphere may be neglected, can be measured by the amount of ice which is melted in this time. When a steady flow is established, the temperature of any point along the rod remains constant. Thus the quantity of heat flowing into any element of volume of the rod in a given time interval is equal to the quantity of heat flowing out of that volume element in the same time interval. If this were not so, the temperature of the volume element would be changing with time, in contradiction to the hypothesis that a steady state was established.

Any two points along the rod of Figure 18-1 are at different temperatures. Let us suppose that the points B and C have temperatures which differ in amount by ΔT, and that these two points are separated by a distance Δs along the rod. The quotient $\Delta T/\Delta s$ is called the *temperature gradient* in this region of the conductor. The greater the temperature gradient, the greater is the amount of heat which flows through this portion of the rod in any given time interval from the region at higher temperature B to the region at lower temperature C. The process of conduction may be thought of as the transfer of heat from any one part of the rod to a neighboring part, because of the difference in temperature existing between these two parts.

The rate at which heat is transferred by conduction is found to depend upon the temperature gradient and upon the cross-sectional area of the rod.

We may write

$$\Delta Q = kA \frac{\Delta T}{\Delta s} \Delta t, \tag{18-1}$$

in which ΔQ is the quantity of heat transferred through a rod of cross-sectional area A in a time interval Δt when the temperature gradient along the rod is $\Delta T/\Delta s$. The factor k is a constant of proportionality which

TABLE 18-1 THERMAL CONDUCTIVITIES

Substance	k in $\dfrac{\text{cal}}{\text{cm sec°C}}$	k in $\dfrac{\text{Btu}}{\text{ft hr°F}}$
Metals		
Aluminum	0.49	118
Brass	0.26	63
Copper	0.91	225
Gold	0.71	169
Iron	0.16	39
Lead	0.084	20
Nickel	0.14	34
Platinum	0.17	41
Silver	0.99	242
Tin	0.15	37
Tungsten	0.38	92
Insulators		
Aluminum foil, crumpled	($\frac{3}{8}$-in. air spaces) 0.0001	0.025
Asbestos, sheets	0.0004	0.097
Insulating brick, kaolin	0.0006	0.15
Glass, window	0.0012–0.0024	0.3–0.6
Snow	0.0011	0.27
Fluids		
Air	0.000054	0.017
Water	0.0015	0.37

depends upon the material of the rod and upon the units in which the other quantities of the equation are measured: k is called the *thermal conductivity* of the rod. Two sets of units are commonly used for expressing the thermal conductivity. In cgs units Q is expressed in calories, A in square centimeters, T in degrees centigrade, s in centimeters, and t in seconds. From Equation (18-1) we see that the appropriate units for k are $\dfrac{\text{cal}}{\text{cm sec°C}}$. In

British units Q is often expressed in Btu's, A in square feet, T in degrees Fahrenheit, s in feet, and t in hours, so that the appropriate units of k are $\frac{\text{Btu}}{\text{ft hr}\,^\circ\text{F}}$. The thermal conductivities of metals are generally greater than those of other solids, and silver is the best conductor of all. It is also interesting to note that those substances which are good conductors of heat are also good conductors of electricity. The thermal conductivities of some more common substances are given in Table 18-1. Although conduction does take place through liquids and gases, their conductivities are very small, gases being among the poorest conductors. Many insulating materials are constructed so that they trap small quantities of air in small closed spaces and thus make use of the poor conductivity of the air for insulation, and at the same time avoid the transfer of heat through the air by convection.

Let us consider the process of conduction by a gas at rest; that is, there is no net flow of the gas, and the average velocity of the molecules of the gas in any direction is zero. If a layer of gas is at a higher temperature than an adjacent layer, the mean kinetic energy of the molecules in the high-temperature layer is greater than the mean kinetic energy of the molecules in the low-temperature layer. In the collision between a molecule from the higher-temperature region and a molecule from the lower-temperature region, energy is transferred to the slower molecule. Macroscopically, we view this process as the transfer of heat by conduction. When the gas is at very low pressure, molecules of the gas may travel from one wall of the container to the other without striking a second molecule of the gas. In a collision between a gas molecule and a wall, we assume that the molecule leaves the wall with the mean kinetic energy of the molecules of the wall. Thus a gas molecule absorbs energy at the high-temperature wall and delivers energy at the low-temperature wall. The molecules of the gas move across the apparatus without making a collision with other gas molecules. One cannot describe any region of the gas as having a temperature, for the molecules of the gas are not in equilibrium with each other, nor is it possible to have a temperature gradient within the gas.

The internal energy of a nonmetallic solid is associated with the vibrations of the atoms or molecules of the crystal about their mean positions in the crystal lattice. The molecules in a region at higher temperature may be thought to be vibrating with greater amplitude than the molecules in an adjacent region at lower temperature. These vibrations are transmitted from molecule to molecule by the forces which hold the crystal together, similar to the way in which a wave is propagated down a string. In a metallic solid the transmission of thermal energy by lattice vibrations is much smaller than the transmission of energy by a second mechanism—the transmission of energy by the free electrons of the metal. When atoms

of a metal are assembled in a solid, the outermost electrons of the metal are relatively free to drift from atom to atom. These free electrons are responsible for the electrical conductivity of metals. These electrons behave like a free-electron gas, and they conduct heat in very much the same way that heat is conducted by the molecules of a gas. This relationship was first observed by Wiedemann and Franz in 1853, who noted that *the ratio of electrical to thermal conductivity was the same for all pure metals, at any given temperature*—a relationship known as the Wiedemann-Franz law.

Illustrative Example. A silver rod of circular cross section has one end immersed in a steam bath and the other end immersed in a mixture of ice and water. The distance between these two ends is 6 cm, and the diameter of the rod is 0.3 cm. Calculate the amount of heat that is conducted through the rod in 2 min.

The quantity of heat which is conducted across any cross section of the rod in a given time interval must be constant. Since the cross-sectional area of the rod is constant, we see, from Equation (18-1), that the temperature gradient must have the same value everywhere along the rod. Thus the temperature gradient is

$$\frac{\Delta T}{\Delta s} = \frac{100°\text{C}}{6\text{ cm}} = 16.7\,\frac{°\text{C}}{\text{cm}}\cdot$$

The thermal conductivity of silver is $0.99\,\dfrac{\text{cal}}{\text{cm sec}°\text{C}}$. We find

$$\Delta Q = kA\,\frac{\Delta T}{\Delta s}\,\Delta t = 0.99\,\frac{\text{cal}}{\text{cm sec}°\text{C}} \times 0.071\text{ cm}^2 \times 16.7\,\frac{°\text{C}}{\text{cm}} \times 120\text{ sec}.$$

Thus

$$\Delta Q = 140\text{ cal}.$$

Illustrative Example. A pipe of length l, made of material of thermal conductivity k, and having inner and outer radii r_1 and r_2, is used to heat a water bath at temperature T_2 by passing steam at temperature T_1 through the pipe. Determine the rate of flow of heat from the inside to the outside of the pipe.

From Equation (18-1) we have

$$\frac{\Delta Q}{\Delta t} = kA\,\frac{\Delta T}{\Delta s}\cdot$$

Let us consider the flow of heat through a cylindrical shell of thickness Δr at radius r. The area of the shell is

$$A = 2\pi rl,$$

and the temperature gradient across the shell is

$$\frac{\Delta T}{\Delta r},$$

so that

$$\frac{\Delta Q}{\Delta t} = 2\pi rlk\,\frac{\Delta T}{\Delta r}\cdot$$

When the pipe is in thermal equilibrium, the rate of flow of heat through each cylindrical shell must be the same as the rate of flow of heat through the pipe. Thus the quantity $\Delta Q/\Delta t$ is a constant and is independent of r. To evaluate this quantity we must apply the known conditions at the boundary of the pipe. To do this we first rewrite the above equation in a form appropriate to infinitesimally thin shells, replacing the quantity $\Delta T/\Delta r$ by dT/dr, and integrate from the inside to the outside of the pipe. Thus we obtain

$$\frac{\Delta Q}{\Delta t}\int_{r_1}^{r_2}\frac{dr}{r} = 2\pi lk\int_{T_1}^{T_2} dT.$$

Remembering that

$$\int\frac{dx}{x} = \log_\epsilon x = \ln x,$$

we find

$$\frac{\Delta Q}{\Delta t}\ln\left(\frac{r_2}{r_1}\right) = 2\pi lk(T_2 - T_1),$$

so that

$$\frac{\Delta Q}{\Delta t} = \frac{2\pi lk(T_2 - T_1)}{\ln\left(\dfrac{r_2}{r_1}\right)}.$$

If the pipe is of copper, $k = 0.91\,\dfrac{\text{cal}}{\text{cm sec}^\circ\text{C}}$. Suppose that the pipe is of inner diameter 2.5 cm and outer diameter 3 cm, and that the temperature of the steam is 120°C while the temperature of the water bath is 20°C. The rate of heat transfer through a pipe 100 cm long is

$$\frac{\Delta Q}{\Delta t} = \frac{2\pi\times 100\text{ cm}\times 0.91\,\dfrac{\text{cal}}{\text{cm sec}^\circ\text{C}}\times(20^\circ\text{C} - 120^\circ\text{C})}{\ln\left(\dfrac{3/2\text{ cm}}{2.5/2\text{ cm}}\right)},$$

$$\frac{\Delta Q}{\Delta t} = -\frac{2\pi\times 100\times 0.91\times 100}{0.182}\,\frac{\text{cal}}{\text{sec}} = -3.14\times 10^5\,\frac{\text{cal}}{\text{sec}}.$$

The minus sign appearing in the expression indicates that the flow of heat is in the direction of decreasing temperature.

Problems of the type illustrated above are known as boundary-value problems; they are very common in physics and engineering. In these problems we know the applicable differential equation and certain conditions at the boundaries, and we cannot obtain a solution to the problem until we perform appropriate integrations.

18-3 Convection

Convection is the transfer of heat from one part of a fluid to another by the flow of the fluid, mixing the warmer parts of the fluid with the cooler parts.

As an example, consider the case of a jar of water which is heated by applying a flame at one side A, as shown in Figure 18-2. Heat is conducted through the glass to the water. As the water in contact with the glass is heated by conduction, its density decreases, and it floats to the top. Colder water moves down to replace it. The colder water, in turn, is heated; once hot, it rises because of its smaller density, thus setting up a circulation of the liquid. During this circulation the warmer parts of the liquid mix with the cooler parts, and in a short time a fairly uniform temperature is established throughout the liquid. This type of heat transfer is called *natural convection* because the motion of the fluid is due to differences in the density of the fluid. In the case of *forced convection*, a fan, a pump, or some other mechanical device is used for stirring and mixing the warmer and cooler parts of the fluid.

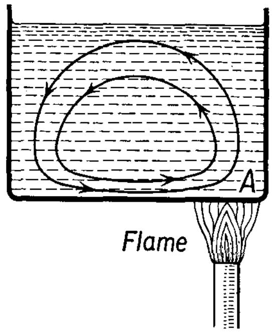


Fig. 18-2 Convection. A circulation is set up in the fluid by heating it near one end.

In almost all cases of the transfer of heat by means of fluids, both convection and conduction must be considered. The heating of a room presents several interesting illustrations of convection and conduction, and, to some extent, of radiation. If the room is heated by means of a "radiator," heat is conducted through the walls of the radiator to the air in contact with it. The warmed air rises and displaces the cooler air, thus establishing a circulation of the air in the room. The warmer air, striking the cooler walls and windows, loses heat to the outside by conduction through the walls and windows. Fortunately, there is always a film of stagnant air close to the walls and windows so that the heat which is conducted to the outside must pass through this film of air as well as through walls and windows. Since air is a very poor conductor, a thin layer of air is sufficient to form a good insulator.

The process of convection is generally much more difficult to formulate in quantitative terms than is the process of conduction. The transfer of heat by convection must be determined by experience in each particular situation. Thus one finds in handbooks typical illustrations of heat loss from vertical steampipes, from horizontal steampipes, from rough pipes, from smooth pipes, and so on, rather than a comprehensive summary of all these situations in a few simple formulas. To minimize the transfer of heat by convection, it is important that the flow of air or other fluid past the heated body be nonturbulent, that the surface be smooth rather than rough, so that a layer of stagnant fluid may be in contact with the surface, providing good insulation. Smooth pipes are often wrapped with a thin

layer of asbestos tape in the mistaken notion that the insulating tape will reduce heat losses. In reality, the rough surface of the tape causes the flow of air past the pipe to be turbulent, and the more intimate contact of the flowing air with the pipe causes greater heat loss.

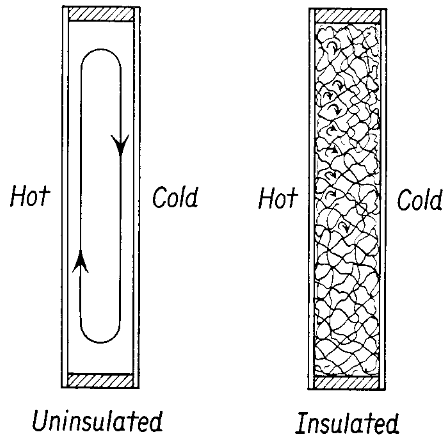


Fig. 18-3 Convection currents within the walls of a house.

There is customarily an open air space between the inner and outer walls of a house, and the transfer of heat from inside to outside in winter takes place by convection. The insulating material between the walls serves to restrict the flow of air by making many small passages in place of one large one, as shown in Figure 18-3. In the earth's atmosphere the convective flow of the air between the equator and the poles tends to equalize the temperature differential generated by the differences in solar radiation in these regions; this convective flow is largely responsible for weather and climate.

18-4 Emission and Absorption of Radiation

The radiation emitted from a warm object, which we call heat radiation, consists of *electromagnetic waves* which are identical in character with light, x-rays, radio waves, and gamma rays. These radiations travel with the speed of light, about 186,000 mi/sec, in vacuum. We shall discuss some aspects of thermal radiation more completely in the following section, while deferring discussions of other forms of electromagnetic radiation to the chapters on electronics, light, and modern physics. In this section we shall be interested in the general relationships between the process of the emission of radiation and the absorption of radiation.

A perfect absorber of radiation is a hole in a wall, for if a hole has no matter on its opposite side, the radiation incident on the hole passes

through and cannot return, for there is no matter to reflect it back. To visible light, such a hole is perfectly black; that is, it is a perfect absorber of radiation. We can approximate a perfect absorber by constructing a box, say 1 ft on a side, as shown in Figure 18-4, and by providing the box with a small hole in one face. Even though the face is painted black, the hole will look blacker. This is the case even if the inside of the box is painted white. Light incident upon the hole is reflected many times around the interior of the box before it emerges, and, though only a small fraction of the light is absorbed in each reflection, the large number of internal

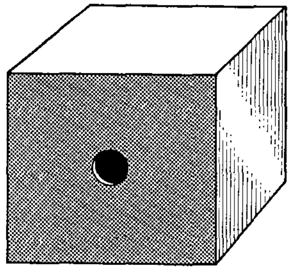


Fig. 18-4 Black body. Even though the inside of the box is painted white, and the front face is painted black, a small hole in the front face looks blacker than the black paint.

reflections greatly diminishes the intensity of the light before it emerges from the hole. The smaller the area of the hole in relation to the total surface area of the box, the more nearly the hole becomes a perfect absorber of radiation. Such a box is called a *black body* and represents a laboratory approximation to a perfect absorber. In practice, any large furnace provided with a small peephole is a good approximation to a perfect absorber, no matter what the furnace contains, and no matter what the materials of which the furnace is constructed.

When radiation is incident upon a body, some of that radiation will be absorbed and some will be reflected. The fraction of incident radiation which is absorbed by a body is called the *absorptivity* of the body. The absorptivity of a black body is unity. The absorptivity of a perfect reflector is zero. The absorptivity of any other body will have a value between zero and 1.

We can gain some insight into the relationship between the effectiveness of a body as a radiator of energy and as an absorber of radiant energy by considering a body suspended in a furnace. When equilibrium is reached, the body is at the temperature of the furnace. The body itself is radiating energy and, at the same time, is acting as an absorber of the energy radiated from the walls of the furnace. Since the body remains at constant temperature, it must radiate exactly as much energy as it absorbs. The body must be as effective in emitting radiation at a particular temperature as it is in absorbing radiation at that temperature. If the body were more effective in absorbing radiation than in emitting radiation, it would

soon become hotter than the furnace, while if it were less effective an absorber of radiation than an emitter, it would continue to grow colder than the furnace. Since these effects are not observed in experience, we must conclude that a black body, which is most effective in absorbing radiation, will also be most effective in radiating energy. The effectiveness of a body in radiating energy is called its *emissivity* e. The emissivity of a body is the rate at which a body radiates energy divided by the rate at which a black body radiates energy at the same temperature. The emissivity of a black body is equal to unity. The emissivity of any other bodies lies between zero and 1 and is identical with the absorptivity of the body.

The argument we have indicated for describing the reciprocal relationship between effectiveness in radiation and absorption is a perfectly general one and does not depend upon the nature of the radiator or upon the kind of radiation. Thus *a good radiator must be a good absorber*, and vice versa. If we wish to find whether a particular radio antenna will be a good transmitting antenna and we do not have suitable instruments available to make this determination, we can study its effectiveness as a receiving antenna. An acoustical enclosure will be an effective radiator of sound if it is a good absorber of sound. One very effective type of loud-speaker baffle is called an infinite baffle, which simply consists of a hole in a wall to which a loud-speaker is bolted. In the same way a simple harmonic oscillator will absorb sound energy most effectively at those frequencies at which it radiates such energy. Similar considerations apply to atoms, molecules, and nuclei, and we can identify the presence of certain atoms in the outer atmosphere of the sun by the light they absorb in the sun's spectrum.

18-5 Radiation

The thermal radiation emitted from a heated body varies in color and intensity with the temperature. According to an analysis by Joseph Stefan (1835–1893) of the radiation emitted from a heated body, the rate R at which energy is radiated from a unit area of a body at absolute temperature T is given by the equation

$$R = e\sigma T^4, \tag{18-2}$$

where σ is a constant which depends upon the units used. In cgs units the value of σ is

$$\sigma = 5.672 \times 10^{-5} \frac{\text{erg}}{\text{sec cm}^2\ \text{deg}^4},$$

while in mks units the value of σ is

$$\sigma = 5.672 \times 10^{-8} \frac{\text{joule}}{\text{sec m}^2\ \text{deg}^4}.$$

When an object of emissivity e and at temperature T_1 is placed within an enclosure at temperature T_2, the body radiates energy to the enclosure and absorbs energy from the walls of the enclosure. The rate at which energy is radiated by the body is given by Equation (18-2). To find the rate at which energy is absorbed from the walls of the enclosure, we observe that, if the body were at the temperature of the enclosure, it would be in thermal equilibrium. The energy the body absorbs is thus equal to the energy it would radiate in the same time if it were at the temperature T_2, the temperature of the enclosure. The net energy radiated by the body per unit time is the difference between the rate at which energy is radiated and the rate at which it is absorbed. Assuming that the emissivity is independent of temperature, the net rate of radiation from a body at temperature T_1 which is within an enclosure at temperature T_2 is given by

$$R = e\sigma(T_1^4 - T_2^4). \tag{18-3}$$

Equation (18-3) represents the rate at which energy, or heat, is transmitted by a body per unit area of surface. If A is the area of the surface, then the rate at which heat is radiated or absorbed is given by

$$\frac{\Delta Q}{\Delta t} = AR = Ae\sigma(T_1^4 - T_2^4). \tag{18-4}$$

In Equation (18-3) a positive value of R indicates that the body is radiating more energy than it absorbs, as would be the case when the body is hotter than its surroundings, while a negative value indicates that the body is radiating less energy than it absorbs, as would be the case if the body is cooler than its surroundings.

Illustrative Example. Calculate the rate at which energy is radiated from a tungsten ribbon filament 1 cm long and 0.2 cm wide which is maintained at a temperature of 2727°C. The emissivity of tungsten at this temperature is 0.35. Neglect the radiation absorbed by the filament from the room at 20°C.

Neglecting the contribution from the room, symbolized by T_2 in Equation (18-4), we find

$$\frac{\Delta Q}{\Delta t} = 1\text{ cm} \times 0.2\text{ cm} \times 0.35 \times 5.672 \times 10^{-5}\,\frac{\text{erg}}{\text{sec cm}^2\,\text{deg}^4} \times (3000°)^4$$

$$= 64.3 \times 10^7\,\frac{\text{erg}}{\text{sec}} = 64.3\text{ watts.}$$

18-6 The Spectrum of Black-Body Radiation

The radiant energy emitted by a black body may be examined by opening a hole in the side of a furnace and letting some of the radiation emerge. It is commonly known that the color of the light emitted by a furnace changes

as the temperature of the furnace rises, glowing first a deep red, then orange, and then white. These are not pure colors; they may be analyzed into their component wavelengths by means of a spectroscope (see Chapter 39). The longest wavelengths are invisible and are in the infrared region, that is, beyond the red. The wavelengths of this radiation decrease progressively

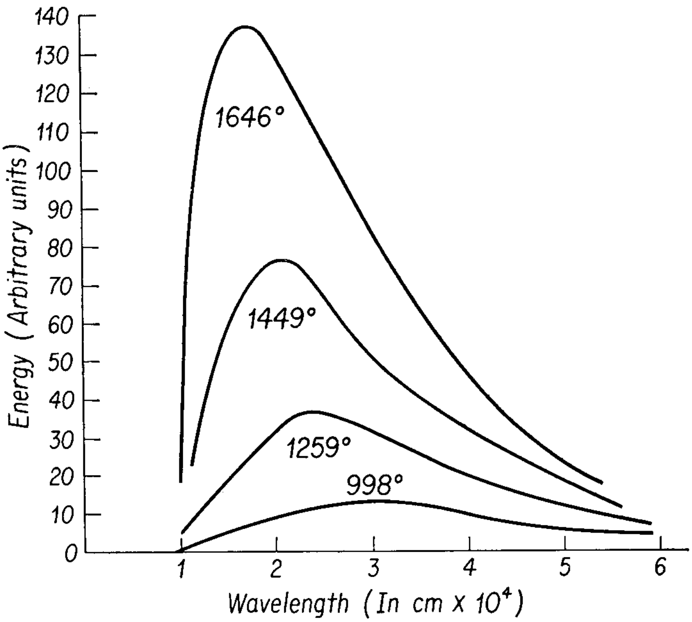


Fig. 18-5 Distribution of the energy among the wavelengths in the spectrum of a black body at different temperatures in °K.

through the colors red, orange, yellow, green, blue, and violet, and finally to the invisible ultraviolet. The distribution of energy as a function of wavelength is shown in Figure 18-5. As the temperature of the furnace increases, the wavelength of maximum intensity steadily decreases, and the intensity at every wavelength increases in such a way that the total energy radiated increases as the fourth power of the absolute temperature, according to Equation (18-2). In the figure the total energy radiated is represented by the area under a curve.

At room temperature the radiation is invisible to the eye, but even at body temperature there is sufficient infrared radiation emitted for a sensitive detector to locate the source of radiation. It has been possible to construct detectors of infrared radiation sensitive enough to locate the chimney of an industrial installation, or the engine of an airplane, and these detectors are of considerable military importance. When the tem-

perature increases to about 700°C, the radiation becomes visible, and the body is said to be red hot. At 1500°C there is a sufficient amount of the shorter-wavelength radiation present in the spectrum for the radiation to appear nearly white. By examining the color of the light from a furnace, it is possible to determine its temperature by means of a device called an *optical pyrometer*. Through the examination of the spectrum of the sun and the stars by means of a spectroscope, it is possible to determine their temperatures. The temperature determined in this way is called the *color temperature*.

The energy emitted from a heated object appears in all parts of the electromagnetic wave spectrum. Thus radio waves have been detected in the solar spectrum as radio noise, and at the other extreme, x-rays, of very short wavelength, have been detected in solar radiation by equipment installed in rockets sent high into the earth's atmosphere. While the relative amounts of energy in these regions of the spectrum are quite small, nevertheless, recent studies of the solar spectrum have made it possible to detect these invisible radiations.

18-7 Heat Insulation

A thorough understanding of the subject of the transmission of heat will enable one to solve the very important problem of heat insulation. This involves the use of proper materials for a given job as well as the development of new insulating materials. For example, gasoline storage tanks are frequently coated with aluminum or other reflecting material to reduce the absorption of radiation from the sun. Insulating materials are constructed so that they contain many small pockets of air to make use of the very low conductivity of the air; there is practically no convection since the air is trapped in these pockets. Crumpled aluminum containing small air pockets is a very good insulator; there is practically no transfer of heat by convection. The transfer by conduction is very slight since the crumpling of the aluminum makes the conducting path very long while the cross-sectional area is very small, and very little heat is transferred by radiation.

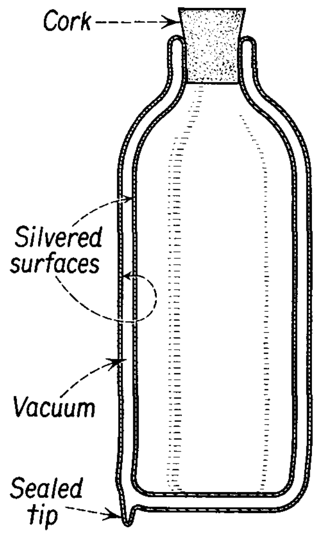


Fig. 18-6 Section of a glass thermos bottle.

The ordinary thermos bottle, sketched in Figure 18-6, is an excellent illustration of heat insulation. The thermos bottle consists of two cylin-

drical glass flasks sealed together at the top. The inside surface of the outer cylinder and the outside surface of the inner cylinder are silvered. Then the air between the two walls is pumped out, and the space is sealed off. If hot food is placed inside the bottle and the bottle is then corked, the food will remain hot for a long time. Very little heat will be conducted along the glass or through the cork to the outside: there is practically no convection, since there is a good vacuum between the walls, and radiation is reduced considerably by the silver coatings.

Problems

18-1. A metal rod 100 cm long and 4 cm^2 in cross-sectional area conducts 40 cal of heat per minute when the ends of the rod are maintained at a difference in temperature of 80°C. Determine the coefficient of thermal conductivity of the metal.

18-2. A copper rod 60 cm long and 8 mm in diameter has one end immersed in steam and the other end immersed in a mixture of ice and water. Determine the amount of heat which will be conducted through the rod in 5 min.

18-3. An aluminum pan has a diameter of 25 cm and is 0.5 cm thick. What is the rate of flow of heat through the bottom if the pan contains boiling water and is transmitting heat to it from a stove at a temperature of 400°C?

18-4. Water in a glass beaker is boiling away at the rate of 35 gm/min. The bottom of the beaker has an area of 200 cm^2 and is 0.2 cm thick. Calculate the temperature of the underside of the bottom of the beaker if $k = 0.002$ cal/cm sec °C.

18-5. An oven used for baking glass x-ray tubes is made of sheets of asbestos 0.75 in. thick. This oven is 4 ft high, 1 ft wide, and 1 ft deep. At what rate must heat be supplied to this oven to maintain the temperature inside at 400°C if the temperature of the air just outside the asbestos is 100°C?

18-6. Calculate the rate at which energy is radiated from a black body whose temperature is 1500° abs if its surface area is 1 cm^2.

18-7. Calculate the rate at which energy is radiated from a tungsten filament which is maintained at the temperature of 2500° abs if its surface area is 0.30 cm^2 and its emissivity at this temperature is 0.30.

18-8. The operating temperature of a 50-watt incandescent bulb is 2500° abs. The emissivity of the filament is 0.30. Find the surface area of the filament.

18-9. If the incandescent bulb of Problem 18-8 is placed within an enclosure at a temperature of 500° abs, what power must be supplied to the bulb to keep it at its normal operating temperature? Neglect convection losses from the surface of the glass bulb.

18-10. A blackened copper sphere initially at a temperature of 0°C is placed within an evacuated furnace which is held at a temperature of 500°C. If the sphere is 1 cm in diameter, what will be the initial rate of change of temperature of the sphere?

18-11. A copper rod 1 m long has been "turned" so that the first 60 cm of its length is 2 cm in diameter, while the last 40 cm of its length is 1 cm in diameter.

The thick end is maintained at a temperature of 100°C, while the thin end is maintained at a temperature of 0°C. Find (a) the temperature at the junction between the thick and thin ends and (b) the rate of heat flow. [HINT: When equilibrium has been reached, the same rate of flow must exist at all points within the rod. If this were not so, some part of the rod would be absorbing heat, and its temperature would rise.]

18-12. A cubical ice chest is to be constructed of sheet asbestos 1 in. thick. The chest is to have an inside volume of 1 ft^3 and is to hold 25 lb of ice. How long will the block of ice last if the outside temperature is 100°F?

18-13. A uniform sheet of concrete of thickness 8 in. separates a reservoir at 40°F from a reservoir at 160°F. The thermal conductivity of the concrete is $k = 12 \frac{\text{Btu in.}}{\text{hr ft}^2\ °\text{F}}$. Find (a) the temperature gradient and (b) the heat flow through the concrete in Btu/ft^2-hr. (c) A 4-in.-thick slab of glass of conductivity $k = 6 \frac{\text{Btu in.}}{\text{hr ft}^2\ °\text{F}}$ is introduced on the low-temperature side of the concrete. Find the temperature of the interface between the concrete and the glass.

18-14. A cube of copper 1 cm on a side and of mass 7 gms is heated to a temperature of 727°C. The cube is in an evacuated box whose walls are held at a constant temperature of 27°C. The cube cools by radiation alone. (a) How much heat has been lost when the temperature of the cube has dropped by 1°C? (b) How long does it take for the temperature to drop by 2°C? The specific heat of copper is 0.1 cal/gm °C, and its emissivity is 0.3.

18-15. The earth receives energy from the sun at the rate of 1.94 calories per minute per cm^2, called the *solar constant.* The radius of the earth's orbit is 149×10^6 km and the diameter of the sun is 1.39×10^6 km. Assuming the sun to be a black body, find the surface temperature of the sun. (See Figure 36-2.)

18-16. Using the data in the second illustrative example of Section 18-2, find the temperature gradient (a) at the inner surface and (b) at the outer surface of the pipe.

19

Heat Engines

19-1 Heat-Engine Cycles

In this chapter we shall consider the physical principles underlying the operations of heat engines because of the intrinsic importance of these principles and because of the part they have played in the development of fundamental physical ideas. *Heat engines* are designed and built to *convert* heat into work. In most cases the heat is obtained from the combustion of a common fuel such as coal, oil, gasoline, or natural gas. An important new source of heat that is just beginning to be used, and will be used more extensively in the future, is the *mass which is converted into energy* by means of a process called *nuclear fission*. Several power plants are now in operation which get the heat for their engines from the nuclear fission of the element uranium.

There are many different types of heat engines; we shall present brief descriptions of the operations of a few of them. In general, a heat engine utilizes a *working substance*, usually steam, or a mixture of fuel and air, or fuel and oxygen, through a series of operations known as a *cycle*. The working substance goes through a series of changes of state in this cycle, as a result of which some of the heat which has been supplied to it from a source at a high temperature is converted into work which is delivered to some external agency. Experience shows that not all of the heat supplied is converted into work; the heat which has not thus been utilized is delivered by the engine to some outside reservoir at a lower temperature.

The actual processes that occur in the operation of a heat engine are fairly complex. We can, however, simplify matters by replacing the actual heat-engine cycle by an ideal cycle which can produce the same transformations of heat and work. In such an ideal engine the working substance starts in some state designated by its pressure, its volume, and its temperature, is taken through a cycle in which its state continually changes, and then is brought back to its original state; the cycle then starts over again.

The operation of an ideal heat engine can be represented schematically

by the diagram shown in Figure 19-1. A quantity of heat Q_1 is delivered to the engine during one cycle by some source of heat, and the engine performs an amount of work $\mathcal{W}$ on some outside agency and rejects an amount of heat Q_2 to another reservoir of heat. Since the substance in the engine returns to its original state at the end of the cycle of operations, it contributes no energy to this cycle. There is therefore no change in the internal energy of the working substance; that is,

$$U_f - U_i = 0.$$

From the first law of thermodynamics applied to this cycle, we get

$$Q_1 - Q_2 = \mathcal{W}. \tag{19-1}$$

The *thermal efficiency* e of a heat engine is defined as

$$e = \frac{\text{work done during one cycle}}{\text{heat added during one cycle}},$$

or

$$e = \frac{\mathcal{W}}{Q_1}. \tag{19-2}$$

Substituting the value for $\mathcal{W}$ from Equation (19-1), we get

$$e = \frac{Q_1 - Q_2}{Q_1} = 1 - \frac{Q_2}{Q_1}. \tag{19-3}$$

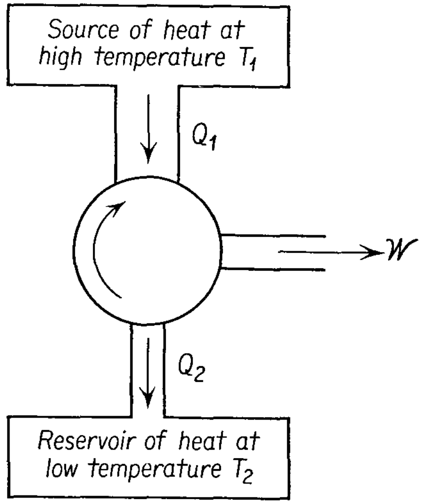


Fig. 19-1 Schematic diagram of the operation of a heat engine.

Equation (19-3) shows that the thermal efficiency of an engine is less than 100 per cent because a quantity of heat Q_2 is not transformed into work during the cycle. Experience shows that every heat engine rejects some heat during the exhaust stroke; one need merely recall the hot gases coming from the exhaust of an automobile engine or the steam exhausted by the engine of a steam locomotive. No engine has been built which takes in a quantity of heat Q_1 from some source and converts it completely into work.

19-2 The *P-V* Diagram

As we have already seen in Section 16-4, the work done by a gas at a pressure P in expanding through an increment of volume ΔV is given by $P\,\Delta V$. It is therefore convenient to plot the behavior of a gas on a diagram having pressure as the ordinate and volume as the abscissa when our interest is focused upon the mechanical work done on or by the gas. As the pressure

or the volume of a gas is changed, the temperature may also be changed; heat may be added to the gas or taken from the gas. To describe properly the changes in the state of the gas, it would be necessary to represent the initial state of the gas as a point on a P-V-T diagram, and to represent the changes in the state of the gas as a curve in this diagram. Such a curve is called the *thermodynamic path* and represents the succession of values of

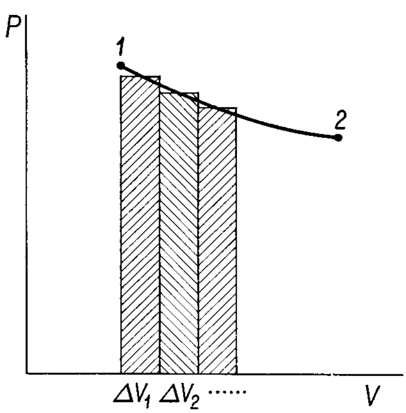


Fig. 19-2 Thermodynamic path projected onto a P-V plane.

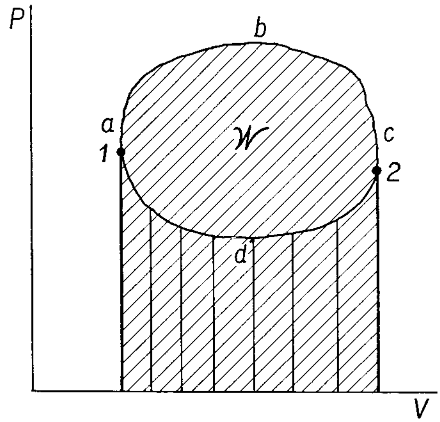


Fig. 19-3 A cycle of operations of an engine.

pressure, volume, and temperature of the gas as its state is changed. If the thermodynamic path is projected onto the P-V plane, as in Figure 19-2, the changes in the temperature of the gas are not shown on the path, but the work done in "moving" the gas along its path from condition 1 to condition 2 through a series of volume increments ΔV_1, ΔV_2, and so on, is clearly the area under the path.

Let us suppose that we alter the state of the gas from condition 1 to condition 2, in Figure 19-3, by either of the two alternate paths represented by *adc* or *abc*. Since the work done is represented by the area under the curve, the work done along the upper route is greater than the work done along the lower route. But since in either case the state of the gas, that is, the internal energy of the gas, at condition 2 at the end of each path is the same, we must assume that different quantities of heat were delivered to the gas along the two routes, to be consistent with the first law of thermodynamics.

If the gas is taken from the point *a* around the path *abcda* back to its initial condition, the device would operate as an engine, for the work done *by* the gas in expanding is greater than the work done *on* the gas when it is compressed, and the net work done by the gas $\mathscr{W}$ is represented by the area within the closed path. Since the gas at the end of the cycle is at the

same pressure, volume, and temperature as it was at the beginning of the cycle, its final internal energy must be the same as its initial internal energy, and, since $U_f = U_i$, we must have

$$\mathscr{W} = Q;$$

that is, the work done by the gas in its passage around the closed path must result from the conversion of an equal quantity of heat Q which was supplied to the gas during the cycle. In the cycle of Figure 19-3 neither the changes in temperature of the gas nor the places where heat entered the gas or left the gas are shown.

The properties of a gas make it a likely substance for use in the conversion of heat energy into mechanical work. The volume changes associated with changes in the pressure or temperature of a gas are essential to the performance of mechanical work, for it is clear that any substance used as a working substance in a heat engine must be capable of changing its volume if it is to be able to do mechanical work.

If the passage around the closed cycle of Figure 19-3 had been made in the counterclockwise direction, as *adcba,* instead of in the clockwise direction, the device would operate as a refrigerator, for the work done by the gas in expanding along the path *adc* would have been less than the work required to compress the gas along the path *cba,* and the work $\mathscr{W}$, equal to the area within the closed path, would have been done on the gas. In accordance with the sign convention established in Section 15-6, work done on a system is negative work, so that the first law of thermodynamics leads to the result that a negative amount of heat has been added to the system; that is, a net quantity of heat Q equal to $\mathscr{W}$ in magnitude has been removed from the gas.

Thus any apparatus which carries a gas through a closed reversible cycle can be operated as an engine, converting some of the heat absorbed in the path *abc,* when the gas was expanding, into mechanical work, the heat not converted into mechanical work being removed from the gas when it was compressed along the path *cda.* The same apparatus operated in reverse would absorb heat during its expansion over the path *adc,* and, as the result of the mechanical work performed on the gas, would deliver the absorbed heat plus an additional quantity of heat equivalent to the work done on the gas during its compression over the path *cba.* A reversible heat engine can thus be operated as a refrigerator simply by reversing the sense of the thermodynamic path.

19-3 The Carnot Cycle

An interesting cycle from the theoretical point of view is the *Carnot cycle.* This consists of two isothermal processes and two adiabatic processes.

Although any material may be used as the working substance, we shall make use of an ideal gas as the working substance. Suppose that this gas is contained in a metal cylinder with a tight-fitting piston, as shown in Figure

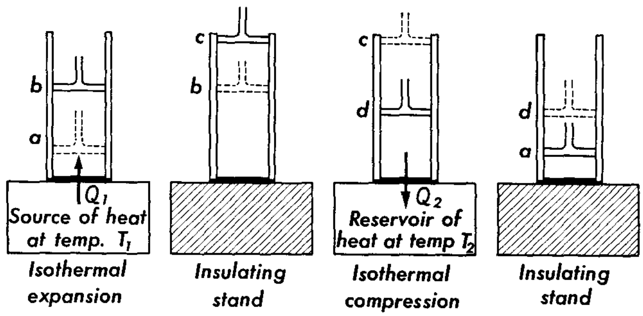


Fig. 19-4 Steps in the operation of a Carnot cycle.

19-4. Let the walls of the cylinder and the top of the piston be covered with thick layers of insulating material. Suppose we place the cylinder on a stove or other source of heat; heat will flow through the bottom of the cylinder into the gas. When equilibrium is reached, the temperature of

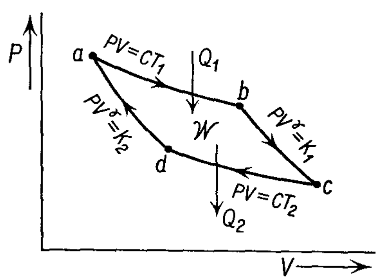


Fig. 19-5 Graphical representation of the steps in a Carnot cycle with an ideal gas as the working substance.

the gas T_1 will be the same as that of the source, and its pressure and volume will be represented by the point a in the graph of Figure 19-5. Now let the gas expand slowly so that its temperature remains T_1 while its volume increases to the point b. During this isothermal expansion, a quantity of heat Q_1 is delivered to the gas. The curve ab represents the isothermal expansion of the gas, and the area under it represents the work done by the gas during this expansion. Now imagine that the cylinder is placed on an insulated plate so that no heat can flow in or out of the cylinder. Let the gas now expand adiabatically from volume b to volume c. The amount of work done during this adiabatic expansion is the area under the curve bc. As a result of this expansion, the temperature of the gas will drop to some value T_2. Now place this cylinder on another reservoir at a temperature T_2; this reservoir may consist of a mixture of

ice and water, for example. Now compress the gas isothermally to point *d*. During this process, work will be done *on* the gas equal to the area under the curve *cd*; at the same time, a quantity of heat Q_2 will be delivered to this reservoir. Now place the cylinder on the insulating stand and compress the gas *adiabatically* until it is back to the state *a*. During this adiabatic compression, work will be done *on* the gas equal to the area under the curve *da*. The cycle has now been completed.

During this Carnot cycle, a quantity of heat Q_1 was delivered to the gas at temperature T_1 and a quantity of heat Q_2 was rejected by the gas to the lower reservoir at temperature T_2, and the net work $\mathscr{W}$ was delivered to the outside. This work is equal to the area enclosed by the curves *abcda*. Since the internal energy of the gas was restored to its original value, the work $\mathscr{W}$ is given by

$$\mathscr{W} = Q_1 - Q_2, \tag{19-1}$$

and the thermal efficiency is

$$e = 1 - \frac{Q_2}{Q_1}. \tag{19-3}$$

19-4 Absolute Thermodynamic Temperature Scale

The Carnot cycle is used to define the *absolute thermodynamic scale of temperature.* The properties of the working substance do not enter into the calculation of the efficiency. The only quantities which enter into this discussion are the temperatures of the two heat sources. Let us now arbitrarily define these two temperatures by the following relationship:

$$\boxed{\frac{Q_1}{Q_2} = \frac{T_1}{T_2};} \tag{19-4}$$

that is, the ratio of these two temperatures is the ratio of the quantities of heat extracted from and delivered to these sources by an engine operating in a Carnot cycle between these two temperatures. The efficiency of the Carnot engine now becomes

$$e = 1 - \frac{T_2}{T_1}. \tag{19-5}$$

We see that its efficiency can be 100 per cent only if the temperature of the lower heat source is 0° on this scale.

We can now choose the size of the degree to suit our convenience. In the scientific scale known as the Kelvin scale of temperature, the difference between the temperature of boiling water at atmospheric pressure and the

temperature of melting ice at atmospheric pressure is set equal to 100°, thus

$$T_s - T_i = 100°. \tag{19-6}$$

This makes the size of the degree on the Kelvin scale the same as that on the centigrade scale. As we shall see, the temperatures on the Kelvin scale defined by Equation (19-4) are identical with the temperatures previously introduced for the absolute gas scale of temperature. But the Kelvin scale is independent of the properties of any particular substance. On this scale, the temperature of the ice point is $T_i = 273.15°K$. From now on, we shall make no distinction between the Kelvin scale and the absolute scale. Another scale used by engineers sets $T_s - T_i = 180°$. This scale is called the absolute Fahrenheit scale of temperature. On this scale, $T_i = 491.8°$ abs F.

For the sake of definiteness, let us suppose that 1 mole of an ideal gas is carried through the Carnot cycle of Figure 19-5. When the thermodynamic path is an isotherm, such as *ab*, or *cd*, its temperature, and therefore its internal energy, remains constant. The heat Q_1 delivered to the gas in its expansion from a to b must be equal to the work $\mathscr{W}_1$ done in this expansion. From Equation (16-14) we have

$$Q_1 = \mathscr{W}_1 = RT_1 \ln \frac{V_b}{V_a}\cdot$$

Similarly, in the isothermal compression in the path *cd*, we have

$$Q_2 = \mathscr{W}_2 = RT_2 \ln \frac{V_c}{V_d}\cdot$$

Dividing the first of these equations by the second, we find

$$\frac{Q_1}{Q_2} = \frac{T_1}{T_2} \frac{\ln \dfrac{V_b}{V_a}}{\ln \dfrac{V_c}{V_d}}\cdot$$

The equations describing the isothermal and adiabatic processes of an ideal gas in the changes of state in the Carnot cycle are $P_aV_a = P_bV_b$; $P_bV_b^\gamma = P_cV_c^\gamma$; $P_cV_c = P_dV_d$; $P_dV_d^\gamma = P_aV_a^\gamma$. Multiplying the left-hand sides of these equations together, and setting this equal to the product of the right-hand sides of the equations, we find, on factoring the product $P_aP_bP_cP_d$,

$$V_aV_b^\gamma V_cV_d^\gamma = V_bV_c^\gamma V_dV_a^\gamma,$$

or

$$(V_bV_d)^{\gamma-1} = (V_cV_a)^{\gamma-1},$$

from which

$$\frac{V_b}{V_a} = \frac{V_c}{V_d}\cdot$$

Substituting this result into the above expression for the ratio of the heat absorbed and emitted over the isothermal portions of the cycle, we find

$$\frac{Q_1}{Q_2} = \frac{T_1}{T_2},$$

in agreement with Equation (19-4). Thus we see that the definition of temperature based upon the properties of an ideal gas is exactly equal to the definition of temperature on the Carnot cycle.

The Kelvin temperature scale represents a remarkable achievement in terms of the abstract ideas involved in the concept of temperature. Heretofore the concept of temperature was intimately bound up with the properties of matter. The very conception of the existence of an absolute zero was based upon experiments on the expansion properties of real gases at ordinary temperatures. Such a concept proved to be useless at very low temperatures where all gases liquefied. Although the practical measurement of temperature is still based upon the properties of matter, the meaning of temperature has become significantly different as the result of the absolute thermodynamic temperature scale, for now we see that temperature is intimately related to the efficiency of an ideal heat engine, the Carnot engine. We see that temperature is a measure of the *quality of heat,* for the success of any attempt to convert a given quantity of heat to mechanical energy depends upon the temperature at which that heat is available, in relation to the *ambient temperature,* the temperature of the surroundings. This, in turn, tells us a great deal about the quality of a fuel or of a combustion process. Two different fuels having the same heat of combustion do not generally produce the same amount of mechanical work. The fuel which burns at the higher temperature produces heat of higher quality, for that heat can be converted more efficiently into mechanical work.

19-5 The Second Law of Thermodynamics

It is a matter of general experience that heat always flows from a hotter to a colder substance, unless some external device is employed. So general is this observation that we could phrase a tentative hypothesis, that *heat of its own accord, will always flow from a higher to a lower temperature.* We might even extend this idea to assert the impossibility of constructing a device which, of itself, would move heat from a colder to a hotter substance. There are many devices which move heat from colder to hotter substances in everyday use, such as household refrigerators, but these are connected to the outside world through an electrical outlet. If the household refrigerator is unplugged from the power line, it no longer is able to remove

heat from the freezer compartment at low temperature and deliver it to the kitchen at higher temperature.

There are many ways to formulate this second fundamental principle of thermodynamics, and they are all equivalent to each other. One form of the second law of thermodynamics is:

It is impossible to construct an engine, which, operating in a cycle, will produce no effect other than the extraction of heat from a reservoir and the performance of an equivalent amount of work.

This statement implies that every engine operating in a cycle which takes in heat from some source or reservoir must deliver some of this heat to a reservoir at a lower temperature. From the discussion of the preceding paragraph, the definition of the absolute zero is based upon the Carnot engine. The absolute zero is thus the temperature of a reservoir to which no heat will be delivered by a Carnot engine operating between some heat reservoir at a higher temperature and the reservoir at the absolute zero. But the statement of the second law of thermodynamics asserts the impossibility of constructing an engine which will eject no heat to a low-temperature reservoir. Thus we may assert that it will be impossible to achieve the absolute zero. Devices may be constructed which will come close to the absolute zero of temperature, but no device can be constructed which can achieve this temperature.

Many would-be inventors have sought to construct perpetual-motion machines in violation of both the first and the second laws of thermodynamics. Perpetual-motion machines, intended to deliver more energy than they receive, are examples of attempts to violate the first law. An inventor who succeeded in circumventing the second law, that is, design a perpetual-motion machine of the second kind, would be able, for example, to drive a ship across the ocean by extracting heat from the ocean's water and converting it to mechanical energy. He could operate electrical generating stations by extracting heat from the earth, or he could propel an airplane by extracting heat from the air. If our inventor could have a reservoir at the absolute zero at his disposal, he might appear to be able to achieve his goal, for his engine would expel no heat to his reservoir, so that there should be little difficulty about keeping the reservoir at the absolute zero once it had been cooled to that temperature. Unfortunately, the reservoir at absolute zero is unattainable.

19-6 The Refrigerator

In principle, a refrigerator may be thought of as a heat engine operated in reverse. As shown schematically in Figure 19-6, heat Q_2 is taken from some source or sources at a low temperature, work $\mathscr{W}$ is done on the engine by means of some outside agency such as an electric motor, and a quantity

of heat Q_1 is delivered to a source at a higher temperature. The source from which heat is extracted is usually the food in the refrigerator. The source which received the heat Q_1 is usually the air surrounding the refrigerator. In the electrically operated refrigerator, the electric motor runs a compressor which consists essentially of a cylinder, a piston, and two valves

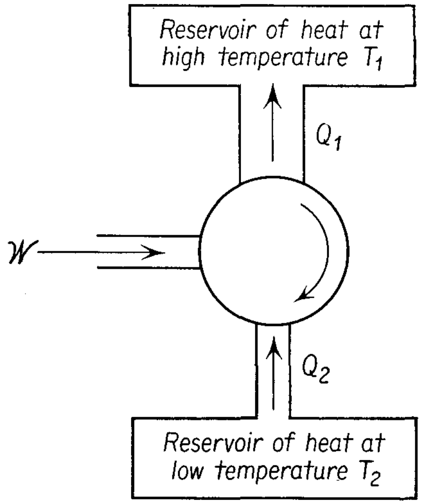


Fig. 19-6 Schematic diagram of the operation of a refrigerator.

just like the cylinder of a steam engine. The working substance used is called the refrigerant and may be ammonia, sulphur dioxide, or any other substance whose boiling point is fairly low. The refrigerant is taken through a cycle of operations which is described below, and at the end of this cycle its internal energy remains unchanged. For the first law of thermodynamics as applied to the refrigerator, we can write

$$Q_1 - Q_2 = \mathscr{W},$$

or

$$Q_1 = Q_2 + \mathscr{W}; \tag{19-7}$$

that is, the heat delivered to the air in the kitchen is greater than that taken from the food.

A typical cycle of operations for the refrigerant is as follows: suppose we start with the refrigerant, say ammonia, as a liquid, at high pressure and at room temperature, and allow some of it to pass through a valve or throttle into a region of lower pressure (see Figure 19-7). This process is called a throttling process (see Section 17-5). During this process the temperature also drops, and some of the ammonia is vaporized. This mixture is now led into the evaporation chamber in which the remaining liquid is vaporized at this low temperature and pressure. It is during this process of vaporization that heat is extracted from the food and water in

the refrigerator and is used to vaporize the ammonia. The ammonia vapor is now taken into the compressor and is compressed adiabatically to a high pressure and a temperature slightly above room temperature. This compressed fluid is then sent through pipes which are cooled by the circulating air around them. It is during this process that the heat Q_1 is given out by the refrigerant and the refrigerant is brought back to its initial state.

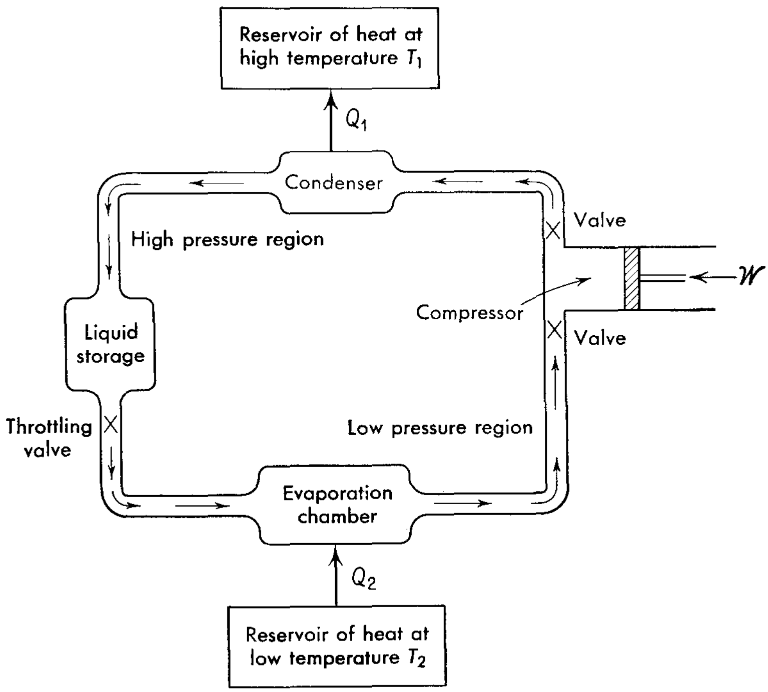


Fig. 19-7 Schematic diagram outlining the processes which occur in a refrigerator using ammonia as the refrigerant.

What is desired in a refrigerator is the extraction of an amount of heat Q_2 from the cold source with the performance of as little work $\mathscr{W}$ as possible. Instead of talking about the efficiency of a refrigerator, engineers use the term *coefficient of performance* of a refrigerator, defined as

$$K = \frac{Q_2}{\mathscr{W}} \tag{19-8}$$

where K is the coefficient of performance. In most practical refrigerators K has the value of 5 or 6. The smaller the amount of work needed to extract a given amount of heat, the greater is the coefficient of performance. For example, if 1,000 cal of heat are extracted from the food in the refrigerator and the motor which operates the compressor performs an amount of work

equivalent to 200 cal, then the coefficient of performance of this refrigerator is 5. The heat Q_1 delivered to the air in the kitchen is 1,200 cal.

19-7 Practical Heat Engines

The Carnot engine and the Carnot refrigerator are imaginary devices. No one has yet built an operating Carnot engine, and it is extremely unlikely that such a device will ever be built. Yet the Carnot engine is of very great practical value, for the imaginary experiments we have conducted with the Carnot engine have enabled us to establish the meaning of temperature more clearly than before and to set limits on the efficiency of real engines, without regard for practical problems such as minimizing friction or achieving perfect fit between piston and cylinder. In fact, we may assert that *no reversible engine operating between two given heat reservoirs at different temperatures can be more efficient than a Carnot engine,* for this would constitute a violation of the second law of thermodynamics. Let us suppose that there were an engine more efficient than the Carnot engine. We shall imagine that the second engine is operated in reverse, as a refrigerator between the two temperature reservoirs, and is driven by the Carnot engine, operating between the same two reservoirs. The two devices together would constitute a single self-acting device which would pump heat from the reservoir at low temperature to the reservoir at high temperature and would produce no other effect, which is in clear violation of the second law.

The steam engine, the gasoline engine, the diesel engine, the turbine, the jet engine, and the rocket engine are all primarily heat engines. Some of these are external-combustion engines, in which the fuel is burned in a combustion chamber and the heat is transferred to the engine proper, while others are internal-combustion engines in which the combustion of the fuel takes place within the engine itself. In all of these engines, it is the heat liberated in combustion rather than any explosion of the fuel which is converted to mechanical work. For this reason the efficiency of all these engines is fundamentally limited by the efficiency of a Carnot engine. Modern engine designers, seeking increased efficiency, are limited by the properties of materials at high temperatures, for only by operation at high temperatures can the efficiency of present-day engines be significantly improved. Much of the effort of metallurgical and ceramic research is concentrated in the study and development of materials suited to the construction of more efficient engines.

19-8 A Heat Pump

The analysis of the refrigeration cycle shows that, by the performance of a certain amount of work $\mathcal{W}$, a quantity of heat Q_2 is taken from a reservoir

at a low temperature T_2, and a larger quantity of heat Q_1 is delivered to a reservoir at a higher temperature T_1. Lord Kelvin, in 1852, suggested that this is just what is desired in the operation of a *heat pump*. It took about 75 years for the first practical heat pump to be put into operation; they are coming into more common use now for heating homes in the winter and cooling them in the summer. A schematic diagram showing the operation

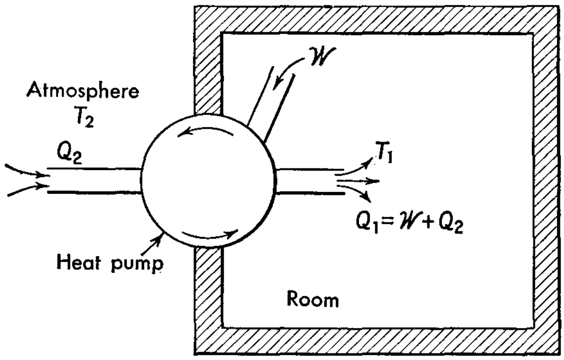


Fig. 19-8 Schematic diagram of the operation of a heat pump which takes a quantity of heat Q_2 from the atmosphere at a low temperature T_2 and pumps a quantity of heat Q_1 into the room at a higher temperature T_1 when an amount of work $\mathscr{W}$ is done on it.

of a heat pump is sketched in Figure 19-8. This heat pump will take heat Q_2 from the atmosphere and pump heat Q_1 into the room at the higher temperature T_1 with the performance of work $\mathscr{W}$ on the compressor of the heat pump. The quantity of heat Q_1 that is delivered to the room is given by

$$Q_1 = \mathscr{W} + Q_2,$$

although the energy that is paid for on the electric bill is represented by the work $\mathscr{W}$ which is done by the electric motor in operating the compressor.

By reversing the flow of the refrigerant in the heat pump, the temperature of the room can be kept below that of the atmosphere, a very desirable feature on hot summer days. In this case a quantity of heat Q_2 will be taken from the room at a low temperature T_2, a quantity of work $\mathscr{W}$ will be done on the compressor to accomplish this task, and a quantity of heat $Q_1 = \mathscr{W} + Q_2$ will be delivered to the outside.

19-9 Entropy

We have seen that in a Carnot cycle the quotient of the quantity of heat Q absorbed from or given out to a reservoir divided by the temperature T of

the reservoir is a constant, for

$$\frac{Q_1}{T_1} = \frac{Q_2}{T_2}. \tag{19-4}$$

If we regard Q as an algebraic quantity which is positive for heat absorbed by a body and negative for heat given out by a body, we may write

$$\frac{Q_1}{T_1} + \frac{Q_2}{T_2} = 0,$$

or

$$\sum \frac{Q_i}{T_i} = 0.$$

If we write ΔQ for the (algebraic) quantity of heat transferred to the engine at the temperature T over a small increment in the thermodynamic path, we have

$$\sum \frac{\Delta Q}{T} = 0. \tag{19-9}$$

The quantity

$$\Delta S = \frac{\Delta Q}{T} \tag{19-10}$$

is called the change in *entropy* ΔS. Thus the entropy change in the working substance in a cycle of operation of a Carnot engine is zero.

Let us consider the change in entropy of the universe associated with the flow of a quantity of heat Q from a reservoir at high temperature T_1 to a reservoir at low temperature T_2. Since heat has flowed out of the high-temperature reservoir, we call the heat change of the reservoir negative, and the entropy change of the high-temperature reservoir ΔS_1 is

$$\Delta S_1 = -\frac{\Delta Q}{T_1},$$

while the change in entropy of the low-temperature reservoir ΔS_2 is

$$\Delta S_2 = +\frac{\Delta Q}{T_2}.$$

The total change in entropy of the universe is

$$\Delta S = \Delta S_1 + \Delta S_2 = \frac{\Delta Q}{T_2} - \frac{\Delta Q}{T_1},$$

and, since T_1 is greater than T_2,

$$\Delta S > 0.$$

Similarly, we may consider the change in entropy of the universe when work $\Delta \mathscr{W}$ is converted to heat by friction against the surface of a reservoir

at temperature T. The work is converted to heat $\Delta Q = \Delta W$ (in appropriate units), and the entropy change of the reservoir is

$$\Delta S = \frac{\Delta Q}{T}.$$

Since no heat has passed into or out of the agent doing the work, ΔS represents the entropy change of the universe, and again ΔS is a positive quantity.

Experience shows that in every thermodynamic process the change of entropy of the universe is equal to or greater than zero, and, in fact, natural thermodynamic processes move in such a direction as to increase the entropy of the universe. This may be written mathematically as

$$\boxed{\Delta S \geqq 0.} \tag{19-11}$$

Further analysis of the concept of entropy of a substance indicates that the entropy is a measure of the statistical probability for the arrangement of molecules of the substance in a particular state. A state of perfect order might be represented as one of zero entropy. In this sense the tendency for the entropy of the universe to increase might be described as a tendency for the universe to proceed from a highly organized to a completely disorganized state, a tendency for the universe to run down. In this very general way the concept of entropy has been used to explain the tendency of isolated organic matter to decompose, and to explain that life processes which seem to increase the degree of organization of nature can only take place if an even greater disorganization is taking place elsewhere in the universe, as in the sun. If a quantity of heat ΔQ is transferred by radiation from the sun at temperature T_s to a solar furnace at temperature T_f, the change in entropy of the sun is $-\dfrac{\Delta Q}{T_s}$, while the change in entropy of the furnace is $\dfrac{\Delta Q}{T_f}$. The change in entropy of the universe in the transfer of heat energy must be greater than or equal to zero, according to Equation (19-11).

Hence
$$\Delta S = \frac{\Delta Q}{T_f} - \frac{\Delta Q}{T_s} \geqq 0.$$

Thus the temperature of the furnace cannot exceed the temperature of the sun.

Problems

19-1. (a) Determine the thermal efficiency of a Carnot engine which operates between the temperatures of 100°C and 0°C. If 1,000 cal of heat are supplied to

it, (b) how much work is done, and (c) how much heat is rejected to the low-temperature reservoir?

19-2. (a) Determine the thermal efficiency of a Carnot engine which operates between two reservoirs whose temperatures are 300°C and 0°C. If 1,000 cal of heat are supplied to it, (b) how much work is done, and (c) how much heat is rejected to the low-temperature reservoir?

19-3. Superheated steam at a temperature of 520°F is supplied to a steam engine which exhausts the steam to a condenser kept at a temperature of 60°F. Determine the maximum thermal efficiency of this steam engine.

19-4. Steam at a temperature of 227°C is supplied to a steam turbine which exhausts it to a condenser kept at a temperature of 13°C. Determine the maximum thermal efficiency of this steam turbine.

19-5. The coefficient of performance of a refrigerator is 4. It takes 3,000 cal out of a quantity of food. (a) How much work is done by the electric motor which operates this refrigerator? (b) How much heat is supplied to the surrounding air?

19-6. A tray containing 2,500 gm of water at 20°C is placed in a refrigerator whose coefficient of performance is 6. The water is changed to ice at −10°C. Determine (a) how much heat is removed from the water in converting it to ice, (b) the work done by the electric motor which operates the refrigerator, and (c) the amount of heat supplied to the surrounding air.

19-7. A two-stage turbine operates at an initial temperature of 327°C. Steam at the upper temperature is passed through the first stage and is expelled to a second stage at 227°C. The steam undergoes a further expansion in the second stage and is expelled to the atmosphere at a temperature of 127°C. Find the maximum theoretical efficiency of the turbine.

19-8. In a two-stage engine a quantity of heat Q_1 is absorbed at a temperature T_1, work $\mathscr{W}_1$ is done, and heat Q_2 is expelled at a lower temperature T_2. The second stage takes the heat expelled by the first, does work $\mathscr{W}_2$, and expels a quantity of heat Q_3 at a lower temperature T_3. Show that the efficiency of the combination is given by $(T_1 - T_3)/T_1$.

19-9. A Carnot refrigerator removes 100 cal of heat from a low-temperature reservoir at −23°C and expels 150 cal to a high-temperature reservoir. What is the temperature of that reservoir?

19-10. A refrigerator delivers heat into a room at 27°C at a rate of 1,200 watts. (a) What is the rate of extracting heat, in watts, from a low-temperature reservoir at −23°C? (b) What is the rating of the motor (in horsepower) required to operate this refrigerator?

19-11. A real engine absorbs 1,000 cal from a high-temperature reservoir at 127°C and expels 800 cal to a low-temperature reservoir at 27°C in each cycle. What is the efficiency of this engine?

19-12. One end of a copper rod is in thermal contact with a heat reservoir at a temperature of 100°C, and the other end of the rod is in contact with a heat reservoir at a temperature of 0°C. Determine the change in entropy (a) of the hot reservoir, (b) of the cold reservoir, (c) of the copper rod, and (d) of the universe when 1,500 cal of heat are transmitted through the rod.

19-13. Referring to Problem 19-1, determine the change in entropy (a) of the hot reservoir, (b) of the cold reservoir, and (c) of the universe. (d) What is the change in entropy of the working substance in a cycle of operation?

19-14. Referring to Problem 19-2, determine the change in entropy (a) of the hot reservoir, (b) of the cold reservoir, and (c) of the universe. (d) What is the change in entropy of the working substance in a cycle of operation?

19-15. A Carnot engine using helium as a working substance operates between temperatures of 27°C and 127°C. The engine does 0.1 joule of work in each cycle. Referring to Figure 19-5, determine the change in the entropy of the working substance (a) in the isothermal expansion *ab*, (b) in the adiabatic expansion *bc*, (c) in the isothermal compression *cd*, (d) in the adiabatic compression *da*, and (e) in the entire cycle.

19-16. A blunt drill is driven against a block of hardened steel by a $\frac{1}{4}$ horsepower motor. Assuming that all of the energy delivered to the motor is converted into heat, find the change in entropy of the universe in 3 minutes, assuming that the block remains at a constant temperature of 27°C.

Part Three

WAVE MOTION AND SOUND

20

Wave Motion

20-1 Vibrating Bodies and Wave Motion

Wave motion is an important method of transferring energy from one place to another without involving the actual transfer of matter. When a pebble is dropped into a still pool, some of the kinetic energy of the pebble is used to generate the ripples which spread out in all directions over the surface of the pool. When the ripples pass by a floating object, such as a bit of cork, the cork bobs up and down, having acquired its kinetic energy of vibration from the ripple system. The ripples thus serve to deliver some of the energy of the pebble to the distant cork. Although the ripple system is seen to move, there is no net flow of the water, for the cork simply bobs up and down and does not move in the direction of the ripples. Since the cork merely provides visible evidence of the behavior of the water, we may infer that the motion of the water is one of vertical oscillation, while the motion of the wave is horizontal; that is, *the oscillations of the water are transverse to the direction of propagation of the wave.* A wave motion in which the vibrations of the medium are perpendicular to the direction of propagation is called a *transverse wave.* The waves set up in a taut string when one end of the string is vibrated in simple harmonic motion are transverse waves.

A second type of wave motion may be demonstrated by the use of a long helical spring. If such a spring is suspended with its axis vertical, and one end of the spring is caused to oscillate in the vertical direction, these oscillations are transmitted down the spring as a succession of compressions and rarefactions of the spring. The direction of oscillation of any part of the spring is parallel to its axis and therefore parallel to the direction of propagation of the disturbance. A wave motion in which the oscillations of the medium are parallel to the direction of propagation of the wave is called a *longitudinal wave.*

Both transverse and longitudinal waves may be propagated within a continuous medium, provided that the medium has appropriate elastic properties. For the transmission of a longitudinal wave, a displacement of one element of the medium in the direction of propagation must be capable of exerting a force on an adjacent element. In a solid, this type of stress-strain relationship is described by Young's modulus of elasticity, while in a fluid medium the bulk modulus is the means through which the displacement of one volume element generates a force on an adjacent element. Thus both solids and liquids may transmit longitudinal waves.

The transmission of a transverse wave in a continuous medium requires the existence of a shear modulus, for the displacement of one part of the medium in a direction perpendicular to the direction of propagation must generate a force on an adjacent element which is transverse to the direction of propagation. Gases are not capable of exerting shearing forces, hence they cannot transmit transverse waves. Since sound is a wave motion generated by a vibrating body and transmitted through air, we must infer that sound waves are longitudinal waves.

20-2 Equation of Wave Propagation

When a wave is propagated in an elastic medium, each particle of the medium vibrates in simple harmonic motion. The *frequency* f of each vibrating particle is the same as the frequency of the source of vibration. Just as in the case of simple harmonic motion, it is sometimes convenient to describe this oscillation by its *period* T, or its *angular frequency* ω. If we examine the appearance of the wave in space at any one instant of time, the displacements of the particles of the medium from their equilibrium positions follow a sine or cosine function of the space coordinates. The adjacent particles of the medium are out of phase with each other. The wavelength λ is the distance between two successive crests, or two successive troughs, or between any two successive particles whose simple harmonic motion is at the same phase angle. We may relate the velocity v at which the wave is propagated to the wavelength λ and the period T by observing that in the time of one period the wave has advanced by a distance of one wavelength. Thus

$$\boxed{v = \frac{\lambda}{T} = \lambda f,} \tag{20-1}$$

for the frequency is the reciprocal of the period: $f = 1/T$.

When a wave moves to the right a small distance Δx in a time Δt, as shown in Figure 20-1, a point P may be displaced upward to a new position

P', while a second point Q is displaced downward to a new position Q'. The particles at P and Q are therefore clearly out of phase with each other.

The general equation of simple harmonic motion was given in Chapter 12 as

$$y = A \sin (\omega t + \phi),$$

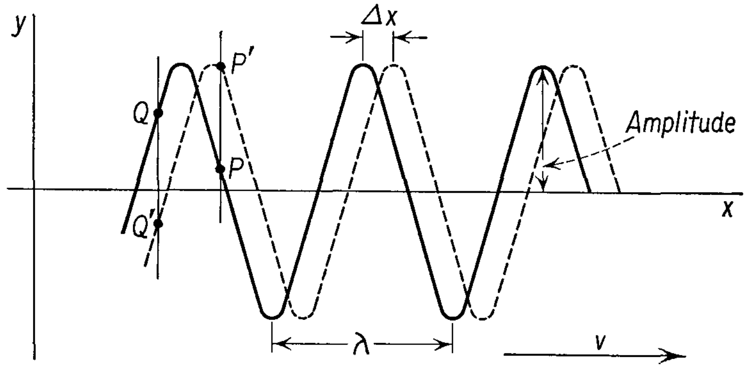


Fig. 20-1 When a wave moves to the right a distance Δx, the point Q of the medium is displaced downward to Q', while the point P is displaced upward to P'.

where A is the amplitude of the motion, and the angular frequency ω is given by

$$\omega = 2\pi f = \frac{2\pi}{T}\cdot$$

Substituting for ω, we find

$$y = A \sin \left(\frac{2\pi t}{T} + \phi\right) \tag{20-2}$$

as another form of the general equation of simple harmonic motion; that is, this equation describes the simple harmonic motion of any one particle when suitable values are inserted for the parameters A, T, and ϕ. As we have seen, the phase of a particle along a wave depends upon its position. The phase changes at a regular rate with displacement along the wave, and in a distance of one wavelength the phase changes in amount by 2π. For a wave which is moving to the right, we may set

$$\phi = -\frac{2\pi x}{\lambda}$$

in Equation (20-2) to obtain

$$y = A \sin 2\pi \left(\frac{t}{T} - \frac{x}{\lambda}\right)\cdot \tag{20-3}$$

For a particular instant of time t, fixed in value, we observe that the displacement y of a particle from its equilibrium position is given by a sine curve of amplitude A. The displacement of a point at any position along the wave is repeated a distance λ farther along. If the position x is held fixed, the value of the displacement y at this point in space varies periodically, repeating its motion after a time interval T; this is the period of its motion.

Let us consider the direction in which the wave described by Equation (20-3) is propagated. When we watch the motion of a system of ripples we can tell which way the wave is going by noting the translational motion of a particular high light of one of the ripples of the system. At the position of the high light, the displacement of the water from its equilibrium position is always the same. Thus the high light is a position of constant phase angle. In Equation (20-3) a constant phase angle is represented by a constant value of the quantity within the parentheses. Thus Equation (20-3) must describe a wave of amplitude A which moves in the direction of increasing x; that is, the wave must move to the right. If, instead of the minus sign in the parentheses, we had written a plus sign, it would have been necessary for x to decrease as t increased in order to keep the value of the parentheses constant. The wave would then have moved to the left.

The speed of the wave represented by Equation (20-3) may be obtained by noting that, if t is increased by one period T, the wave need have moved a distance of one wavelength λ in that time for the parentheses to remain at constant value. Thus the speed of the wave is given by Equation (20-1). The speed v represented by the equation $v = \lambda/T$ is called the *phase velocity* of the wave, for it represents the speed of points of constant phase.

Equation (20-3) is a general equation of wave motion and is used much as one uses the general equation for a straight line in analytic geometry. Specific cases of wave motion are described by substituting appropriate numbers for A, λ, and T into the general equation.

Illustrative Example. A sine wave of amplitude 1 m and of wavelength 4 m is moving to the right with a speed of 2 m/sec. Find the displacement of a point 25 cm to the left of the origin when $t = 1.5$ sec.

A sine wave moving to the right is described by the equation

$$y = A \sin 2\pi \left(\frac{t}{T} - \frac{x}{\lambda}\right).$$

As indicated in the problem, $A = 1$ m, $\lambda = 4$ m; T may be obtained from the equation

$$v = \frac{\lambda}{T},$$

so that

$$T = \frac{\lambda}{v} = \frac{4\text{ m}}{2\text{ m/sec}} = 2\text{ sec.}$$

We wish to find y when $x = -0.25$ m and $t = 1.5$ sec. Substituting into the equation, we have

$$\begin{aligned} y &= 1\text{ m} \times \sin 2\pi\left(\frac{1.5\text{ sec}}{2\text{ sec}} - \frac{-0.25\text{ m}}{4\text{ m}}\right) \\ &= 1\text{ m} \times \sin 2\pi\left(\frac{13}{16}\right) \\ &= 1\text{ m} \times \sin 292^\circ \\ &= -0.93\text{ m}. \end{aligned}$$

20-3 Huygens' Principle

When a source of waves is placed in a uniform continuous medium, waves spread out from the source in all directions. The locus of points of constant phase is a spherical surface, just as the locus of points of constant phase in a two-dimensional ripple is a circle. Such a wave in a continuous medium

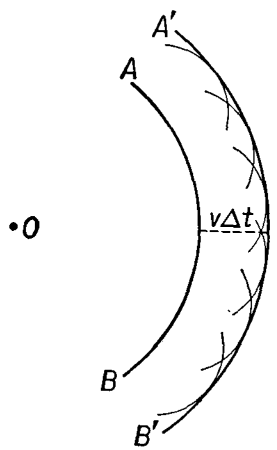


Fig. 20-2 Huygens' construction for determining the position of wave front AB after a time interval Δt. $A'B'$ is the new position of the wave front.

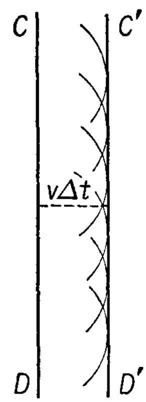


Fig. 20-3 Huygens' construction for determining $C'D'$, the new position of the plane wave front which started at CD at a time Δt earlier.

is called a *spherical wave*. We may restrict the spherical wave from a distant source by means of appropriate apertures so that only a very small portion of the wave passes through the aperture. In this case we may approximate the wave front by a plane surface. Such a wave is referred to as a *plane wave*.

The progress of a wave under a variety of conditions can be predicted

with the aid of a principle first enunciated by the Dutch physicist Christian Huygens (1629–1695). According to *Huygens' principle, each point in a wave front can be considered as a source of waves*, and the new position that the wave front will occupy after the lapse of a small time interval Δt can be found by drawing the envelope to all of the small waves from all of the individual points on the initial wave front at the beginning of this time interval. To illustrate the use of Huygens' principle, let us consider the progress of a spherical wave front from a point source O, as shown in Figure 20-2. If the arc of circle AB represents the position of a section of the wave front at a certain time t, each point on the wave front is imagined to emit spherical waves which are of radius $v\ \Delta t$ at time $t + \Delta t$. The new wave front $A'B'$ is the envelope of all these small spheres. The same type of construction can be used to find the subsequent position of any type of wave front. Figure 20-3 shows a plane wave front CD progressing with speed v. The new position of the wave front after a short time interval Δt is $C'D'$, the envelope of all the small waves, each of radius $v\ \Delta t$ emitted by each of the points on the wave front CD.

20-4 Reflection of a Wave

When a wave which is traveling in one medium reaches the surface of a second medium, part of the wave is *reflected* back into the first medium, and the rest penetrates into the second medium and is said to be *refracted* into the second medium. Let us suppose that a plane surface S separates two media I and II in which the velocities of propagation of the wave are v_1 and v_2 respectively, as shown in Figure 20-4.

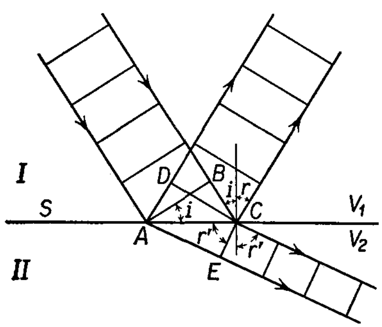


Fig. 20-4 Reflection and refraction of a plane wave at a surface separating two media.

The position of the reflected wave and the direction of its motion relative to that of the incident wave can be found with the aid of Huygens' principle. The angle that each incident wave front makes with the surface is called the *angle of incidence* i, while the angle each reflected wave front makes with the surface is the *angle of reflection* r. A line drawn in the direction of motion of the wave is called a *ray;* a ray is perpendicular to a wave front. A line drawn perpendicular to the interface between the two media is called a *normal.* Thus the angle of incidence is equal to the angle between the incident ray and the normal, while the angle of reflection is the angle between the reflected ray and the normal. Figure 20-5 shows

the plane wave advancing toward the surface S at an angle of incidence i. Let us consider the plane wave front represented by the line AB such that the point A has just reached the surface S, while B is still moving toward it; The angle between AB and the surface S is the angle i. If the surface had not been there, the wave front AB would have advanced to the position $A'C$ in a time Δt. But as the different parts of the wave front reach the surface, each point on the surface, such as A, e, f, g, h, . . . C, becomes a source of waves. By the time the point B on the incident wave front reaches the position C on the surface S, the Huygens wavelet which started

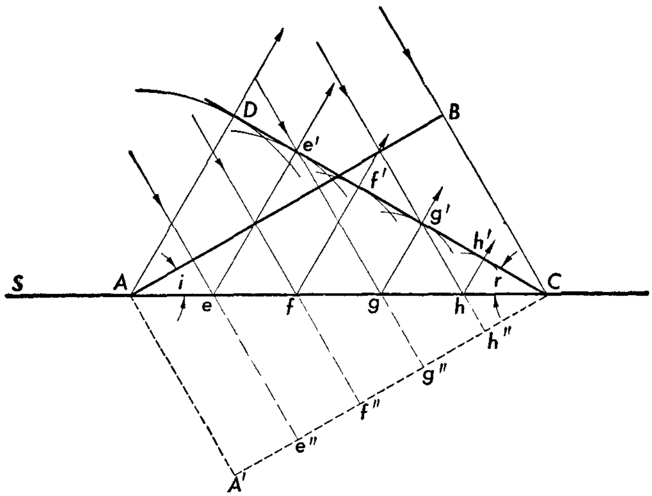


Fig. 20-5 Huygens' construction for determining the position and direction of motion of the wave front reflected by surface S. AB is the incident wave front and CD is the reflected wave front.

from A has grown to radius AD. Similarly, the wavelet initiated at point e when the incident wave front struck the surface has grown to radius ee', and so on. The envelope $De'f'g'h'C$ tangent to these spherical wave fronts is the plane wave front reflected by the surface. The angle of reflection r is the angle between CD and the surface S. Since the velocity of propagation is a property of the medium and is the same for the incident and reflected waves, we see from the construction that the *angle of incidence is equal to the angle of reflection.* The direction of the reflected wave is indicated by the arrows on the rays AD, ee', and so on. The reflected waves appear to originate somewhere behind the reflecting surface.

If a wave from some point source P strikes a flat surface and is reflected from it, the reflected wave will appear to come from a point P' behind the surface. This point P' is the *image* of P and is located directly

behind the surface so that the surface is the perpendicular bisector of the line PP', as shown in Figure 20-6. An observer in the region QQ' who wishes to determine the source of the waves does so by extending the direction of the rays reaching him in the backward direction until they intersect in a point. The direct rays PQ and PQ' thus appear to originate at point P, while the reflected rays CQ and $C'Q'$ appear to originate at the point P', the image point. It may be readily shown from the construction that the distance OP is equal to the distance OP' by making use of the fact that the angle of incidence is equal to the angle of reflection.

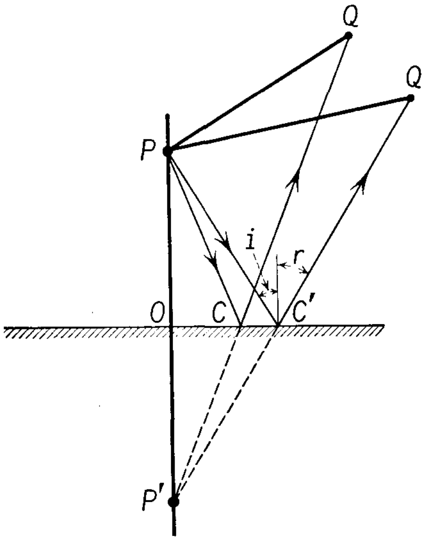


Fig. 20-6 Image formation by a plane mirror; P' is the image of P.

The walls of a room usually act as mirrors for the sounds produced in the room. In rooms of average size these reflections add to the intensity of the sounds. In a large auditorium the reflected sound may reach the hearer a considerable time after he has received the directly transmitted sound. In the case of speech this effect may be very objectionable; in the case of music the overlapping of different sounds may be pleasing to the ear.

If a taut string is struck a blow near one end, a single *pulse* is formed which travels down the string toward the other end, as shown in Figure 20-7(a). When the pulse strikes the wall at the end of the string, it is reflected so that both the direction of propagation and the direction of the deflection are opposite to that of the original pulse as shown in Figure 20-7(b). If we regard the upward deflection of the string as positive, the downward deflection of the reflected pulse is negative, and we may describe the alteration in the sense of the pulse by saying that *a phase change of* $180°$ *occurs on reflection.*

Another way of viewing the phenomenon of reflection is to imagine that the surface from which a wave is reflected is truly a mirror, as in Figure 20-7(c), and that all events which occur in fact on the left-hand side are reproduced in the opposite sense on the right-hand, or image, side. Thus if an upward pulse is initiated at the point P of a real string, a downward pulse is initiated at the image point P' of the image string. Both pulses may be imagined to move toward and through the wall at the same speed. When the two pulses reach the wall, the two opposite deflections

cancel, and there is no deflection of the string, in agreement with the physical restriction that the string is tied to the wall. We may imagine that the image pulse then passes through the wall and becomes visible as a real reflected pulse, as in Figure 20-7(d).

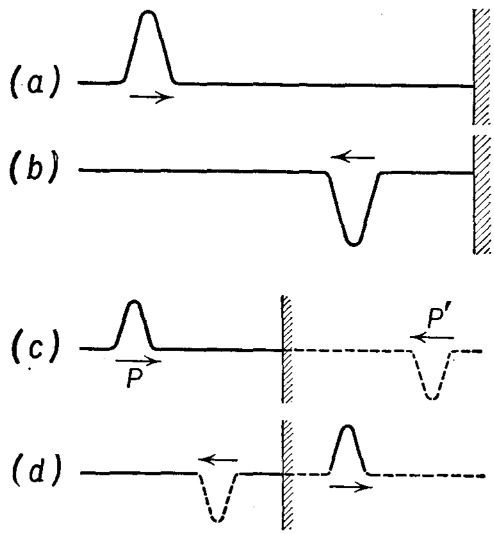


Fig. 20-7 Change of phase of a pulse on reflection from a wall.

20-5 Standing Waves

When a particle is simultaneously subjected to two vector displacements, its resultant displacement is the vector sum of the two individual displacements. At a point in a medium where the paths of two waves intersect, the medium is simultaneously displaced by the two waves, so that its resultant displacement is the vector sum of the individual displacements. This is known as the *principle of superposition.* We have already utilized the concept of superposition in the last paragraph of the preceding section.

When a wave strikes a reflecting surface normally, that is, at zero angle of incidence, the wave is reflected back at the same angle, and consequently along the same line. If a continuous wave is propagated in this medium, the incident and reflected waves will *interfere* with each other. If the two waves traveling in opposite directions through the medium have the same wavelength and the same amplitude, their effect is to set up steady vibrations, called *standing waves.*

The existence of standing waves may be demonstrated analytically by use of the general equation of wave motion. Two waves of the same wavelength, velocity, and amplitude, moving in opposite directions, may be

represented by the equations

$$y_1 = A \sin 2\pi \left(\frac{t}{T} + \frac{x}{\lambda}\right)$$

for the wave moving to the left, and

$$y_2 = A \sin 2\pi \left(\frac{t}{T} - \frac{x}{\lambda}\right)$$

for the wave moving to the right. The resultant displacement y of the medium by the simultaneous application of these two waves is the sum of the deflections due to the individual waves. Thus

$$y = y_1 + y_2.$$

From trigonometry we recall that

$$\sin (x \pm y) = \sin x \cos y \pm \cos x \sin y.$$

Applying this formula to the equations for y_1 and y_2, and adding to find y, we obtain

$$y = 2A \cos 2\pi \frac{x}{\lambda} \sin 2\pi \frac{t}{T} \cdot \tag{20-4}$$

Unlike Equation (20-3), the above equation no longer displays a wave in motion, for there is no longer a moving point of constant phase. Instead we see that the amplitude of vibration of a point in the medium of coordinate x is given by the factor multiplying $\sin 2\pi \frac{t}{T}$ and is given by $2A \cos 2\pi \frac{x}{\lambda}$. The amplitude of vibration varies from place to place on the string, and, at positions separated by half a wavelength, the amplitude of the vibration is zero. Points of zero vibration are called *nodes*. Midway between two nodes the vibration is a maximum, at points called *loops* or *antinodes*. From the above equation we note that two successive nodes are separated by a distance of half a wavelength.

If two wave motions are simultaneously imposed on a string by attaching identical tuning forks to its opposite ends, or, more simply, by attaching a vibrating tuning fork to one end of the string while the other end is attached to a rigid reflecting wall, the conditions for the generation of standing waves are fulfilled. If the string is properly adjusted, it appears to vibrate in segments. Since the eye cannot follow the vibrations with sufficient rapidity, an observer sees the string clearly where it is moving slowly and sees a blur where it is moving rapidly. Consequently, the observer sees only the envelope within which all vibrations take place, described by the function

$$y = \pm 2A \cos 2\pi \frac{x}{\lambda} \cdot$$

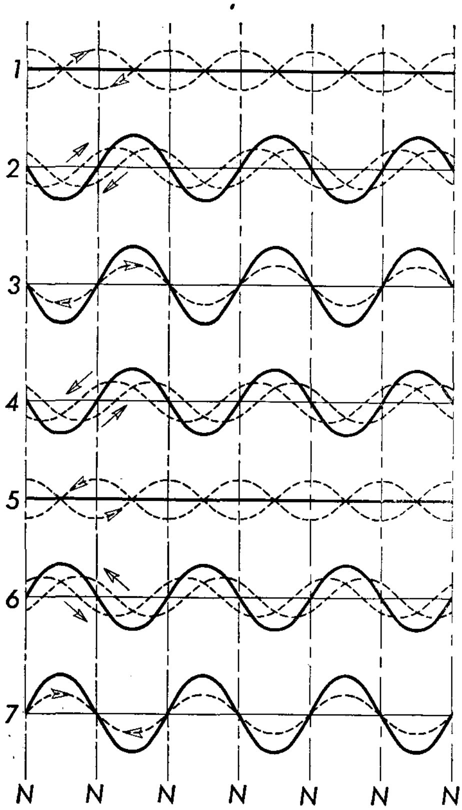


Fig. 20-8 Method of adding two waves of equal wavelengths and equal amplitudes traveling in opposite directions to produce a stationary wave.

The addition of two waves of equal amplitude and wavelength traveling in opposite directions is illustrated graphically in Figure 20-8. The dashed lines indicate the individual waves moving in opposite directions, while the heavy lines show the sum of the two waves.

If a source of sound of constant frequency and intensity is operated in a room, the walls, floor, and ceiling of the room are reflectors of sound waves, and standing waves are set up in the room. At some positions in the room, the sound intensity is a minimum; these are the nodes of the standing wave pattern. At other positions in the room, the intensity is a maximum; these are the positions of the antinodes.

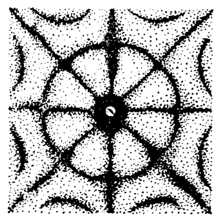


Fig. 20-9 Standing wave pattern of a vibrating plate.

The vibrations of tuning forks, of strings in musical instruments, of air columns such as those in organ pipes, and of bars and plates can all be analyzed into sets of standing waves in the substance. To detect the standing waves on a vibrating plate, sand or fine powder is sprinkled onto the plate. The standing wave pattern obtained on a horizontal plate under one specific set of conditions is shown in Figure 20-9. Each particular standing wave pattern is called a *mode of vibration* and is excited at a particular frequency. If a structure, or part of a structure, is subjected to vibration at its own resonant frequency, a standing wave pattern is generated. At the position of an antinode, the structure is subjected to alter-

nating strains; it experiences large alternating stresses, and is likely to crack. This process is known as *fatigue failure*. Cracks which develop in strange places in old automobile fenders can often be explained by this mechanism.

20-6 Diffraction of Waves

Many common phenomena, such as the bending of sound waves around obstacles and the spreading of sound waves after passing through a small aperture, are examples of the *diffraction of waves*. We use the term "diffraction" to describe wave phenomena in which the wave front is limited, as by a barrier or an aperture. Diffraction phenomena occur with all types of wave motion and may be most readily understood through the applica-

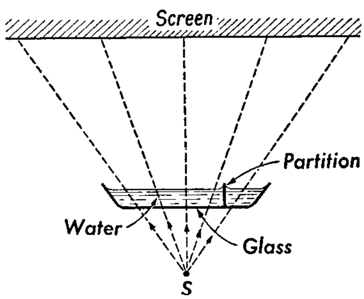


Fig. 20-10 Shallow rectangular tank for demonstrating wave phenomena. Side view showing source of light S under the glass bottom of the tank, for projection of wave phenomena.

tion of Huygens' principle. When the size of a barrier is small compared to the incident wavelength, the wave bends around the barrier. If the barrier is large compared to the wavelength, the barrier seems to cast a sharp shadow. Thus sound waves, whose length ranges from a few centimeters to many meters, may be heard in all parts of a room when sound comes in through an open window. Light coming through the same window directly illuminates only a small part of a room, for the wavelength of light is in the neighborhood of 5×10^{-5} cm. We may infer that man's physical size determines his use of sound as a communication medium and light as a means of localizing objects in space, for sound waves bend easily around man-sized objects, while light waves cast sharp shadows of these objects. The ability to produce diffraction effects is one of the criteria used to determine whether we are dealing with a wave phenomenon.

A simple method for demonstrating the diffraction of waves is illustrated in Figure 20-10. A shallow glass tank containing water to a depth of about 1 in. is illuminated from a light source S below the tank. Waves, in the form of ripples, may be generated in the tank by an oscillating source, which forms circular ripples when the source is a small rod, while plane waves are generated by a flat stick. The progress of a wave can be followed

by viewing its image on a screen. Photographs obtained with such a ripple tank are shown in Figure 20-11. The photograph is brightest where the layer of water is thinnest, as in the trough of a ripple, while the darkest regions correspond to the crest of a ripple. The distance between two successive bright bands is the wavelength. In Figure 20-11(a) a plane wave is shown incident upon an aperture of nearly the same width as the wave-

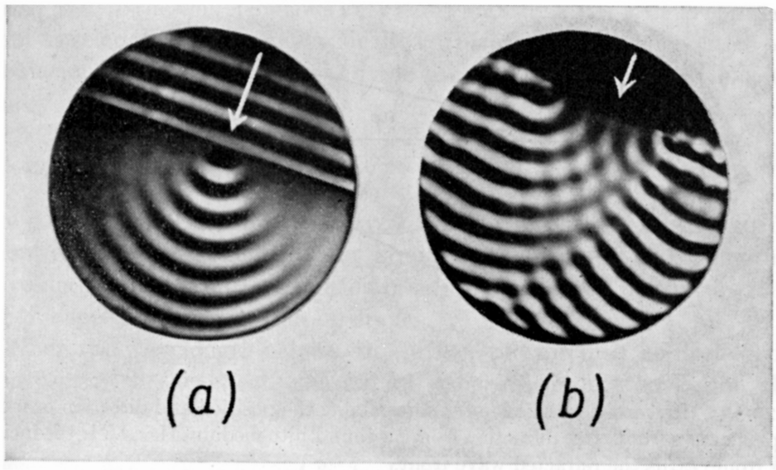


Fig. 20-11 (a) Diffraction of plane waves through an aperture (8–10 mm) in a metal barrier. (b) Diffraction of plane waves round an obstacle about 20 mm wide and production of diffraction fringes. (*Courtesy the Ealing Corporation.*)

length of the ripples. The waves are diffracted into the shadow area beyond the aperture. In Figure 20-11(b) the waves are diffracted into the shadow area behind a barrier. If we think of each point on the wave front of the incident plane wave as a source of circular wavelets, whose effects must be added in accordance with their phase relationships by the principle of superposition, we may account for these diffraction phenomena. We will consider diffraction effects in a quantitative way in Chapter 40, Light as a Wave Motion.

20-7 Refraction of a Wave

When a wave traveling in a medium I, in which its velocity is v_1, reaches an interface S between the first medium and a second in which the wave velocity is v_2, part of the wave enters the second medium at an angle r', called the *angle of refraction*. We may determine the relationship between the angle of incidence i and the angle of refraction r' with the aid of Huygens' principle, as illustrated in Figure 20-12. Suppose that the incident

wave is a plane wave, and that a portion of a wave front is represented by the line AB. The point A on the wave front has just reached the surface S. The position of the wave front in medium II, after a time interval Δt when the part of the wave front at B has reached the interface at C, may be found from Huygens' construction. From A as center we draw a spherical wave front of radius $v_2\,\Delta t$. Carrying out this construction for all points on the

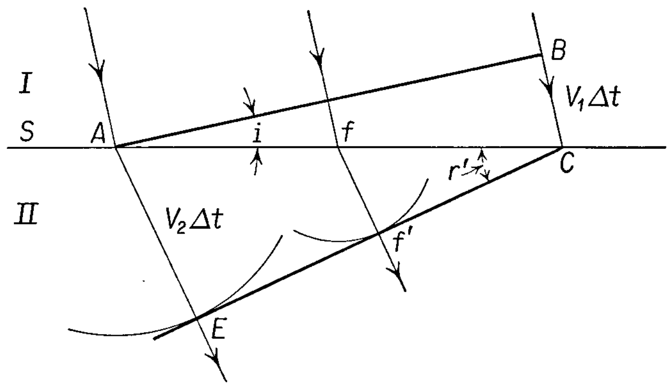


Fig. 20-12 Huygens' construction for determining the position and direction of motion of a plane wave front after refraction from medium I into medium II. AB is the incident wave front; EC is the refracted wave front.

incident wave front, we find the refracted wave front to be along the line EC, and the direction of propagation of the refracted plane wave front is normal to the wave front, or along the direction ff'. If the speed of the wave v_2 in medium II is greater than the speed of the wave v_1 in medium I, then the transmitted part of the wave will be bent away from the normal. From the figure we find

$$BC = v_1\,\Delta t,$$

$$AE = v_2\,\Delta t,$$

$$\sin i = BC/AC,$$

$$\sin r' = AE/AC,$$

from which

$$n_r = \frac{\sin i}{\sin r'} = \frac{v_1}{v_2}\,; \tag{20-5}$$

that is, the ratio of the sine of the angle of incidence to the sine of the angle of refraction is constant for all plane waves transmitted from medium I to medium II. This constant is represented by the symbol n_r, called the *relative index of refraction* of the two media.

If a wave is refracted into a medium in which its speed v_2 is less than

the speed v_1 in the first medium, the wave will be bent toward the normal: that is, the angle r' will be smaller than the angle of incidence i, according to Equation (20-5).

20-8 Surface Waves on Water

The speed of surface waves on water depends upon the depth and the wavelength. Where the wavelength is very short and the water is deep compared to the wavelength, the waves are called *ripples* and are propagated by a combination of surface tension and the force of gravity with a velocity v given by

$$v = \left(\frac{g\lambda}{2\pi} + \frac{2\pi T}{\lambda\rho}\right)^{1/2}, \tag{20-6}$$

where g is the acceleration of gravity, λ is the wavelength, T is the surface tension, and ρ is the density of the liquid. Equation (20-6) is sometimes used to determine the surface tension of a liquid of known density by measuring the speed of propagation of ripples.

When the wavelength is long, the surface tension may be neglected in comparison with the effect of gravity. The speed of propagation then depends on the wavelength and the depth and is given by

$$v = \left(\frac{g\lambda}{2\pi}\tanh\frac{2\pi D}{\lambda}\right)^{1/2}, \tag{20-7}$$

where D is the depth of the water and the symbol tanh means hyperbolic tangent and is a symbol for the expression

$$\tanh x = \frac{e^x - e^{-x}}{e^x + e^{-x}}.$$

Equation (20-7) may be approximated for short waves in deep water, when the depth is greater than half a wavelength, by the expression

$$v = \left(\frac{g\lambda}{2\pi}\right)^{1/2}, \tag{20-8}$$

so that the speed depends directly on the square root of the wavelength. The appropriate approximation to Equation (20-7) for long waves in shallow water, where the wavelength is greater than about 25 times the depth, is

$$v = (gD)^{1/2}. \tag{20-9}$$

Equation (20-9) is of great interest in determining the depth and the character of beaches, and it accounts for the riffled nature of the water above a reef. As a wave passes from one medium to a second in which it travels with a lower speed, the wave is refracted, and, in addition, its wave-

length is changed in accordance with Equation (20-1), the frequency remaining constant. Since the water is shallow over a reef, the wave velocity, and therefore the wavelength, is reduced. In the invasion of North Africa in World War II, the nature of the shore line and the location of invasion beaches was determined by measurement of the wave velocity of water waves along the coast line by observation from the air.

The motion of water waves is complex and cannot be adequately covered here. The motion of small volume elements of the water is circular or elliptical rather than purely longitudinal or transverse, and extends in depth beneath the surface, but at a depth of one wavelength the amplitude is less than 1/500 of the amplitude at the surface. Generally speaking, there is no translational motion of the water in a water wave. When the depth of water is too shallow for wave propagation, the energy of the waves is converted into translational energy of water, giving rise to the familiar breakers and surf along the shore line.

Ocean waves are generated by winds and storms at sea. Since the waves travel at much greater speed than the storm, wave observation stations have been set up along the coasts of Britain to determine the direction from which the waves come and the amplitude with which waves reach the coast. By comparing data from several stations, the progress of storms at sea can be followed, and useful weather information can be transmitted to ships at sea.

20-9 Transverse Waves

The waves transmitted down a string are transverse waves. If a slotted board is passed over a vibrating string, the slot will permit passage of the waves when it is parallel to the direction of vibration of an element of the string, but it will prevent passage of the waves when it is perpendicular to the direction of vibration. A transverse wave motion in which all of the vibrations of the medium are in the same direction is said to be *polarized.* From an experimental point of view, a wave is said to be polarized if a slit or slitlike device can be found which will permit passage of the wave in one orientation, but which will forbid passage of the wave in a second orientation at right angles to the first. Since a longitudinal wave will pass through a slit, however oriented, longitudinal waves are not polarized. The wave propagation in air, which we call sound, must be a longitudinal wave motion, by virtue of the fact that air does not possess a shear modulus. Sound waves in air are not polarized and are not polarizable.

If a wire is given an impulsive blow, the deflection of the wire is transmitted as a single pulse down the wire, with a speed v, and the pulse may be assumed to retain its shape as it moves. Any short section of the deflected part of the wire may be approximated by the arc of a circle. To

analyze the motion we imagine that the deflected portion of the wire is stationary, and that the wire moves past a template of the form of the pulse in a direction opposite to the motion of the pulse with a speed v, as shown in Figure 20-13(a). If the mass per unit length of the wire is m, the mass of a short section of the wire, whose arc is of radius r and which sub-

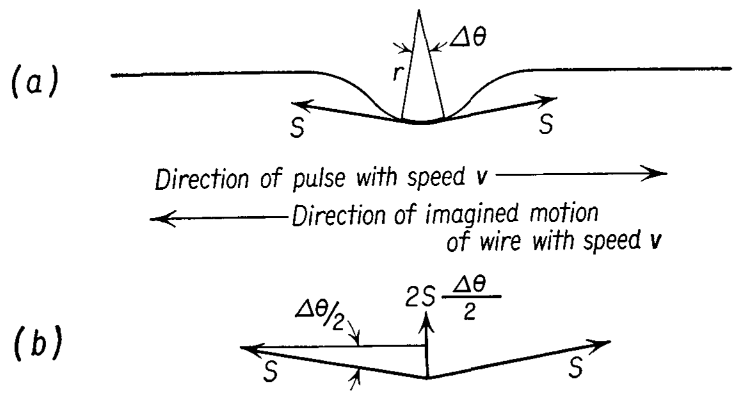


Fig. 20-13

tends an angle $\Delta\theta$ at the center of its circle of curvature, is $mr\,\Delta\theta$. The centripetal force F required to keep this element of wire moving in a circular arc is

$$F = \frac{mr\,\Delta\theta\,v^2}{r}.$$

From Figure 20-13(b) the required centripetal force is supplied by the radial components of the tension S in the wire at both ends of the circular arc. For small angles these are given by

$$F = 2S\frac{\Delta\theta}{2} = S\,\Delta\theta.$$

Thus we have

$$m\,\Delta\theta\,v^2 = S\,\Delta\theta,$$

or

$$v = (S/m)^{1/2}. \tag{20-10}$$

Illustrative Example. A long steel wire under a tension of 10^5 dynes has one end attached to a prong of a tuning fork which vibrates with a frequency of 128 vib/sec. The linear density of the wire is 0.03 gm/cm. Determine (a) the speed of the transverse wave in the wire and (b) the wavelength of this wave.

The speed of the wave in the wire can be obtained from Equation (20-10), which yields

$$v = \sqrt{\frac{10^5\text{ dynes}}{0.03\text{ gm/cm}}} = 1{,}830\text{ cm/sec}.$$

Solving Equation (20-1) for λ, we get

$$\lambda = \frac{v}{f},$$

from which $$\lambda = \frac{1{,}830}{128}\ \text{cm} = 14.3\ \text{cm}.$$

20-10 Longitudinal Waves

Waves can be set up in any elastic body by producing some disturbance at a point in the body. For example, consider a helical spring stretched between two fixed points A and B, as shown in Figure 20-14, and suppose that we compress a small section of the spring near the bottom by taking one of its coils and pushing it down. When this coil is released, it will be

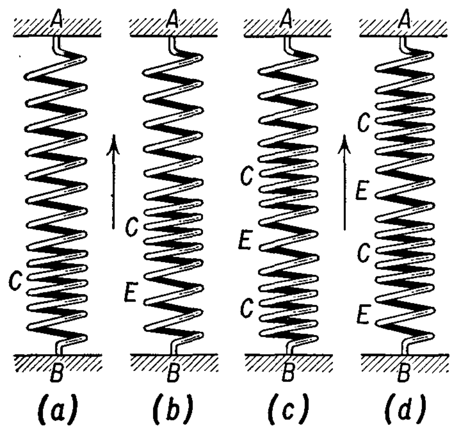


Fig. 20-14 Setting up a longitudinal wave in a coiled spring stretched between two points A and B. (a) Portion of spring C near the bottom is compressed. (b) Compression C has moved up the spring and is followed by an extension E below it. (c) and (d) Compressions C and extensions E moving up the spring.

pulled up by the coil above it and pushed up by the coil below. The displaced coil will move up to its equilibrium position and will continue on upward because of its inertia. Thus the compressed region of the spring will be displaced upward, and an *extension* will appear in the part of the spring which formerly was in *compression*. While the compressions and extensions of a coil spring move along the spring, the individual coils vibrate up and down with simple harmonic motion; the vibration is parallel to the direction of propagation of the wave, so that the wave is longitudinal.

The wave motion in air is also a longitudinal wave. Let us consider the wave in air produced by the vibratory motion of a stick, as shown in Figure 20-15. When the end of the stick moves to the right, the air next to it is compressed; this, in turn, produces a *compression* in the next layer of air, and so on, so that a compression travels out from the vibrating stick. While the compression is moving out to the right, the end of the stick starts

moving back, leaving a *rarefaction,* or region of reduced pressure, on its right. The adjacent layer of air starts moving back, so that the rarefaction travels out from the vibrating stick. This succession of compressions and rarefactions traveling out from the stick constitutes a *longitudinal* wave in air. The layers of air vibrate with the same frequency as the stick.

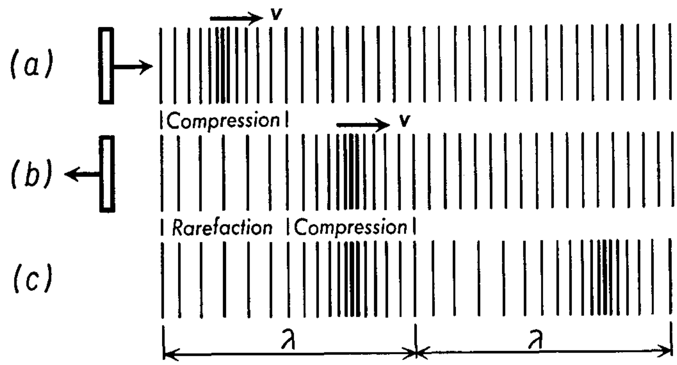


Fig. 20-15 (a) Compression in the air produced by the motion of the end of the rod to the right, followed by (b) a rarefaction produced by the motion of the end of the rod to the left. (c) Longitudinal wave in air consists of a series of compressions and rarefactions moving outward with velocity v.

The changes in pressure which occur during the compressions and rarefactions are generally very small in comparison with the atmospheric pressure. A graphical method for representing the wave is shown in

Fig. 20-16 Representation of a longitudinal wave.

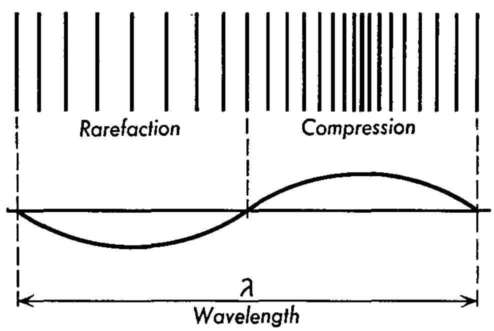


Figure 20-16, in which the change in pressure from normal atmospheric pressure is plotted along the vertical axis, and the position where this pressure difference exists is plotted along the horizontal axis. The section of the curve below the axis represents the rarefaction, while the section above the axis represents the compression.

It may be shown that the speed of propagation of a longitudinal wave within a fluid medium is given by

$$v = (B/\rho)^{1/2}, \tag{20-11}$$

where B is the bulk modulus of the fluid and ρ is its density. Although there is a temperature gradient between a compression and a rarefaction, the conductivity of a gas is so low, and the distance between them, which is a half wavelength, is so long, that very little heat can be transmitted in the available time of half a period. The processes of compression and rarefaction in a sound wave are practically adiabatic at audio frequencies. Hence we must use the value for the *adiabatic bulk modulus* in Equation (20-11). This is given by

$$B_{(\text{adiabatic})} = \gamma P, \tag{20-12}$$

where γ is the ratio of the specific heats of the gas and P is its pressure. Thus the speed of propagation of longitudinal waves in gases is given by

$$v = \left(\frac{\gamma P}{\rho}\right)^{1/2}. \tag{20-13}$$

From the general gas law

$$PV = nRT,$$

where n is the number of moles of the gas, and R is the gas constant per mole. If a mass m of a gas of molecular weight M is enclosed in a container, the number of moles of gas within the container is given by

$$n = \frac{m}{M}.$$

If we substitute this expression into the gas law and rearrange terms, we find that

$$\frac{P}{m/V} = \frac{RT}{M},$$

and, since the term on the left-hand side of the above equation is equal to P/ρ, we may substitute this result into Equation (20-13) to obtain

$$v = \left(\frac{\gamma RT}{M}\right)^{1/2}. \tag{20-14}$$

Thus the velocity of sound in a gas depends upon the absolute temperature T, but is independent of the pressure of the gas. The velocities of propagation of longitudinal waves in several substances are given in Table 20-1.

A longitudinal wave is transmitted in a wire or rod with a speed given by

$$v = (Y/\rho)^{1/2}, \tag{20-15}$$

TABLE 20-1 SPEED OF SOUND (Longitudinal Waves)

Substance	Temperature in °C	Speed in meters per sec
Air	0	331.46
Hydrogen	0	1,262
Carbon dioxide	0	258.0
Water	15	1,447
Sea water	13	1,492
Glass	0	5,500
Steel		4,700–5,200

where Y is Young's modulus. In many problems in engineering design, it is important to know Young's modulus for a particular substance at elevated and reduced temperatures. One of the simplest methods for making this determination is to measure the velocity of propagation of a longitudinal wave in the substance at the desired temperature, and to apply Equation (20-15) to this measurement.

Problems

20-1. A steel wire is stretched between two pegs 80 cm apart under a tension of 10^6 dynes. The linear density of the wire is 0.25 gm/cm. Determine the speed of a transverse wave in this wire.

20-2. One end of a horizontal string is attached to a prong of an electrically driven tuning fork which is vibrating with a frequency of 256 vib/sec, while the other end passes over a pulley and has a weight of 6 lb attached to it. The weight of 1 ft of string is 0.02 lb. (a) Determine the speed of transverse waves in the string. (b) Determine the wavelength of the waves set up in the string.

20-3. A copper wire 2 m long whose mass is 8 gm has one end attached to a fixed post and the other end attached to a prong of a tuning fork which vibrates with a frequency of 1,000 cycles/sec. A standing wave is set up in the wire, and the distance between the nodes is 8.0 cm. Determine (a) the wavelength of the transverse wave in the wire, (b) the speed of the wave, and (c) the tension in the wire.

20-4. A stone is dropped in a well, and the sound of the stone's splash is heard 4.0 sec later. How deep is the well?

20-5. A steel pipe 200 ft long is struck at one end. A person at the other end hears the sound that traveled through the pipe and also the sound that traveled through the air. Determine the time interval between the two sounds.

20-6. (a) Determine the index of refraction of hydrogen with respect to air for a sound wave. (b) Determine the index of refraction of glass with respect to air for a sound wave. A glass partition 0.5 cm thick separates a volume of air from a volume of hydrogen, each at atmospheric pressure. A sound wave from the air strikes the glass surface at an angle of 3° with respect to the normal. Determine the angle at which the sound wave is refracted (c) into the glass and (d) into the hydrogen.

20-7. Derive an equation similar to Equation (20-4) which displays the existence of standing waves by beginning with the cosine representation of a wave

$$y = A \cos 2\pi \left(\frac{t}{T} - \frac{x}{\lambda}\right).$$

20-8. Show that Equation (20-3) satisfies an equation of the form

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

which is called the *wave equation.*

20-9. A standing wave is set up in a string. The distance between the nodes is 50 cm, and the measured width of the vibration at the antinodes is 10 cm. The standing wave pattern is set up by a tuning fork which vibrates at a frequency of 100 cycles/sec. (a) Rewrite Equation (20-4), substituting appropriate numerical values. (b) Find the width of the vibration pattern in the string at a distance of 20 cm from a node.

20-10. A transverse wave in a string is represented by the equation

$$x = 5 \cos (30t - 15y),$$

where all dimensions are in cgs units. Find (a) the direction of propagation of the wave, (b) the wavelength, (c) the frequency, and (d) the velocity.

20-11. A wind blows from the north at a speed of 50 ft/sec. In what direction should a beamed source of sound be pointed for the sound to travel due east?

20-12. Assume a ship to have vertical sides near the water line. (a) Find the natural frequency of vibration of a ship of cross-sectional area A and mass M when floating in water of density ρ. (b) When the ship is at anchor in deep water, what is the wavelength of ocean waves which will excite the ship to resonance?

20-13. A plane wave of length 25 cm, velocity 100 cm/sec, and having an amplitude of 5 cm, is propagated in the $-x$ direction. At time $t = 0$, a point at the origin of coordinates experiences its maximum positive displacement. (a) Find the displacement of a point whose coordinates are (0,5 cm) when $t = 5$ sec. (b) Find the displacement of a point whose coordinates are (+5 cm, −3 cm) when $t = 8$ sec.

20-14. A wave motion is described by the equation

$$y = 10 \sin\left(2t - 5x + \frac{5\pi}{6}\right).$$

(a) In which direction and along which coordinate axis is the wave moving? (b) Is the wave longitudinal or transverse? (c) What is the frequency? (d) The wavelength? (e) the velocity? (f) At $t = 0$ what is the displacement and the direction of motion of a point located at $x = 0$?

20-15. The motion of a vibrating string is given by

$$y = 3 \cos \frac{2\pi}{5} x \sin \pi t.$$

Find the velocity of the segment of the string located at $x = 2.5$ when $t = 0.5$ sec.

21

Vibrations and Sound

21-1 Sound

There are two aspects of sound: one is the physical aspect which involves the physics of the production, propagation, reception, and detection of sound; the other, which is the sensation of sound as perceived by the individual, depends upon physiological and psychological effects. It is not desirable to separate the two aspects of sound completely, but the main emphasis in this book must necessarily be on the physical aspect. In this chapter we shall consider mostly musical sounds. A vocabulary has been developed to describe the sensation experienced when a musical sound is heard. Such terms as the *pitch* of a sound, its *loudness*, and its *tone quality* or *timbre* are used to describe the musical sound. The physicist, on the other hand, speaks of the *frequency* of the sound, its *intensity*, and the *number* and *intensities* of the *overtones* present in a musical sound. Unfortunately, there is not a one-to-one correspondence between the terms used by the physicist and the terms used by the musician. A great deal of progress has been made in recent years as a result of tests involving thousands of persons which attempt to correlate the sensation of sound with the physical properties of sound. Some of these results will be mentioned at appropriate places in this chapter.

21-2 Frequency of a Musical Tone

A musical tone is regarded as a pleasing sound, while a noise is usually thought of as disagreeable; there are some sounds which are difficult to classify. A musical sound, for example, can be produced by a series of regular blasts of air, while a noise results when these blasts occur at irregular intervals. This can be demonstrated by means of a disk containing five concentric rings of circular holes, as shown in Figure 21-1. In the innermost ring these holes are irregularly spaced; the next four rings have circular holes which are regularly spaced. There are 40, 50, 60, and 80 holes in

these rings, respectively. When this disk is rotated at uniform angular speed and a stream of air is directed at the innermost ring of holes, an unpleasant noise will be heard. But when the stream of air is directed against any of the other rings, a pleasant musical tone will be heard. When the stream of air is directed against the four outer rings from the second to the fifth in sequence, the *pitch* of the sound coming from the third ring will be higher than that from the second ring of holes; the pitch of the sound from the fifth ring will be heard as an octave higher than that from the second ring. The physicist's method of describing these tones is in terms of the *frequency* of the sound produced. For example, if the disk is rotating at the rate of 10 rps, the frequency of the sound produced by the ring with 40 holes in it is 400 vib/sec; the next ring produces 500 vib/sec; the one after that, 600 vib/sec; and the last one, 800 vib/sec. To a first approximation we can say that the pitch of a tone depends upon its frequency, the tone with the higher pitch having the higher frequency. Two tones an *octave* apart have a frequency ratio of 2:1, for example. A musician will recognize these tones, which have frequencies in the ratios 4:5:6:8, as the tones comprising a *major chord.*

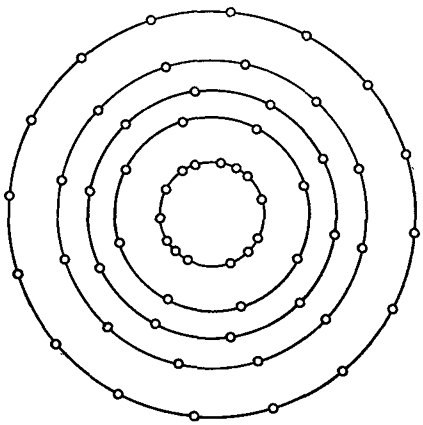


Fig. 21-1 Construction of a wheel for showing the difference between musical tones and a noise.

21-3 Resonance

An interesting phenomenon occurs when a body which is capable of vibrating at a definite frequency receives small impulses of the same frequency. These impulses set the body into vibration, with each succeeding impulse building up the amplitude of the vibration. This phenomenon, known as *resonance,* has been discussed in Section 12-7. A simple way of demonstrating resonance is to take two tuning forks having the same natural frequency and place them a short distance apart. One tuning fork is set vibrating by a hammer blow. After a short time interval, it will be found that the other tuning fork is vibrating and emits sound. The compressions and rarefactions produced in the air by the first tuning fork set the second tuning fork vibrating. Since the sound wave and the tuning fork have the same frequency, the impulses on the tuning fork are properly

timed to build up its amplitude of vibration. A steady state is reached when the energy radiated by the second tuning fork is equal to the energy it receives from the first one.

Resonance can occur between any two bodies which can vibrate with the same natural frequency. An interesting example is shown in Figure 21-2, which illustrates resonance between a tuning fork and an air column. A hollow cylindrical glass tube is inserted in a jar of water. The vibrating system is the air in the hollow tube; the length of the air column can be varied by moving the tube up or down in the water. The air column ends at the surface of the water. If a tuning fork vibrating at a known frequency is held over the open end of the hollow tube and the hollow tube is raised, there will be a marked increase in loudness of the sound at some position.. At this position, the air column in the tube is set into vibration with the same frequency as the tuning fork; the two are in resonance. We may think of the process of changing the length of the air column as "tuning" it to the frequency of the wave incident upon it. This may be compared to the tuning of a radio circuit to the same frequency as the incident electromagnetic wave.

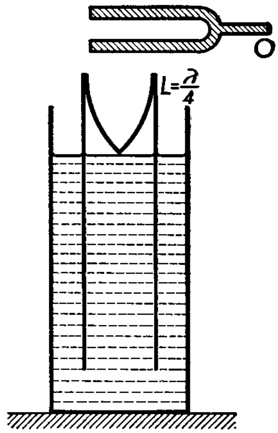


Fig. 21-2 Resonance between an air column and a tuning fork.

A tuning fork which vibrates with a frequency f emits a wave of length λ given by

$$V = f\lambda,$$

where V is the speed of sound in air. When an air column which is closed at one end is set into vibration, standing waves are produced in the air with a node at the closed end and an antinode at the other end. Since the distance between two successive nodes is half a wavelength, the distance between a node and the adjacent antinode is a quarter of a wavelength. Thus the shortest length of tube L in which the air can be in resonance with a wavelength λ is

$$L = \frac{\lambda}{4}. \tag{21-1}$$

If the tube is long enough, it will be found that resonance will occur again when the length of the tube is three quarters of a wavelength, for this length of air column will also have a node at the closed end and an antinode at the open end.

Tuning forks are often mounted on boxes whose air columns are in resonance with the sounds emitted by these forks. More energy is radiated per second from this system than from the tuning fork alone.

21-4 Beats

When two bodies having slightly different natural frequencies are set into vibration, the two waves emitted by them will *interfere* with each other. At some instant the two waves will be in the same phase, and there will be a reinforcement of the waves, resulting in a wave of increased amplitude. At

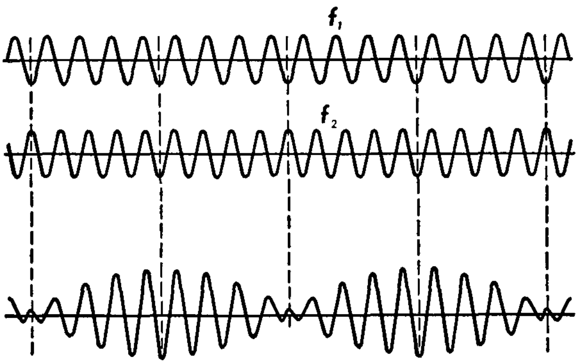


Fig. 21-3 Beats produced by the interference of two waves of slightly different frequencies.

some other instant the two waves will be completely out of phase; that is, a compression and a rarefaction will meet, resulting in a decreased amplitude producing a sound of very low intensity. The addition of two waves of slightly different frequency is illustrated in Figure 21-3. If one wave has a frequency f_1 and the other a frequency f_2, the number of times that these waves will get out of phase in a unit time can be shown to be $f_1 - f_2$; this is called the number of *beats* per unit time. If the two frequencies differ slightly, a series of beats will be heard; that is, the loudness of the sound will decrease noticeably $f_1 - f_2$ times in 1 sec. For example, if one tuning fork is emitting 256 vib/sec and another is emitting 260 vib/sec, 4 beats will be heard each second.

The phenomenon of beats is frequently used in tuning two sources of sound to the same pitch. This is a very accurate method of tuning, since the ear can perceive beats which occur only once in about 10 sec. When there are only a few beats per second, the sound produces an unpleasant effect. When the difference in frequencies is large, no beats can be distinguished in the sound produced.

21-5 Doppler Effect

In the previous discussions it was tacitly assumed that the source of sound and the observer were at rest with respect to each other. But when the source of sound is moving with respect to the observer, the pitch of the sound appears to be different from that when the two are stationary with respect to each other. There are two distinct cases to be considered: one in which the source is moving and the observer is at rest, and the other in which the observer is moving and the source is stationary. In both cases the frame of reference is fixed in the air.

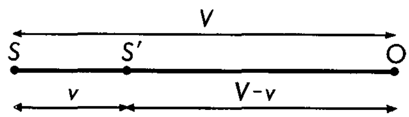


Fig. 21-4 The sound waves which are emitted in unit time by the source as it moves from S to S' with speed v toward the observer at O are contained in the distance $S'O$.

Suppose that an *observer is stationary at O*, as shown in Figure 21-4, and that the *source of sound is stationary at S*. If the source emits f vib/sec, the length of the wave λ emitted by it will be

$$\lambda = \frac{V}{f},$$

where V is the velocity of the sound. For simplicity, let us choose S to be at a distance from O equal to the distance traveled by sound in 1 sec; that is, $SO = V \times 1$ sec, where V is the speed of sound. Then when the source and the observer are both stationary, there will be f waves in the distance SO, each of length λ.

Let us now suppose that the *source is moving* with speed v *toward O*. At the end of 1 sec the source will have moved to S', where $SS' = v \times 1$ sec. During this time the source has emitted f vibrations; the first one has already reached the observer at O, and the last one has just left the source at S'. These f vibrations are therefore located in the region $S'O$, whose length is

$$S'O = (V - v) \times 1 \text{ sec}.$$

Since f waves have been emitted in this second, the length of these waves is

$$\lambda' = \frac{V - v}{f}. \tag{21-2}$$

These waves travel with the velocity of sound V, and the frequency f' with which they reach the ear is therefore

$$f'\lambda' = V. \tag{21-3}$$

Eliminating λ' from Equations (21-2) and (21-3) yields

$$\boxed{f' = f\frac{V}{V - v}\cdot} \tag{21-4}$$

In other words, more waves will now reach the ear per second than reach it when the source is stationary. This will be interpreted as a sound of higher pitch. The change in pitch produced by the relative motion of source and observer is known as the *Doppler effect.*

The same reasoning can be applied to show that when the *source is moving away from the observer* with a velocity v, the frequency f' of the sound reaching the observer is given by

$$\boxed{f' = \frac{V}{V + v}f.} \tag{21-5}$$

The pitch of the sound in this case is lower than the pitch of the sound when the source is stationary.

It is instructive to analyze the Doppler effect in terms of the waves emitted by the moving source. Let us assume that the source emits spherical waves which, in Figure 21-5, are drawn as circles with successive positions of the source as centers. These successive positions are shown at time intervals equal to T, the period of the vibrations emitted by the source. In the figure, S is the present position of the source, S_1 is the position of the source at a time T earlier than S, S_2 the position at a time $2T$ earlier, and S_3 the position at a time $3T$ earlier. The wave emitted when the source was at S_1 has traveled a distance VT, where V is the speed of sound; hence this wave is represented by a circle of radius VT. Similarly, the wave emitted when the source was at S_2 is drawn as a circle of radius $2VT$, that emitted from S_3 is drawn as a circle of radius $3VT$. The source of sound is moving toward the right with a speed v less than V.

An observer in front of the moving source will receive more waves per second than if the source had been at rest. Conversely, an observer behind the moving object will receive fewer waves per second than if the source had been at rest. The observer in front of the moving source will hear a higher pitched sound than the observer behind the source. When the moving source passes the observer, he will always note a drop in the pitch of the sound.

The frequency of the sound f' received by the observer in front of the source is given by Equation (21-4) and can be derived very simply by

referring to Figure 21-5. The distance between successive wave fronts which reach this observer is $(V - v)T$ and is therefore the length of the wave λ' perceived by the observer, that is,

$$\lambda' = (V - v)T,$$

but, for the source, $$T = \frac{1}{f}.$$

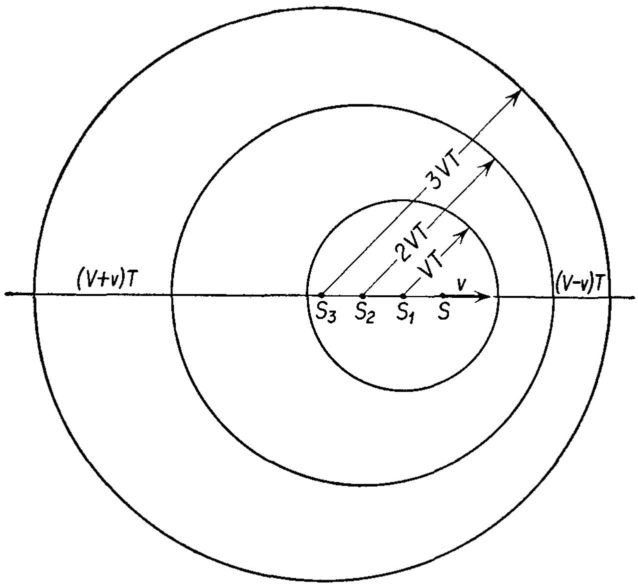


Fig. 21-5 Waves emitted by source moving to the right with speed v.

Therefore $$\lambda' = \frac{V - v}{f},$$

and since these waves travel with speed V in the air, the frequency f' of these waves is given by

$$f'\lambda' = V, \tag{21-3}$$

so that $$f' = f\frac{V}{V - v} \tag{21-4}$$

follows immediately.

Equation (21-5) can be derived in a similar manner.

Figure 21-5 can be used to determine the pitch of the sound heard by a stationary observer who is not in the line of motion of the source. The distance between wave fronts increases from the smallest value $(V - v)T$

for an observer in front of the source to its largest value $(V + v)T$ for an observer behind the source. To an observer at right angles to the line of motion, the distance between wave fronts is simply VT or λ; that is, it is exactly the same as if the source were stationary.

When the source is stationary and the *observer* is *moving* toward the source with a velocity v, the pitch is higher than that heard when the observer is stationary, but the actual value of the new frequency is slightly different from that given by Equation (21-4). The wavelength of the sound in air remains unchanged, but as the observer moves toward the source, he receives more waves per second than he receives when standing still. If he moves toward the source with a velocity v, he will receive v/λ additional waves per second, or a total of

$$f' = \frac{V}{\lambda} + \frac{v}{\lambda},$$

or

$$f' = \frac{V + v}{\lambda}.$$

Now since

$$f\lambda = V,$$

we get

$$\boxed{f' = \frac{V + v}{V} f,} \tag{21-6}$$

which gives the new frequency of the sound heard by the observer.

In a similar manner, if the observer is moving away from the source of sound, it can be shown that the frequency f' of the sound heard by the observer is given by

$$\boxed{f' = \frac{V - v}{V} f,} \tag{21-7}$$

which is a sound of lower pitch than that heard by the observer when stationary.

It must be emphasized that, in any one of these cases, the observer hears only one tone; he does not hear a change in pitch. Only when the motion is changed can he hear a change in pitch. Such a change in pitch can be observed when a train which is sounding its whistle passes an observer; the observer will hear a drop in pitch as the train passes him.

An interesting combination of the Doppler effect and the phenomenon of beats can be produced by moving a tuning fork rapidly toward a wall. A stationary observer will receive two sounds, one directly from the tuning fork and one reflected from the wall. The apparent source of sound of the

reflected wave is the image of the tuning fork formed by the wall acting as a plane mirror. While the tuning fork is moving away from the observer, its image is moving toward him. The direct wave from the tuning fork will have a lower pitch than the wave coming from its image, and the observer will hear beats.

Illustrative Example. The siren of a fire truck is emitting a tone whose frequency is 1,200 vib/sec. The fire truck is traveling with a speed of 60 mi/hr. A man in the street notices a drop in pitch as the truck passes him. Determine the change in frequency of the tone heard by this observer.

While the fire truck was moving toward the observer at a speed of 88 ft/sec, he heard a tone whose frequency was higher than 1,200 vib/sec. This frequency f_1' can be determined from Equation (21-4),

$$f_1' = f\frac{V}{V - v} = 1{,}200 \times \frac{1{,}100}{1{,}100 - 88}\,\frac{\text{vib}}{\text{sec}},$$

$$f_1' = 1{,}304\,\frac{\text{vib}}{\text{sec}}.$$

As the fire truck passed the observer, it moved away from him with a speed of 88 ft/sec, and the tone he heard had a frequency lower than 1,200 vib/sec. This frequency can be determined from Equation (21-5),

$$f_2' = f\frac{V}{V + v} = 1{,}200\,\frac{1{,}100}{1{,}100 + 88}\,\frac{\text{vib}}{\text{sec}},$$

$$f_2' = 1{,}111\,\frac{\text{vib}}{\text{sec}}.$$

Hence the drop in pitch of the tone heard by the observer was due to a change in frequency of

$$f_1' - f_2' = 193\,\frac{\text{vib}}{\text{sec}}.$$

21-6 Velocity of Source Greater than Velocity of Sound

When a body such as a projectile, a jet plane, or a rocket moves with a velocity v greater than the velocity of sound V in the medium, it sets up a compressional wave, as shown in Figure 21-6. The wave front, sometimes called *a shock wave*, is a cone, with the moving body at its apex S. The cone inside which the sound waves travel can be constructed by drawing spherical waves which originated at various positions of the source during its motion. In Figure 21-6 S is the present position of the source, S_1 its position at a time t earlier, S_2 its position at a time $2t$ earlier, and S_3 its position at a time $3t$ earlier, where t is an arbitrary time interval; let us call it one unit of time. With S_1 as a center we draw a circle of radius $V \times 1$, with S_2 as a center

we draw a circle of radius $V \times 2$, and so forth. These circles represent the present positions of the compressions which started from S_1, S_2, and S_3. The wave front is the tangent to these circles. In a three-dimensional diagram this wave front would be a cone whose elements were tangent to spheres

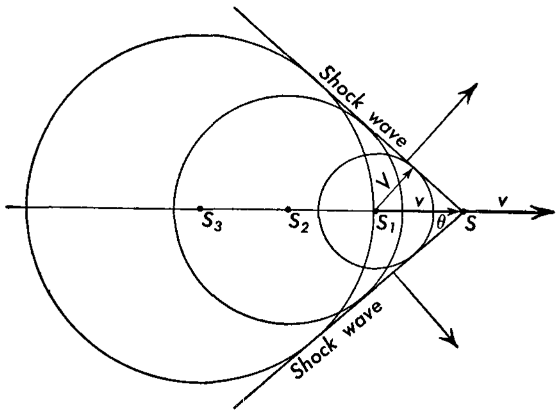


Fig. 21-6 Waves emitted by source moving with speed $v = 1.5\ V$.

of radii V, $2V$, and $3V$, respectively. In the unit of time that the wave progresses a distance V, the source moves a distance v, as shown in the figure. The angle θ that an element of the cone makes with the direction of motion of the source is

$$\boxed{\sin\theta = \frac{V}{v},} \qquad \textbf{(21-8)}$$

since V is at right angles to the element of the cone and v is the hypotenuse of the right triangle.

The cone of sound moves through the medium with the speed v of the source. Outside the cone, no sound can be heard.

When an airplane wing moves through the air, it produces a change in the pressure, or a pressure pulse, which travels through the air with the speed of sound. If the speed of the airplane is less than the speed of sound, the pressure pulse travels ahead of the wing and, in effect, sets up the flow pattern of the air ahead of it (see Figure 9-9). When the speed of the airplane is greater than the speed of sound through the air, the wing meets the air head on, producing a shock wave which travels across the wing. This shock wave increases the drag on the wing and also sets up great stresses in it. In aircraft engineering the ratio of the speed v of a plane to the speed V of sound in air through which it is traveling is called the *Mach number*.

Thus the Mach number chosen for Figure 21-6 is 1.5. The bow wave, or the wake from a speedboat, is a similar phenomenon.

21-7 Intensity and Loudness

The *intensity* of a wave at any point in space is defined as *the amount of energy passing perpendicularly through a unit area at this point in unit time.* The intensity can be expressed in ergs per cm^2 per sec or in watts/cm^2. The intensity of the sound received from any source depends upon the rate at which the source emits energy, upon the distance of the observer from the source, and upon the reflections which the waves undergo from the walls, ceiling, floor, and objects in the room. If the size of the source is small in comparison with its distance from the observer, and if no reflection or absorption takes place, the intensity of the sound at any place will vary inversely as the square of its distance from the source, but this is rarely the case with sound waves. In terms of the sound wave which reaches the observer, it can be shown that *the intensity depends upon the square of the amplitude of vibration of the particles in the wave and upon the square of its frequency.*

The *loudness* of a sound is a sensation experienced by the observer, and although loudness is related to the intensity of the sound, the relationship between the two is not a simple one. Waves in air may be detected by the normal human ear if their frequencies lie between about 20 cycles/sec and 20,000 cycles/sec and if their intensities are within a certain range; the range of intensities audible to the ear also depends on the frequency of the wave. Those waves which can be heard are called *sound waves.* Figure 21-7 shows the range of frequencies and their intensities which are perceived as sound by the normal human ear; the intensity of the wave is plotted along the y axis, while the frequency of the wave is plotted along the x axis. Because of the wide range of intensities, these are plotted not on a uniform scale but on a logarithmic scale. One scale shows the intensities in watts/cm^2. Another scale shows the intensities in terms of the pressure changes in the wave which strikes the eardrum; since the pressure in a wave varies sinusoidally, the *effective* or *root mean square* (abbreviated rms) values of the pressure changes are used. The lower curve represents the *threshold of audibility.* A point on this curve represents the smallest intensity of a sound of given frequency which is just audible to the average ear. The ear is most sensitive to sounds of about 3,000 cycles/sec. At a certain intensity, known as the *threshold of feeling,* the sound is not heard but is felt by the ear as a tickling sensation. Above the threshold of feeling the intensity may be so great as to be painful. The region between the two curves represents the range of hearing. The range of intensities to which the ear is sensitive is about a millionfold. Because of this large

range of intensities, a *logarithmic scale* has been adopted for expressing the *level of intensities of sound*, taking the zero level at about the limit of audibility of sound. The intensity level B of a sound is defined as

$$\boxed{B = 10 \log \frac{I}{I_0},} \tag{21-9}$$

where I is the intensity of the sound and I_0 is the zero level of intensity which is taken arbitrarily to be equal to 10^{-16} watt/cm^2 or 10^{-12} watt/m^2.

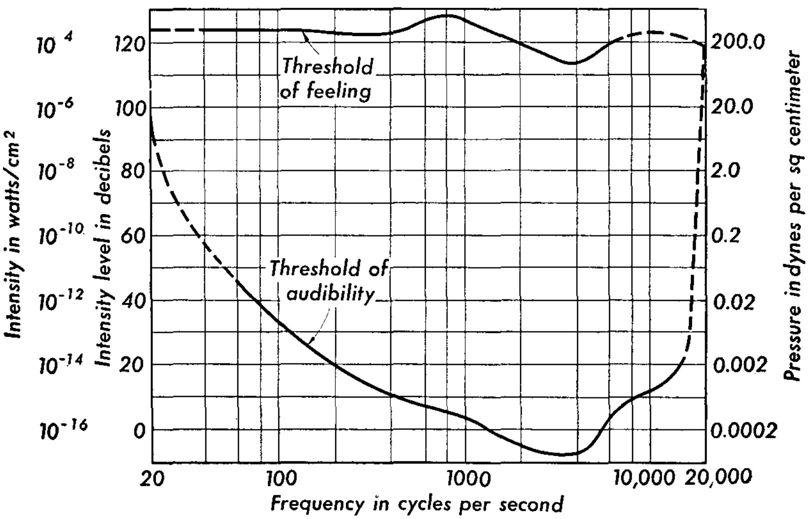


Fig. 21-7 Range of frequencies and their intensities which are perceived by the human ear. (After H. Fletcher, *Reviews of Modern Physics*, January, 1940.)

The intensity level B is expressed in *decibels* (db). Thus if a sound has an intensity $I = 10^{14}$ watt/cm^2, its intensity level is

$$B = 10 \log \frac{10^{-14}}{10^{-16}} \text{ db},$$

or

$$B = 10 \log 100 \text{ db},$$

from which

$$B = 20 \text{ db}.$$

Sound levels have been measured at various places under a variety of conditions. For example, inside some noisy subway cars the sound level is about 100 db, while the threshold of feeling (or pain) is about 120 db; the sound level of a whisper is about 15 db.

The shape of Figure 21-7 is of interest in connection with the high-fidelity reproduction of music. It is common experience that a radio or phonograph sounds better when played at high volume, for at low volume a significant range of frequencies is reproduced below the threshold of audibility. It is unreasonable to expect reproduced sound to have the fullness of orchestral music unless the intensity of sound in the home is equal to that of the concert hall. In modern high-fidelity phonographs an attempt has been made to compensate for the response of the ear by introducing a contoured volume control which decreases the intensity at both high and low frequencies at a lesser rate than the middle frequencies, as the volume control is turned down.

21-8 The Ear

Figure 21-8 is a diagram showing the essentials of the structure of the human ear. Sound waves enter the ear through the auditory canal and strike the eardrum. The pressure variations of the sound wave are trans-

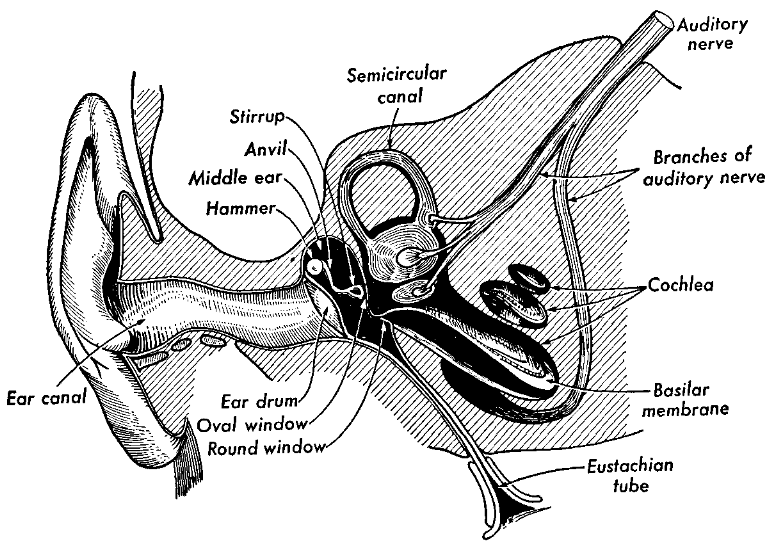


Fig. 21-8 Semidiagrammatic section of the right ear.

mitted from the eardrum by a system of three small bones, the hammer, anvil, and stirrup, to the oval window of the inner ear. The latter is filled with a liquid. The cochlea, a spiral-shaped part of the inner ear, contains the *basilar membrane* which runs along its entire length and divides it into two sections. The nerve endings of the auditory nerve are attached to one edge of the membrane. The entire length of the basilar membrane is about

30 mm (about 1.2 in.), and there are about 30,000 nerve endings attached to it. The vibrations which are transmitted to the liquid of the inner ear set the basilar membrane into vibrations, different tones affecting different sections of the membrane. These vibrations stimulate the nerve endings attached to it, and these transmit the signal along the auditory nerve to the brain.

The sense of balance is associated with the semicircular canal, and it is interesting that there are three of these natural "carpenter" levels in each ear, approximately at right angles to each other. In this way nature has apparently recognized the independence of the three rectangular components of a force, or the three mutually perpendicular directions of space.

Another aspect of the design of the ear which is of interest is the length of the ear canal. If we consider the ear canal as a pipe, closed at one end, we find that its resonant wavelength is given by Equation (21-1) as

$$\lambda = 4L. \tag{21-1}$$

The length of the ear canal is about 3.3 cm. Taking the velocity of sound as approximately 330 m/sec, we find the resonant frequency of the ear canal as

$$\begin{aligned} f &= \frac{V}{\lambda} \\ &= \frac{3.3 \times 10^4 \text{ cm/sec}}{4 \times 3.3 \text{ cm}} \\ &= 2{,}500 \text{ cycles/sec,} \end{aligned}$$

in good agreement with Figure 21-7.

If we assume that the hearing mechanism of most animals is similar to that of man, that the length of the ear canal is roughly proportional to the size of the animal, and that animals speak at frequencies appropriate to their organs of hearing, we may infer that large animals will speak in bass voices while smaller animals will utter sounds of higher frequencies.

21-9 Quality of a Musical Sound

When two tones of the same pitch and same loudness are produced by two different musical instruments, such as a violin and a clarinet, the sensations produced by them are decidedly different. We recognize this difference because of the difference in *quality* or *timbre* between these two musical sounds. One of the main reasons for this difference in quality is that each sound produced by an instrument is not a tone of a single frequency but is a complex sound consisting of several different frequencies. Another reason

for the difference in quality is the manner in which the human ear responds to tones of different frequencies and different loudness. In this section we shall consider only the effect produced by the complexity of the sound on the quality of a tone emitted by vibrating bodies.

As shown previously, the vibrations in a body can be analyzed in terms of stationary waves which are set up in it by the interference of two waves traveling in opposite directions. The *fundamental* mode of vibration of a body corresponds to the longest wave, or wave of lowest frequency, which

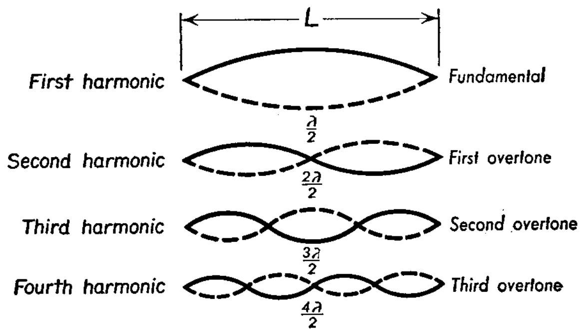


Fig. 21-9 Modes of vibration of a string which is fastened at both ends.

can be set up in the body. For example, in the case of a string which is fastened at both ends, the fundamental mode of vibration is one in which the string vibrates as a whole, as shown in Figure 21-9. Since the two ends are fastened, the standing wave must have nodes at these points, while the antinode is midway between them. The length of the string L is thus half the wavelength $\lambda/2$ of the transverse wave traveling in the string. The pitch of the sound emitted by the string when vibrating as a whole is called the *fundamental tone* of the string. The string can also vibrate in two parts, with a node in the center; the frequency of this sound is twice that of the fundamental tone. When vibrating in this manner, the pitch of the tone will be an octave higher than the fundamental. Other modes of vibration of the string are shown in the figure. When a string is set into vibration by plucking it or bowing it, several of these modes of vibration will be set up simultaneously; in addition to the fundamental tone, some of the higher pitched tones or *overtones* will be emitted by the string. *The quality of the tone will depend upon the number of overtones produced and their relative intensities.*

The same thing is true of other vibrating bodies; the quality of the sound depends upon the number and relative intensities of the overtones produced. Although, in the case of string instruments and wind instruments, the frequencies of the overtones are whole multiples of the frequency

of the fundamental tone, this is not generally true of other musical instruments such as bells, chimes, and drums.

Another effect which helps to determine the quality of a musical tone is its transient character. A note sounded by a musical instrument is rarely sustained at constant volume for more than a moment or two. In a percussion instrument such as the piano, the intensity of the note diminishes rapidly after the note is struck. It is interesting to listen to the recording of a single piano note played backward. With modern tape-recording techniques it is possible to play a simple tune on the piano, to snip out each note, and then to paste them together so that each note is played backward but in the correct order. Each note increases in intensity and then abruptly ends. The effect is quite unrecognizable as a piano piece.

It is perhaps the transient nature of the typical musical instrument's sound which makes many electronic musical instruments seem rather unsatisfactory and lifeless, for, while an electronic instrument may generate the correct frequencies and even the correct distribution of harmonics, it is difficult to simulate the correct transient response as well.

21-10 Vibrations of Strings

The frequencies of the various modes of vibration of a string are in the ratios of whole numbers; such vibrations form a *harmonic series* with the fundamental vibration as the first harmonic, the overtone of twice this frequency as the second harmonic, and so forth. Any wavelength λ and its associated frequency f of a wave in a vibrating string are related by the equation

$$V = f\lambda,$$

where V is the speed of the transverse wave in the string. This speed is given by

$$V = \sqrt{\frac{S}{m}}, \tag{20-10}$$

where m is the linear density of the string, that is, its mass per unit length, and S is the tension in the string. Now the length of the wave in the fundamental mode of vibration is twice the length L of the string, that is,

$$\lambda = 2L; \tag{21-10}$$

therefore the frequency f_1 of the fundamental or first harmonic is

$$f_1 = \frac{V}{2L}, \tag{21-11}$$

from which

$$f_1 = \frac{1}{2L}\sqrt{\frac{S}{m}}. \tag{21-12}$$

The frequency of vibration and hence the pitch of a string can be varied by changing its length or its tension. Increasing the tension four times will double the frequency of vibration or produce a tone an octave higher. Decreasing the length of the string—for example, by pressing the string against a board with the finger—will increase the frequency or pitch of the tone emitted. Two strings of the same length and under the same tension, but of different linear densities, will have different frequencies for their fundamental tones.

The number of overtones set up in a string can be controlled to some extent by the method of bowing or plucking the string. For example, if the string is plucked in the center, those modes of vibration will be most intense which have antinodes at this point; the overtones will consist of the odd harmonics. If the string is plucked at a distance of about one seventh of its length from one end, the seventh harmonic will be absent, but most of the even harmonics will be present; the quality of this tone will be noticeably different from that heard when the string is plucked at the center, even though in each case the pitch heard is that of the fundamental tone.

The presence of overtones in a vibrating string can be shown very easily by plucking the string at one quarter of its length from one end and then placing a finger lightly on the center of the string; this will stop the fundamental mode of vibration, but the second harmonic can continue to vibrate, since it has a node at the center. When the finger is placed at the center, the fundamental tone will no longer be heard, but its octave will be heard, showing that the second harmonic is present.

Another method for demonstrating the presence of overtones in a vibrating string is illustrated in Figure 21-10. A steel wire is stretched between two posts A and B on a board, and these two points are connected to the primary coil of a step-up transformer. The secondary coil of this transformer is connected to an amplifier and a loud-speaker. Several U-shaped magnets are placed on the board so that the wire can vibrate freely between the poles of these magnets. Suppose that these magnets are so placed that their north poles are all on one side of the wire and their south poles are on the opposite side, as shown in Figure 21-10(a). When the string is plucked near the center, an induced electromotive force (emf) will be set up in the wire. This will be amplified, and the fundamental tone will be heard coming from the speaker.

To show the presence of the second harmonic, we use only two magnets, each placed at a distance of about one quarter of the length of the string from the ends A and B, but with opposite poles on the same side, as shown in Figure 21-10(b). The magnets are placed in this position because the string vibrates in two parts, with the portions of the string on either side of the central node moving in opposite directions. If the string is now plucked at a point near one of the magnets, a tone twice the frequency of the funda-

mental will be heard. If, while the string is still vibrating, one of the magnets is reversed so that like poles are on the same side, the fundamental tone will be heard. If the magnet is again reversed, the octave will be heard. This clearly demonstrates that both the fundamental tone and the first overtone are present at the same time.

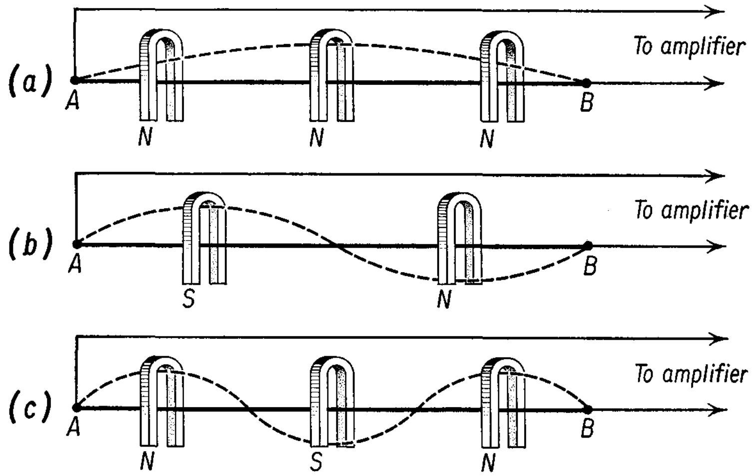


Fig. 21-10 Method for demonstrating the presence of overtones in a vibrating wire. The ends of the string AB are connected to the primary of a transformer whose secondary coil is connected to an amplifier and loud-speaker. N and S represent those poles of the magnets which are in front of the wire.

To show the presence of the second overtone, or third harmonic, three magnets are used, as shown in Figure 21-10(c). If the string is plucked at a point near one of the end magnets, the third harmonic will be heard clearly. If, while the string is still vibrating, the center magnet is reversed, the fundamental tone will be heard, thus showing that the first and third harmonics are both present. This method can be extended to show the presence of several of the other harmonics.

The amount of energy transferred directly from the vibrating string to the air is very small. To permit a greater transfer of energy, the strings are mounted on various types of solid boards, called *sounding boards,* usually made of metal or wood. These sounding boards are set into *forced* vibration by the vibrating string; the vibrations of these boards set larger quantities of air into motion, thus producing a more intense sound. In some string instruments of the violin type, there are air columns in the instrument which also vibrate. All of these vibrating systems make their contribution to the quality of the sound emitted by the instrument.

Illustrative Example. One of the steel strings of a piano is 50 cm long and has a linear density of 0.60 gm/cm. When struck with the hammer, it emits a

tone whose fundamental frequency is 520 vib/sec. Determine (a) the tension in the string and (b) the frequencies of the first and second overtones of this string.

The tension in the string can be determined from Equation (21-12). Solving this equation for S, we get

$$S = 4L^2f^2m.$$

Substituting the numerical values, we get

$$\begin{aligned} S &= 4 \times 2{,}500 \times 270{,}000 \times 0.60 \text{ dyne} \\ &= 16.2 \times 10^8 \text{ dynes.} \end{aligned}$$

The first overtone of a string is its second harmonic, and the second overtone is its third harmonic, therefore their frequencies are

$$f_2 = 1{,}040 \text{ vib/sec},$$

and

$$f_3 = 1{,}560 \text{ vib/sec}.$$

21-11 Vibrating Air Columns

Wind instruments such as the clarinet, the trumpet, and the pipe organ all have vibrating air columns to reinforce some of the sounds produced by the source of sound. In this discussion we shall consider only cylindrical pipes such as those commonly used in pipe organs. There are two general classes of such pipes, the *open* pipe, that is, a pipe open at both ends; and a *closed* pipe, that is, a pipe closed at one end; the end containing the source of vibrations is always considered an open end. The vibrations can be produced in one of several ways, such as blowing air against a reed and setting it vibrating, or blowing a thin sheet of air against a thin lip at one end and setting the air into vibration. Whatever the method of setting up the vibrations, the column of air in the pipe will reinforce those modes of vibration corresponding to the standing waves which can be set up in this column.

Figure 21-11(a) shows several modes of vibration which can be set up in a closed organ pipe. The method for determining these modes of vibration depends upon the fact that only those vibrations can exist in the air column which have a node at the closed end and an antinode at the open end. It can be seen that the fundamental mode of vibration corresponds to a wavelength λ which is four times the length L of the air column. The first overtone possible in this case is one whose wavelength λ is four thirds of the length L of the air column. The frequency of the first overtone is therefore three times the frequency of the fundamental tone. The frequency and the wavelength are related by the usual equation

$$V = f\lambda,$$

where V is the speed of sound in air.

An examination of all other possible modes of vibration of an air column in a closed pipe shows that all the overtones are odd harmonics; their frequencies are in the ratios of 1:3:5:7:... and so on.

A pipe open at both ends must have antinodes at these ends. The fundamental mode of vibration of the air in an open pipe is shown in Figure 21-11(b); its wavelength λ is twice the length of the pipe. The first overtone has a wavelength λ which is equal to the length L of the pipe; its frequency is therefore twice that of the fundamental. An examination of the other possible modes of vibration shows that all the harmonics may be set up in this air column.

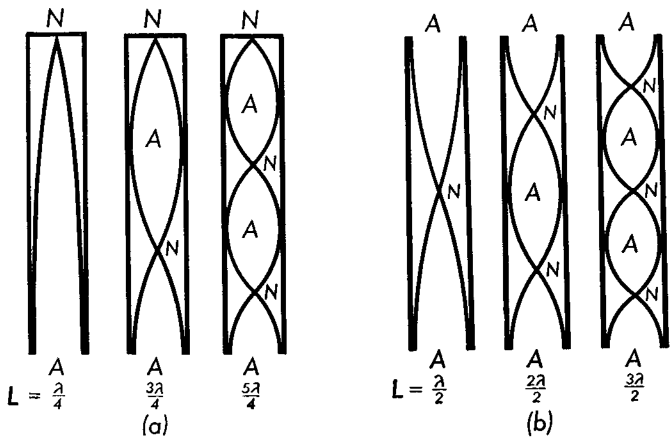


Fig. 21-11 (a) Modes of vibration of an air column closed at one end. (b) Modes of vibration of an air column open at both ends. The nodes are at the positions marked N, and the antinodes are at the positions marked A.

In practice, small corrections need to be applied to the actual length of a pipe to yield an effective length. We shall ignore these corrections in this and subsequent discussions.

A simple comparison of the open and closed pipes will show that if two such pipes of the same length are emitting their fundamental tones, the open pipe will emit a tone an octave higher than the closed pipe.

The pitch of a sound produced by an air column can be varied without varying the length of the air column. For example, when the air is blown harder against the reed, one of the overtones produced in the air column may have a greater intensity than the fundamental, and the pitch of the sound will correspond to this overtone. A bugle, for example, has a fixed length, but different tones can be obtained from it by changing the tension of the lips and the manner of blowing the air through them.

21-12 Acoustical Problems

From our discussion of the natural frequency of organ pipes and of the phenomenon of resonance, we see that any open pipe has a natural frequency of vibration, and that such a pipe will be excited to large amplitudes only when it is excited at its natural frequency. In many engineering problems it is necessary to maintain a free flow of air, while at the same time it is desired to prevent the passage of sound. The air-intake section of an aircraft wind tunnel is a typical illustration. If the propellers which drive the air of the wind tunnel are maintained at a fixed rotational speed, it is possible to admit air to the wind tunnel through a nest of tubes whose dimensions are chosen so that their natural frequencies are far from the frequency of the sound generated by the propellers. Such a nest of tubes constitutes an *acoustic filter* which discriminates against sound waves of certain frequencies. Other examples of sound filters in common use are the mufflers of automobiles and guns. The familiar effect produced by shouting into a barrel is an example of acoustic filtering.

Let us consider the effect of the size of a room on the sound produced within the room by a musical instrument. As a first approximation we may consider the room to be broken up into an imaginary nest of tubes whose length is the longest dimension of the room. Since these tubes are closed at both ends, the resonant wavelength of each tube will be

$$\lambda = 2L,$$

where L is the length of the room. A typical living room about 22 ft long will have a resonant frequency of about 50 cycles/sec. It should not be expected that a recorded organ recital will be reproduced with the same tonal balance in such a living room as it was originally performed in a cathedral.

Another aspect of auditorium acoustics which is of great importance is called the *reverberation time*, which is arbitrarily defined as the time required for the sound intensity to fall to one millionth of its original intensity. The persistence of sound in a room depends upon the absorption of sound at the walls at each reflection. In a room covered with heavy draperies, the reverberation time may be too short, and the sound seems dull. In general, the desirable reverberation time ranges from about 1 sec for a large room to about 2 sec for a large auditorium. An empty, untreated auditorium may have a reverberation time as long as 6 or 7 sec. If the reverberation time is too long, sounds reaching the listener from a speaker by a direct path and by reflected paths are of comparable intensity, and a meaningless garble results.

21-13 Ultrasonics and Supersonics

The term *supersonic* is often used to designate speeds greater than the velocity of sound. The so-called "sound barrier" to aircraft flight has nothing basically to do with sound but rather is associated with the velocity of propagation of a disturbance in air, which we call the velocity of sound. When an aircraft moves through air at speeds higher than the velocity of sound, shock waves, which are regions of considerable turbulence, are developed. The flow is no longer streamline, and control surfaces on the airplane no longer respond in a normal fashion.

The term *ultrasonic* is used to describe sound waves of frequencies greater than the human ear can hear, generally greater than about 20,000 cycles/sec. Such frequencies may be produced by a vibrating quartz crystal plate of appropriate size and shape. The plate is excited electrically through the *piezoelectric effect*, in which a dimensional change in the quartz is produced by electrical excitation. Ultrasonic frequencies may also be generated by the vibrations of a rod of magnetic material which is excited magnetically by *magnetostriction*, in which a change of dimension is produced magnetically.

Sound waves of ultrasonic frequencies have been found to disperse colloids and to produce various biological effects, such as the destruction of microorganisms, and have been used to sterilize and homogenize milk. Because of their short wavelength, such waves tend to move more nearly in a straight-line path and may cast sound shadows and be reflected from small obstacles. Thus ultrasonic waves have been used to detect flaws of suitable dimensions in opaque objects. Sound is reflected from cracks and other discontinuities, and their presence may be readily detected.

By listening for the echo, bats determine the presence of obstacles and insects by the reflection of ultrasonic sound which they emit. In order to obtain good reflections, the wavelength of the sound must not be appreciably greater than the dimensions of the obstacle. Bats feed upon beetles and moths and can sense the presence of ropes strung in a cave. They do not detect gnats and smaller insects and will fly into fine wires strung across a cave, for the dimensions of these objects are appreciably smaller than the wavelength of the ultrasonic wave the bat emits. The highest frequency a bat can utter is about 100,000 cycles per second, corresponding to a wavelength of about 0.13 in.

Problems

[NOTE: The speed of sound in air at room temperature may be taken as 1,100 ft/sec or as 330 m/sec.]

21-1. A siren wheel has 20 uniformly spaced holes near its rim and is rotated

by means of a stream of air. (a) What is the frequency of the sound emitted when its speed is 44 rps? (b) What is the wavelength of the sound wave in air?

21-2. An air column closed at one end is in resonance with a sound wave whose frequency is 128 vib/sec. Determine the length of this air column.

21-3. What is the lowest-frequency note that can be produced by an organ pipe 12 ft. long which is closed at one end?

21-4. A metal tube 4 ft long has a piston placed in it near one end; the position of this piston is adjustable. A vibrating tuning fork whose frequency of vibration is 440 vib/sec is held near the open end of the tube. At what distances from the open end must the piston be placed to produce resonance?

21-5. A siren on a fire-engine truck emits a sound whose frequency is 1,000 vib/sec. What will be the frequency of the tone heard by a spectator (a) when the truck is moving toward him with a speed of 45 mi/hr and (b) when it is moving away from him with a speed of 45 mi/hr?

21-6. A fire siren in a village is emitting a sound whose frequency is 880 vib/sec. What is the frequency of the tone heard by the firemen approaching this source at a speed of 60 mi/hr?

21-7. A loud-speaker is hung from a cord attached to the ceiling of a lecture room and is connected by flexible wire to a 1,000-cycle oscillator. While emitting this note, the loud-speaker is swung toward the front wall with a speed of 3 ft/sec. (a) What is the frequency of the tone coming from the loud-speaker as heard by a student sitting in the rear of the room? (b) What is the frequency of the tone coming from the image of this loud-speaker as heard by this student? (c) How many beats per second will this student hear?

21-8. A string 100 cm long has a linear density of 0.04 gm/cm. When vibrating transversely with a node at each end, it has a frequency of 200 vib/sec. Determine (a) the speed of the wave in the string and (b) the tension in the string.

21-9. Two tuning forks are vibrating simultaneously, one with a frequency of 512 vib/sec and the other with a frequency of 516 vib/sec. How many beats are produced?

21-10. Two tuning forks A and B are observed to produce beats at the rate of 5 per second. Fork A has a frequency of 440 vib/sec. If fork B is loaded with a bit of putty, the number of beats increases to 8 per second. What is the frequency of vibration of fork B when it is not loaded?

21-11. An open organ pipe sounds the note A whose frequency is 440 vib/sec. (a) What is the length of the air column? (b) What are the frequencies of the first and second overtones produced by this air column?

21-12. A closed organ pipe sounds the note A whose frequency is 440 vib/sec. (a) What is the length of the air column? (b) What are the frequencies of the first and second overtones produced by this air column?

21-13. A steel wire 80 cm long is fastened at the ends. The mass of the wire is 2.5 gm. When plucked, it emits a tone whose fundamental frequency is 520 vib/sec. (a) Determine the tension in the wire. (b) What are the frequencies of the first and second overtones of this string?

21-14. A steel wire 3 ft long is rigidly fastened to two posts. The wire weighs 0.02 lb. When bowed, it emits a tone whose fundamental frequency is 110 vib/sec. (a) Determine the tension in the wire. (b) What are the frequencies of the first and second overtones produced by this string?

21-15. A bullet is fired with a velocity of 2,700 ft/sec. Determine (a) its Mach number and (b) the angle that the shock wave makes with the line of motion of the bullet.

21-16. A siren mounted on a car emits a note whose frequency is 500 vib/sec. (a) Determine the frequency of the sound heard by a stationary observer when the car approaches him with a speed of 70 ft/sec. (b) Determine the frequency heard by an observer moving toward the car with a speed of 70 ft/sec while the car remains stationary.

21-17. Assuming that the ear can detect sounds whose frequencies range from 20 vib/sec to 20,000 vib/sec, determine the range of wavelengths that the ear can detect.

Part Four

ELECTRICITY AND MAGNETISM

22

Electrostatics

22-1 Introduction

One simple phenomenon of electricity was known to the ancients: that when a piece of amber was rubbed, it acquired the property of attracting small pieces of paper and other light particles. Records show that Thales of Miletus (circa sixth century B.C.) knew of this property of amber; the Greek word for amber is *elektron*, hence the name *electricity*. There was practically no further development of this subject until about the seventeenth century. Otto von Guericke (1602–1686) of Magdeburg built a large sulphur sphere, which, when rotated about an axis and rubbed with his hand, gave off electric sparks. In the eighteenth century it was found that there were two kinds of electricity, one similar to the kind acquired by amber when rubbed with wool, and the other similar to that acquired by glass when rubbed with silk, called *resinous* and *vitreous* electricity, respectively. They are now known as *negative* and *positive* electricity, names first introduced by Benjamin Franklin (1706–1790). Franklin made many important contributions to the subject, experimentally and philosophically. He showed the electrical character of lightning and designed lightning rods for the protection of buildings. The subject of electricity was put on a firm mathematical foundation as a result of the experiments of Coulomb (1785) on the law of force between electrically charged bodies.

22-2 Electrified or Charged Bodies

Any substance, when rubbed, becomes *electrified* or *charged electrically.* If a glass rod is rubbed with a piece of silk, both the rod and the silk become electrified or charged. It is now known that the purpose of rubbing two substances together is to bring parts of their surfaces into sufficiently close contact so that electrically charged particles, called *electrons*, can be trans-

ferred from one body to the other. The process of charging bodies by rubbing them together used to be called frictional electrification; it is now more properly called *electrification by contact* or *charging by contact.*

In order to study the behavior of electrically charged bodies, let us suspend a glass rod A by means of a string and then charge the rod electrically by rubbing it with a piece of silk. Suppose we take another glass rod B, which has been rubbed with a piece of silk, and bring it near the first rod A, as shown in Figure 22-1. We shall find that there is a force of *repulsion* between them. If we perform a similar experiment with two rubber rods, each of which has been rubbed with a piece of fur or wool, we shall find that there is a force of repulsion between the two charged rubber rods. If we now take one of the charged rubber rods and bring it near the charged glass rod A, we shall find that there is a force of *attraction* between them. From this set of simple experiments we can conclude, first, that *there are two kinds of electric charges;* and second, that *like charges repel each other and unlike charges attract each other.*

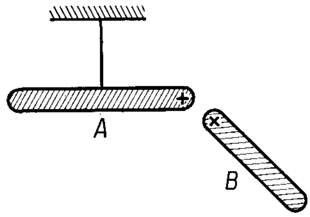


Fig. 22-1 Two glass rods, charged in the same manner, repel each other.

We can use the charged glass rod and the charged rubber rod as our standards and compare all other charged bodies with them. We shall find that there are only two kinds of charges: those which are similar to the charge on the glass rod and are repelled by it, and those which are similar to the charge on the rubber rod and are repelled by it. To distinguish between these two kinds of charges, the charge carried by the glass rod which has been rubbed with silk is arbitrarily called a *positive charge,* and the charge on the rubber rod which has been rubbed with wool or fur, a *negative charge.* The ultimate test of the nature of the charge on any body is to bring this charge near a properly charged glass rod or rubber rod. If the charged body is repelled by the positively charged glass rod, its charge is positive. If it is not repelled by the charged glass rod, it may be brought near the negatively charged rubber rod; if it is repelled by the charged rubber rod, the body carries a negative charge.

22-3 Insulators and Conductors

We have seen that all bodies can be charged electrically by being rubbed. But if a metal sphere is held in the hand and rubbed with a piece of fur and then tested, it will be found to be uncharged. However, if the metal sphere is mounted on a hard-rubber stand or a glass stand and then rubbed

with a piece of fur, it will be found to be charged. The reason for this behavior is that metals are *good conductors* of electricity while rubber and glass are *nonconductors* or *insulators*. The human body is also a conductor of electricity, and so is the ground. When the metal sphere was held in the hand, the charge produced on it by rubbing was conducted away by the hand through the body to the ground. But when the sphere was placed on an insulator, the charge produced by rubbing remained on the sphere.

Materials can generally be classified as either conductors or insulators with some degree of accuracy, although there are many borderline cases, including a class of materials called semiconductors which have recently assumed considerable importance in electronics. The quantitative measurement of the conductivity of substances will be considered in a later chapter. For the present, it will be sufficient to give a qualitative discussion of conductivity. All metals in the solid state have been found to be good conductors of electricity and are also fairly good conductors in the liquid state. Some nonmetals, such as carbon and selenium, are also good conductors. Solids such as glass, rubber, amber, and most plastic materials are good insulators. All gases, under ordinary conditions, are poor conductors of electricity. Certain liquid solutions of materials, which in chemistry are known as acids, bases, and salts, are conductors of electricity, while oils, liquid solutions of organic substances such as sugar, and pure water are nonconductors.

22-4 The Electrical Theory of Matter

During the past half century a great deal of information has been obtained concerning the structure of matter and its behavior under the action of various external forces. It will be worth while at this point to outline briefly the electrical theory of matter to aid us in understanding the phenomena to be discussed later. The fact that all bodies can be charged electrically suggests that electric charges are common to all matter. Since, under normal conditions, a body is electrically *neutral*, that is, uncharged, such a body must contain equal amounts of positive and negative electricity so arranged that no electrical effects can be observed outside the body. From investigations of the structure of the atom, it is now known that each atom consists of a *small positively charged massive nucleus made up of positively charged protons and uncharged neutrons*, surrounded by groups of small negatively charged particles called *electrons*. The mass of an electron is very small in comparison with that of the nucleus, it is about 1/1,840th mass of a single proton. The force of attraction between the nucleus and any electron in the atom depends upon the position of the electron with respect to the nucleus; the force on an outer electron is much smaller than that on an inner electron. The charges which can be removed from the

atom most readily are the outermost negative electrons. We may think of the process of rubbing two bodies together as bringing electrical forces into play during the time of contact between the two surfaces. These forces produce a displacement of the negative electrons from one body to the other. The body from which electrons have been removed becomes positively charged, while the one to which electrons have been added becomes negatively charged.

When atoms are brought close together, as they are in liquids and solids, the electric charges exert forces on each other, causing some rearrangement among the charges, particularly among the outermost electrons. In the case of metals this rearrangement is such that some of the outermost electrons in each atom can move freely from atom to atom. These are sometimes referred to as the *free electrons* of the metal. In the case of insulators the electrons are tightly bound to the atoms and can be moved from atom to atom only by the application of a comparatively large force.

22-5 Charging by Induction

If a negatively charged rubber rod is brought near an insulated metal sphere A, as shown in Figure 22-2, the electrons in the metal will be repelled; the farther side of the sphere will become negatively charged, and the side near the rubber rod will become positively charged. If the experimenter

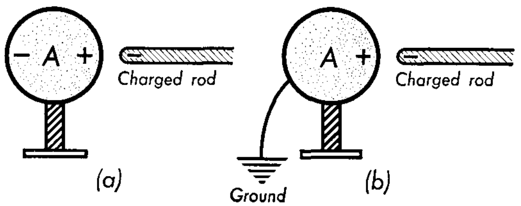


Fig. 22-2 Method of charging a metal sphere by induction.

touches the sphere with his hand for a short time, some electrons will be repelled through him into the ground, thus leaving the sphere with a net positive charge when the hand and rod are removed successively. This method of charging a body by bringing it close to a charged body, without making actual contact, is called *charging by induction.* The fact that the sphere is positively charged can be verified by bringing it near a positively charged glass rod and noting the force of repulsion between them. Grounding of the sphere can also be done by connecting a wire directly from the ground to the sphere. If the insulated positively charged sphere is removed from the neighborhood of other charged bodies and then again connected to ground, electrons will be attracted up from the ground until the sphere

is neutral. In this case we say that the sphere has been *discharged* by connecting it to ground.

To charge a metal sphere negatively by induction, we can bring a positively charged body near the sphere. Some of the electrons will be attracted toward the positively charged body and will go to the surface nearest it, leaving the rest of the surface positively charged. If the sphere is momentarily connected to ground, electrons will be attracted up from the ground during the time of contact; the sphere is now negatively charged. If the original positively charged body is now removed, the excess electrons will redistribute themselves uniformly over the surface of the sphere because of the forces of repulsion between them.

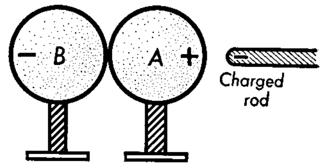


Fig. 22-3 Charging two metal spheres by induction.

An interesting variation of this experiment which does not involve the use of a ground connection is illustrated in Figure 22-3. Two metal spheres A and B, each on an insulated stand, are brought together so that their surfaces touch each other. If a negatively charged body is now brought near A, electrons will be repelled from it to B. If the spheres are now separated, A will be positively charged and B will be negatively charged. Since the total charge on the two bodies was originally zero, the negative charge on B must be numerically equal to the positive charge on A.

22-6 The Electroscope

The electroscope is a simple and sensitive electrical instrument which can be put to a variety of uses. It consists essentially of a metal rod to which two pieces of very thin, light, gold or aluminum foil are attached. The rod is passed into the case through an insulating bushing. The two pieces of foil, or *leaves*, as they are usually called, are enclosed in a container made either of glass or of metal, provided with glass windows, as shown in Figure 22-4. Part of the metal rod projects outside the case so that contact can be made with it. The rod may be provided with any kind of top such as a metal sphere or a flat metal plate. The electroscope can be charged by contact between the rod AB and a charged body. The charge distributes itself over the top, the rod, and the leaves. Since the two leaves acquire like charges, they repel each other and diverge until equilibrium is established. The angle of divergence can be used as a measure of the force between the leaves and also as a measure of the charge on them.

The electroscope can also be charged by induction. Let us suppose that a positively charged body is brought close to the rod of the electroscope. Some of free electrons of the rod-leaf assembly will be attracted to

the rod by the positively charged body, with the result that the leaves will be positively charged. The leaves will diverge. By touching the hand, or a ground connection, to the electroscope rod, additional negative charge is attracted to the leaves, and the leaves will collapse. If the hand is removed from the rod, and then the positively charged body is removed, the leaves of the electroscope will diverge again. The entire leaf-rod assembly has now been negatively charged by induction.

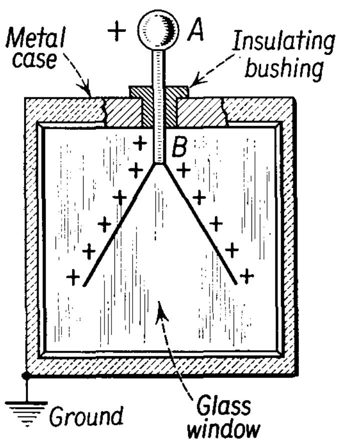


Fig. 22-4 A gold-leaf electroscope.

A charged electroscope is very convenient for determining the sign of the charge on any body brought near it. Suppose that the electroscope is positively charged and that the leaves diverge through a definite angle. If a positively charged body is brought near the electroscope, the leaves will diverge still more. The reason for this is that negative electrons will be attracted up from the leaves by the positively charged body, thus increasing the positive charge on the leaves. Of course if a negatively charged body is brought near the positively charged electroscope, the angle of divergence of its leaves will be decreased.

22-7 The Electrophorus

The process of charging by induction is the basis for the design and construction of many electrostatic machines used for supplying electricity for certain special needs. The *electrophorus* is a simple type of electrostatic machine. It consists of a flat disk made of some insulating material such as hard rubber, sealing wax, or resin, and a metal disk with an insulating handle. The disk A, made of insulating material, is first charged by rubbing it with wool or fur; this gives the disk a negative charge. The metal disk B is now brought very close to the charged disk A. (In actual practice, B is placed on A, but because of the roughness of its surface, contact is made at very few points; the charge transferred at these points is negligibly small.) The electrons in the metal disk are repelled to its upper surface, leaving the lower surface positively charged (see Figure 22-5). The plate B is now grounded momentarily by touching it with the hand or with a grounded wire; the electrons will go to ground, leaving the plate B positively charged. As long as this plate is close to the charged disk A, the positive charge will remain on the lower surface of B, but if this metal disk is removed far

enough from the plate A, the charges in the metal disk will redistribute themselves so that both upper and lower surfaces are positively charged. This positive charge is now available for use in any manner we please. For example, the plate can be discharged through a tube containing neon so that it causes the emission of light characteristic of neon.

If we examine the operation of the electrophorus once again, we note first that practically no charge has been removed from the original charged disk A; hence this plate can be used over and over again. Second, in removing the disk B from the vicinity of disk A, we had to work against the forces of attraction between the charges on the two disks. It was because of this work that there was energy available to operate the neon lamp or any other appropriate device. This process can be repeated indefinitely and as often as desired, the work being supplied by the person who separates the charges.

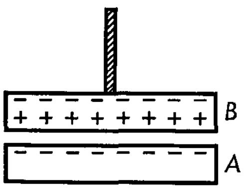


Fig. 22-5 The electrophorus.

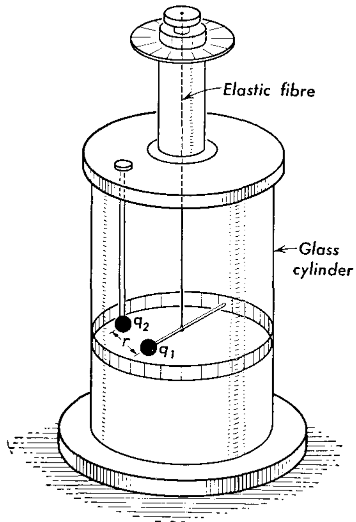


Fig. 22-6 Coulomb type torsion balance for determining the force between two charged bodies.

22-8 Coulomb's Law of Force between Charges

The first quantitative measurement of the force between charged bodies was made in 1785 by Charles Augustin de Coulomb (1736–1806) using an apparatus similar to that used by Cavendish for the determination of the gravitational constant G_0 (Section 6-15). A rod with a small charged sphere at one end was suspended in a horizontal position from an elastic fiber, and another small charged sphere was brought near it, as shown in Figure 22-6. The force between the two charged bodies was measured by noting the amount of twist in the fiber supporting the rod.

Coulomb found that *the force between two small spheres charged with*

electricity was inversely proportional to the square of the distance between them; that is,

$$F \propto \frac{1}{r^2}, \tag{22-1}$$

where F was the magnitude of the force exerted by each charged body upon the other when the two were separated by a distance r. Although Coulomb's determinations were made with equipment which could be considered crude by modern standards, more recent attempts to determine the law of force between charged particles have shown that the exponent of r in Equation (22-1) does not differ from 2 by more than 1 part in 1 billion. Experiments on the scattering of alpha particles from nuclei have shown that the inverse square law of force between charged particles is correct at distances of the order of 10^{-12} cm. This is of the order of the diameter of the nucleus.

Further research on the law of force between small charged bodies in *vacuum* showed that the force was proportional to the magnitudes of the charges as well. This can be put in the form of an equation

$$\boxed{F = k\frac{q_1 q_2}{r^2},} \tag{22-2}$$

where q_1 is the magnitude of the charge on the first body, q_2 is the magnitude of the charge on the second body, and k is some constant of proportionality which depends upon the system of units used. The force is directed along the line joining the two bodies and is repulsive if the bodies are of like charge but is attractive if the bodies are of unlike charge. While Equation (22-2) is rigorously true only for vacuum, it is very nearly correct for air at atmospheric conditions.

Equation (22-2), which is known as *Coulomb's law,* is in exactly the same form as the law of universal gravitation, as stated in Equation (6-17). There is one important difference between the law of gravitation and Coulomb's law. The constant G_0 of Equation (6-17) relates quantities which had been previously related and defined through Newton's second law. In Equation (22-2) both the constant k and the unit of charge remain to be defined and evaluated. We must decide whether to assign arbitrarily some value to k and thus let Equation (22-2) provide the basis for a definition of a unit of charge, or whether to define arbitrarily a unit of charge and use Equation (22-2) to provide the value of k. In practice, both methods are used.

22-9 The CGS Electrostatic Unit of Charge

In the cgs electrostatic system of units, the constant k of Coulomb's law is arbitrarily set equal to 1 for a vacuum; it is a dimensionless quantity. *The unit of electric charge is defined as one which, when placed one centimeter from a like equal charge in vacuum, will repel it with a force of one dyne.* The unit of charge is called the *statcoulomb* (stcoul), and is sometimes also referred to as the *esu of charge,* as an abbreviation of "electrostatic unit." The esu of charge, or the statcoulomb, is widely used in the older literature of physics and engineering and is currently used in atomic and nuclear physics, but it is too small in magnitude for practical purposes. In the cgs electrostatic system of units, Coulomb's law is written as

$$F = \frac{q_1 q_2}{r^2}, \tag{22-2a}$$

where F is expressed in dynes, q is expressed in statcoulombs, and r is expressed in centimeters.

The unit of charge which has been universally adopted for practical application is called the *coulomb* (coul) which may be defined for our present purposes as approximately 3×10^9 stcoul. More exactly,

$$1 \text{ coul} = 2.998 \times 10^9 \text{ stcoul}.$$

The number of statcoulombs in a coulomb has been chosen as numerically equal to one tenth the velocity of light in free space, expressed in centimeters per second.

Illustrative Example. A small body carrying a charge of $+20$ stcoul is placed 6 cm from another small body carrying a charge of $+30$ stcoul. Both bodies are in vacuum. Determine the force acting on each body.

Since the physical quantities are stated in cgs units, we apply Coulomb's law in the form of Equation (22-2a)

$$F = \frac{20 \text{ stcoul} \times 30 \text{ stcoul}}{36 \text{ cm}^2}$$
$$= 16.7 \text{ dynes}.$$

The force on the 20-stcoul charge is 16.7 dynes and is directed away from the 30-stcoul charge along the line joining the two charges. The force acting on the 30-stcoul charge is equal in magnitude but is opposite in direction.

22-10 Rationalized MKS System of Units

We have already seen that the mks system of units incorporates units of practical size to a far greater extent than does the cgs system of units.

The mks system of units has been extended to include electrical quantities by incorporation of the coulomb as a unit of charge of practical size. The arbitrary choice of a unit of charge implies that the choice of the constant k in Equation (22-2) can no longer be arbitrary but must be determined experimentally. Both the magnitude of k and its dimensions are involved. When the charge is given in coulombs, the value of k is no longer unity but is found to be approximately $9 \times 10^9 \dfrac{\text{nt m}^2}{\text{coul}^2}$. More exactly,

$$k = 8.987 \times 10^9 \frac{\text{nt m}^2}{\text{coul}^2}.$$

In practical engineering computations, equations which are developed from Coulomb's law are used much more frequently than is Coulomb's law itself. In order to simplify these equations, it is convenient to incorporate a factor of 4π into Coulomb's law in order that this factor may be eliminated from many other equations used in routine computation. For this reason the proportionality factor k in Coulomb's law has been redefined in terms of a new constant ϵ_0 (epsilon zero) as

$$k = \frac{1}{4\pi\,\epsilon_0}, \tag{22-3}$$

so that

$$\epsilon_0 = \frac{1}{4\pi k} = 8.85 \times 10^{-12} \frac{\text{coul}^2}{\text{nt m}^2}.$$

Coulomb's law in mks units is therefore written as

$$\boxed{F = \frac{1}{4\pi\epsilon_0}\frac{q_1 q_2}{r^2},} \tag{22-2b}$$

where F is expressed in newtons, q is expressed in coulombs, and r is expressed in meters. The constant ϵ_0 is called the *permittivity* of vacuum. A system of units in which the factor $\dfrac{1}{4\pi}$ appears in Coulomb's law is called a *rationalized system of units*. The system of units introduced in this section is generally known as the *rationalized mks system of units*. The units of ϵ_0 are sometimes stated in terms of other electrical units than those given above, but these all reduce to $\dfrac{\text{coulomb}^2}{\text{nt m}^2}$, as, for example, $\dfrac{\text{coulomb}^2}{\text{joule m}}$.

Illustrative Example. Two equally charged spheres, each having a mass of 1 gm, are suspended from a common point by silk threads 1 m long, as shown in Figure 22-7(a). Find the charge on each sphere if the angle between the threads is 10°.

The forces acting on each sphere are its weight mg, the tension in the string T, and the electrical repulsive force F_e, as shown in Figure 22-7(b). Since each sphere is in equilibrium under the action of these forces, their vector sum is zero, thus the vector diagram is a triangle. From Figure 22-7(c) we see that

$$F_e = mg \tan \frac{\theta}{2}.$$

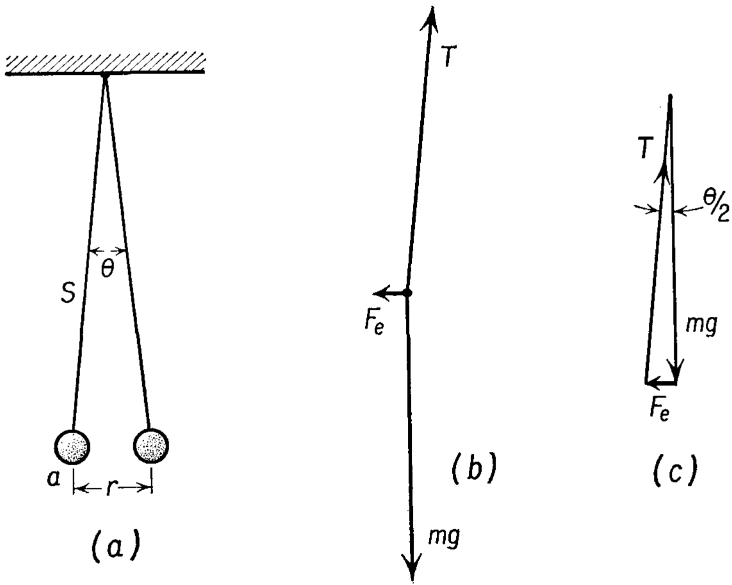


Fig. 22-7 (a) Electrified spheres hung from a common point. (b) Forces on the sphere at a. (c) The vector sum of the forces is zero so the vectors form a triangle.

Putting in numerical values, we get

$$F_e = 0.001\text{ kg} \times 9.8 \frac{\text{m}}{\text{sec}^2} \times 0.087 = 85 \times 10^{-5}\text{ nt}.$$

Now
$$\begin{aligned} r &= 2s \sin 5° \\ &= 2 \times 1\text{ m} \times 0.087 \\ &= 0.174\text{ m}. \end{aligned}$$

From Coulomb's law

$$q^2 = 4\pi\epsilon_0 F_e r^2,$$

so that
$$q^2 = 1.11 \times 10^{-10} \frac{\text{coul}^2}{\text{nt m}^2} \times 85 \times 10^{-5}\text{ nt} \times (0.174)^2\text{ m}^2.$$

Therefore
$$q = 5.3 \times 10^{-8}\text{ coul}.$$

22-11 Vector Form of Coulomb's Law

The formulas displayed thus far as statements of Coulomb's law have indicated only the magnitude of the force between two charged bodies *in vacuo*. For this reason numerical values without sign have been substituted for q in the illustrative examples. The direction of the force was given by statements not included in the formula itself.

To develop a self-contained vector statement of Coulomb's equation, we must first define a *unit vector*. If $\mathbf{r}$ is the vector drawn from an origin located at the charge q_1 to the charge q_2, as shown in Figure 22-8, the unit vector $\mathbf{1}_r$ may be defined by the relationship

$$\mathbf{1}_r = \frac{\mathbf{r}}{r}. \tag{22-4}$$

The unit vector $\mathbf{1}_r$ is a vector of unit length which is directed from q_1 to q_2.

Coulomb's law may be expressed in vector form as

$$\mathbf{F}_2 = k\frac{q_1 q_2}{r^2}\mathbf{1}_r, \tag{22-5}$$

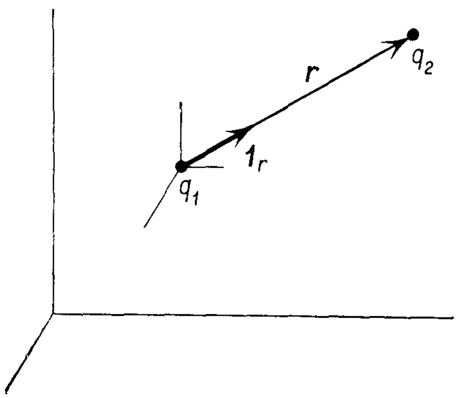


Fig. 22-8 The unit vector $\mathbf{1}_r$ is a vector of unit length directed from q_1 to q_2.

where $\mathbf{F}_2$ is the force acting on the charge q_2 as a result of the presence of charge q_1. Since Equation (22-5) is a vector equation, it is necessary to write both the magnitude and the sign of the charge when substituting values into it. Thus, if q_1 and q_2 are of the same sign, the force $\mathbf{F}_2$ is parallel to the unit vector $\mathbf{1}_r$, while if the charges are of opposite sign, the force is in the direction of $-\mathbf{1}_r$, or oppositely directed to $\mathbf{1}_r$.

22-12 Atomicity of Charge; Conservation of Charge

Measurements made on a macroscopic scale seem to indicate that the electric charge on a body may have any arbitrary value. However, careful measurements of the electric charges on small droplets show that every electric charge consists of an integral multiple of a certain quantity of charge called the *charge of an electron* and designated by the letter e. The numerical value of this charge is

$$\begin{aligned} e &= 4.802 \times 10^{-10} \text{ stcoul} \\ &= 1.602 \times 10^{-19} \text{ coul.} \end{aligned}$$

No charge smaller than this has ever been found. All charged particles, both positive and negative, no matter how large or small, have always been found to have electric charges which are integral multiples of e.

According to present theory, all matter is composed of one or more elements; a few more than 100 elements are now known (see Table of Elements in the Appendix). Each element is composed of atoms which have certain common properties. The atoms of any one element have the same number of protons; this number is called the *atomic number* of the element and is designated by the letter Z. These Z protons are in a very small volume of the entire atom known as the *nucleus*. The charge of a proton is positive and equal to e; thus the nucleus has a charge Ze. In an electrically neutral atom there are Z electrons outside the nucleus. These electrons extend to distances from the nucleus equal to about 10,000 or 100,000 times the nuclear radius. The nucleus, in spite of its small size, also contains neutral particles, called *neutrons;* the only exception to this is the nucleus of the ordinary hydrogen atom, which is simply the proton. Atoms of any one element may differ in the number of neutrons in the nucleus. The neutron is slightly more massive than the proton.

Nuclei heavier than hydrogen are assembled out of the building blocks of protons and neutrons. The mass of the final nucleus is less than the mass of the appropriate number of unassembled neutrons and protons, but the charge of the nucleus is simply the summed charge of all the constituent protons. Although mass is converted into other forms of energy in the process of assembling the nucleus, the electric charge is conserved.

It is a fundamental concept of electricity that the total charge of the universe is constant; that charge is neither created nor destroyed. This may be called the principle of *conservation of charge*. No violation of this principle has ever been observed.

The principle of conservation of charge can be applied to any closed system. If some object inside this system acquires an excess of one kind of charge, another object or objects in this system must acquire an equal and opposite charge.

When we speak of a substance as being electrically neutral, we mean that the total number of positive charges is equal to the total number of negative charges. Since electrons are comparatively far from the nucleus, it is easier to remove electrons from an atom than it is to remove protons. When a body is charged positively, for example, it generally is done by removing electrons from it; if it is charged negatively, electrons are usually added to it. In *annihilation* reactions studied in nuclear physics (see Chapter 46), an electron and positron are observed to disappear simultaneously. Their mass is converted to radiant energy, called annihilation radiation. But the total charge of the universe remains constant, for the total charge of the electron-positron pair is zero before and after the event.

The process inverse to annihilation is known as *pair production,* in which electromagnetic radiation may be converted into mass by the conversion of the radiant energy into an electron-positron pair. Once again, charge is conserved. No event has been observed in which a single charged particle has either been annihilated or has materialized, and we may assert with some confidence that no such event will be observed in violation of the principle of conservation of charge.

Problems

[NOTE: 1 microcoulomb (abbreviated μcoul) = 10^{-6} coulomb.]

22-1. Two small spheres have charges of +600 stcoul and +300 stcoul and are 4 cm apart. Determine the magnitude and direction of the force exerted by the 300-stcoul charge on the 600-stcoul charge.

22-2. Two small spheres have charges of +200 μcoulombs and −50 μcoul and are 50 cm apart. Determine the magnitude and direction of the force exerted by the −50-μcoul charge on the 200-μcoul charge.

22-3. Determine the force, in newtons, exerted by the proton on the electron of the hydrogen atom if each has a charge of 1.60×10^{-19} coul and their separation is 0.53×10^{-8} cm.

22-4. In the Bohr model of the hydrogen atom, the electron is supposed to rotate about the proton in a circular orbit. Assuming that the centripetal force required for the circular motion is supplied by the force of electrostatic attraction, find (a) the frequency of rotation and (b) the linear speed of the electron in its orbit. Use the constants given in Problem 22-9.

22-5. Three point charges of 3, 4, and 5 stcoul, respectively, are located at the corners of an equilateral triangle of sides 10 cm long. Find the magnitude and direction of the force on the 5-stcoul charge. Place the 5-stcoul charge at the origin, and locate the 4-stcoul charge at coordinates (10, 0).

22-6. Two small spheres of mass 0.1 gm are hung from a common point on silk threads 50 cm long. When the spheres are equally charged, the angle between the threads is 15°. Find the charge on each sphere.

22-7. Two small spheres of equal mass are hung from threads 1 m long. The points of support of the threads are separated by a distance of 10 cm. When the charge on each sphere is 1 μcoul, the angle between the threads is 30°. Determine the mass of each sphere.

22-8. A point charge of 10 stcoul is located at the origin and a second charge of −40 stcoul is located at a point along the x axis 2 cm to the right of the first charge. (a) What is the force on a third charge of magnitude 1 stcoul which is located at coordinates (0, 4 cm)? (b) Where along the x axis may this third charge be placed so that the force acting on it is zero?

22-9. In the Bohr model of the hydrogen atom, an electron of mass 9.107×10^{-28} gm and of charge -4.803×10^{-10} stcoul is located at a distance of 0.529×10^{-8} cm from a proton of mass 1.672×10^{-24} gm and charge $+4.803 \times 10^{-10}$ stcoul. Determine the ratio of the gravitational force of attraction to the electrical force of attraction between the two particles.

23

The Electric Field

23-1 The Electric Field

We have previously described the gravitational field as one way of thinking about gravitational forces (Section 6-16). If an object of mass m at rest at a point P experienced a force, we could attribute that force to the presence of the gravitational field. In a similar way we may attribute the force experienced by an electric charge at rest at a point P to the presence of an *electric field* at that point. From the preceding chapter we recognize that the existence of a force on a charged particle is due to the presence of other charged particles in the vicinity, but for many purposes it is unnecessary to have a precise knowledge of the positions of these charges or their magnitudes. In the same way we have utilized a knowledge of the gravitational field intensity g to analyze the trajectories of projectiles without detailed knowledge of the mass distribution which gave rise to g.

We can use a very small body containing a small charge q as a means of exploring the electric field in any region of space. If the charge q experiences a force $\mathbf{F}$ at a given point, the *electric field intensity* $\mathbf{E}$ at this point is defined by the equation

$$\boxed{\mathbf{E} = \frac{\mathbf{F}}{q}.} \tag{23-1}$$

The electric field intensity at a point in an electric field is the force per unit charge at this point. The test charge or probe charge q should be sufficiently small so that it will not change the distribution of charge that gives rise to the field to be measured. Ideally, one can take smaller and smaller test charges q, measure the force on each test charge at the given point, and take the limit of the ratio of F/q as q gets smaller as the electric field intensity at the point.

The electric field intensity is a *vector* quantity; it is the result of divid-

ing force, a vector, by charge, a scalar. Equation (23-1) is analogous to Equation (6-18), which has been used to define the gravitational field. In using Equation (23-1) we must remember to substitute the sign of q as well as its numerical value to get the correct directional relationship between the vector quantities **E** and **F**. The direction of the electric field is opposite to the direction of the force on a negatively charged particle. The units of electric field intensity in the mks system may be expressed as *newtons per coulomb.*

As we have seen in Chapter 22, the equations relating electrical quantities take on different appearances in the cgs and mks systems of units.

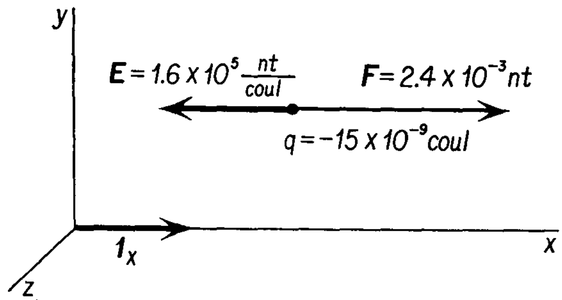


Fig. 23-1 The electric field intensity is opposite in direction to the force on a negatively charged particle.

In order to avoid complication in the body of the chapter, the equations in this and subsequent chapters will be developed using mks units. Discussions relating the two systems of units will be in smaller type to distinguish them from the principal development. The principal equations of the chapter are repeated in both systems of units in Table 23-1, and the relationships between the units of the two systems are stated in Table 23-2, at the end of the chapter.

In defining the intensity of the electric field, we must emphasize that the test body used to probe the field is *at rest.* We shall see in a subsequent chapter that a moving charged particle may experience a force *proportional to its speed.* Such a force is due to the presence of a *magnetic* field. The magnetic field does not exert a force on a charged particle at rest. The electric field may therefore be defined by the force on a *stationary* particle. Once the electric field intensity is known, it may be used to compute the force on a charged particle without regard to whether the particle is at rest or in motion. The force on a charged particle at a given point due to the electric field is determined from Equation (23-1) as

$$\mathbf{F} = \mathbf{E}q \tag{23-1a}$$

and is independent of the speed of the particle.

Illustrative Example. A small particle having a charge of -15×10^{-9} coulomb experiences a force of 2.4×10^{-3} nt in the positive x direction, as shown in Figure 23-1. Find the electric field intensity at that point.

The magnitude of the electric field intensity is

$$E = \frac{F}{q} = \frac{2.4 \times 10^{-3}\text{ nt}}{15 \times 10^{-9}\text{ coul}}$$

$$= 1.6 \times 10^{5}\,\frac{\text{nt}}{\text{coul}}\cdot$$

The electric field is in the negative x direction, opposite to the direction of the force on a negatively charged particle. If we write $\mathbf{1}_x$ for a unit vector in the positive x direction, the electric field intensity may be written as

$$\mathbf{E} = -1.6 \times 10^{5} \times \mathbf{1}_x\,\frac{\text{nt}}{\text{coul}}\cdot$$

23-2 Electric Charge as the Source of Electric Field

In the mks system of units, the vector form of Coulomb's law may be written as

$$\mathbf{F}_2 = \frac{q_1 q_2}{4\pi\epsilon_0 r^2}\,\mathbf{1}_r.$$

We may consider the charge q_2 as a probe charge used to explore the electric field. The electric field intensity $\mathbf{E}$ at the point where the charge q_2 is located may be found by dividing both sides of the above equation by q_2. Thus

$$\mathbf{E} = \frac{\mathbf{F}_2}{q_2} = \frac{q_1}{4\pi\epsilon_0 r^2}\,\mathbf{1}_r,$$

and, dropping the subscript 1 from q_1, the electric field intensity at a distance r from a point charge q, in vacuum, is given by the equation

$$\boxed{\mathbf{E} = \frac{q}{4\pi\epsilon_0 r^2}\,\mathbf{1}_r.} \tag{23-2}$$

The unit vector $\mathbf{1}_r$ is directed from the charge q which generates the field to the point P where we imagine the test charge to be located, as shown in Figure 23-2.

Illustrative Example. A point charge of -10 μcoul is located at the origin. Find the electric field intensity at a point in the x-y plane whose coordinates are (3 m, 4 m).

The location of the charge and the field point P are shown in Figure 23-3.

The unit vector $\mathbf{1}_r$ is shown directed from the charge q at the origin to the point P. Substituting into Equation (23-2), we find

$$\mathbf{E} = \frac{-10 \times 10^{-6}\ \text{coul}}{4\pi \times 8.85 \times 10^{-12}\,\dfrac{\text{coul}^2}{\text{nt m}^2} \times 25\ \text{m}^2} \times \mathbf{1}_r,$$

$$\mathbf{E} = -3.6 \times 10^3\,\frac{\text{nt}}{\text{coul}} \times \mathbf{1}_r.$$

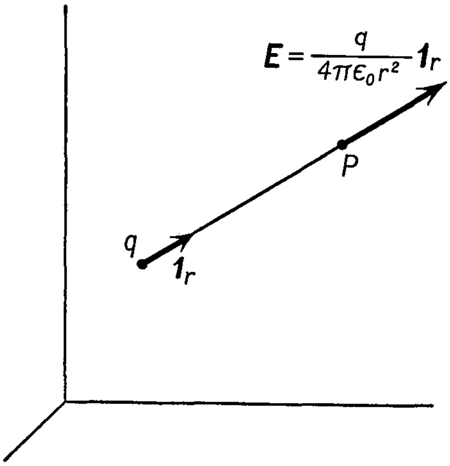


Fig. 23-2

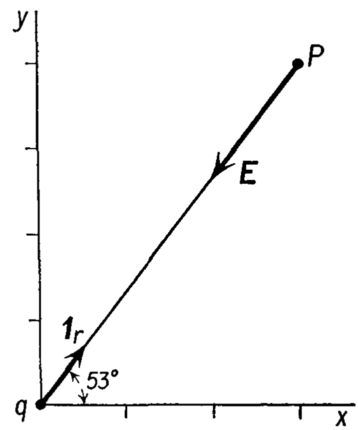


Fig. 23-3

Thus the magnitude of $\mathbf{E}$ is given by

$$E = 3.6 \times 10^3\,\frac{\text{nt}}{\text{coul}},$$

and its direction is in the direction of $-\mathbf{1}_r$, that is, toward the origin along a line making an angle of 53° with the positive x axis, as shown in the figure.

23-3 Electric Field Due to a Collection of Point Charges

A probe charge q placed at a point P in the neighborhood of any number of point charges $q_1, q_2, q_3, \cdots$ will experience a force which is the vector sum of the forces produced by the individual charges on it. Thus

$$\mathbf{F} = \mathbf{F}_1 + \mathbf{F}_2 + \mathbf{F}_3 + \cdots,$$

where $\mathbf{F}$ is the resultant force on charge q and $\mathbf{F}_1$ is the force on it produced by the charge q_1, $\mathbf{F}_2$ the force produced by charge q_2, and so forth.

Since
$$\mathbf{E} = \frac{\mathbf{F}}{q} = \frac{\mathbf{F}_1}{q} + \frac{\mathbf{F}_2}{q} + \frac{\mathbf{F}_3}{q} + \cdots,$$

therefore

$$\mathbf{E} = \mathbf{E}_1 + \mathbf{E}_2 + \mathbf{E}_3 + \cdots. \tag{23-3}$$

Hence the electric field intensity at any point P produced by a set of point charges in its neighborhood is the vector sum of the electric field intensities produced by the individual charges at the same point.

Illustrative Example. Two point charges $q_1 = 5\ \mu$coul and $q_2 = -5\ \mu$coul are separated by a distance of 0.08 m, as shown in Figure 23-4. Find the electric field intensity (a) at point a located on the line joining the two charges at a dis-

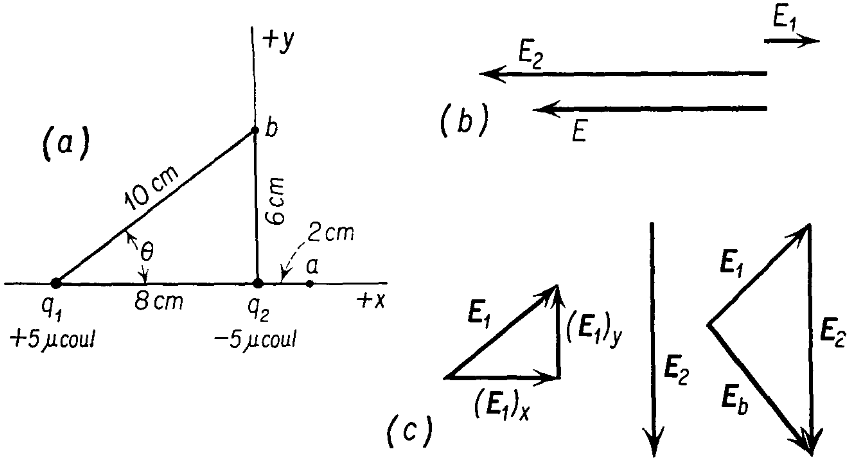


Fig. 23-4 (a) Location of charges and field points. (b) The fields E_1 due to charge q_1 and E_2 due to charge q_2 at the point a. (c) The field at b due to q_1, to q_2, and their resultant field E_b.

tance 10 cm from q_1 and 2 cm from q_2 and (b) at point b located at the vertex of a right triangle at a distance of 10 cm from q_1 and 6 cm from q_2.

(a) The electric field at a due to q_1 is

$$\mathbf{E}_1 = \frac{5 \times 10^{-6}\ \text{coul}}{4\pi \times 8.85 \times 10^{-12}\ \dfrac{\text{coul}^2}{\text{nt m}^2} \times (0.10\ \text{m})^2}\ \mathbf{1}_x$$

$$= 4.5 \times 10^6 \times \mathbf{1}_x \frac{\text{nt}}{\text{coul}}.$$

The electric field at point a due to q_2 is

$$\mathbf{E}_2 = \frac{-5 \times 10^{-6}\ \text{coul}}{4\pi \times 8.85 \times 10^{-12}\ \dfrac{\text{coul}^2}{\text{nt m}^2} \times (0.02\ \text{m})^2}\ \mathbf{1}_x$$

$$= -112.5 \times 10^6 \times \mathbf{1}_x \frac{\text{nt}}{\text{coul}}.$$

The resultant electric field is therefore

$$\mathbf{E} = \mathbf{E}_1 + \mathbf{E}_2$$

$$= (4.5 - 112.5) \times 10^6 \times \mathbf{1}_x \frac{\text{nt}}{\text{coul}},$$

$$\mathbf{E} = -108 \times 10^6 \times \mathbf{1}_x \frac{\text{nt}}{\text{coul}}.$$

(b) The electric field at b due to charge q_1 is

$$E_1 = \frac{5 \times 10^{-6}\,\text{coul}}{4\pi \times 8.85 \times 10^{-12}\,\dfrac{\text{coul}^2}{\text{nt m}^2} \times (0.10\,\text{m})^2}$$

$$= 4.5 \times 10^6 \frac{\text{nt}}{\text{coul}},$$

and the x and y components of this field are

$$(E_1)_x = E_1 \times \cos\theta = 4.5 \times 10^6 \frac{\text{nt}}{\text{coul}} \times \tfrac{8}{10};$$

$$(E_1)_x = 3.6 \times 10^6 \frac{\text{nt}}{\text{coul}}.$$

$$(E_1)_y = E_1 \times \sin\theta = 4.5 \times 10^6 \frac{\text{nt}}{\text{coul}} \times \tfrac{6}{10};$$

$$(E_1)_y = 2.7 \times 10^6 \frac{\text{nt}}{\text{coul}}.$$

The electric intensity at b due to q_2 is

$$\mathbf{E}_2 = \frac{-5 \times 10^{-6}\,\text{coul}}{4\pi \times 8.85 \times 10^{-12}\,\dfrac{\text{coul}^2}{\text{nt m}^2} \times (0.06\,\text{m})^2}\,\mathbf{1}_y;$$

$$\mathbf{E}_2 = -12.5 \times 10^6 \times \mathbf{1}_y \frac{\text{nt}}{\text{coul}}.$$

The components of the resultant electric intensity are therefore given by

$$E_x = (E_1)_x = 3.6 \times 10^6 \frac{\text{nt}}{\text{coul}}.$$

$$E_y = (E_1)_y + (E_2)_y = (2.7 - 12.5) \times 10^6 \times \frac{\text{nt}}{\text{coul}};$$

$$E_y = -9.8 \times 10^6 \frac{\text{nt}}{\text{coul}}.$$

The resultant field may be expressed as the vector sum of its x and y components as

$$\mathbf{E} = (3.6 \times 10^6 \times \mathbf{1}_x - 9.8 \times 10^6 \times \mathbf{1}_y) \frac{\text{nt}}{\text{coul}}.$$

23-4 Electric Field Due to a Continuous-Charge Distribution

When the electric field is established by a continuous distribution of charge rather than by a collection of point charges, we may compute the electric field intensity by imagining the charge distribution to be cut up into small volume elements in which the entire charge of an element is considered to be concentrated at some point of the volume element, say its center. The electric intensity may then be computed by applying Equation (23-3) to this collection of charges. More generally, we apply the methods of the calculus and replace the sum by an integral in the limit of an extremely fine subdivision.

If $d\mathbf{E}$ is the contribution to the electric intensity at the point P from a volume element whose charge is dq, located at a distance r from the point P, we may write

$$\mathbf{E} = \int d\mathbf{E} = \int \frac{dq}{4\pi\epsilon_0 r^2}\,\mathbf{1}_r, \tag{23-4}$$

where $\mathbf{1}_r$ is the unit vector directed from the element of charge dq to the point P. Let us define the *charge density* ρ of a continuous-charge distribution as the quantity of charge per unit volume. The charge dq in a volume element dv is then given by

$$dq = \rho\, dv,$$

and Equation (23-4) becomes

$$\mathbf{E} = \int \frac{\rho\, dv}{4\pi\epsilon_0 r^2}\,\mathbf{1}_r. \tag{23-4a}$$

In Equations (23-4) the integration must be carried out over the entire charge distribution. These equations are useful for symbolic purposes only. In order to carry out the integration, it is necessary to replace them by equations which yield the components of the electric intensity at the point P. Thus

$$E_x = \int dE_x,$$

$$E_y = \int dE_y,$$

and so on.

Illustrative Example. Calculate the electric intensity at a point on the axis of a uniformly charged narrow ring of charge.

Let us locate the ring, of radius a, in the x-y plane. The axis of the ring is along the z coordinate axis. Let the charge of the ring be q. Its linear charge density is therefore $\dfrac{q}{2\pi a}$. The electric field intensity dE, contributed by an ele-

ment of the ring to a point P located at (0, 0, z), is shown in Figure 23-5. The element subtends an angle $d\theta$ and has a charge dq given by

$$dq = a\,d\theta \cdot \frac{q}{2\pi a} = \frac{q\,d\theta}{2\pi}.$$

The intensity dE is given by

$$dE = \frac{\dfrac{q\,d\theta}{2\pi}}{4\pi\epsilon_0(a^2 + z^2)}.$$

Since the electric intensity contributed by each element of the ring is in a different direction in space, we cannot integrate without first finding components along the coordinate axes. The z component of dE is given by

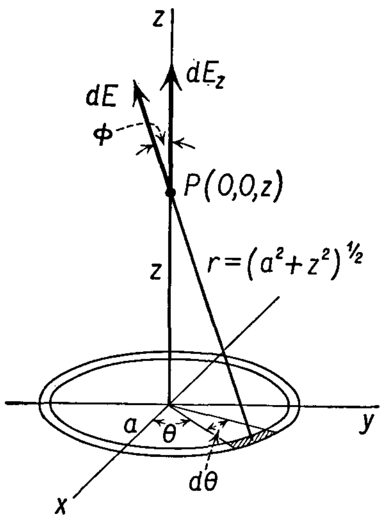


Fig. 23-5

$$dE_z = dE\cos\phi = dE\frac{z}{(a^2 + z^2)^{1/2}}$$

$$= \frac{1}{4\pi\epsilon_0}\frac{q}{2\pi}\frac{z}{(a^2 + z^2)^{3/2}}\,d\theta.$$

Since a, the radius of the ring, and z, the coordinate of the field point, are fixed, we may integrate over the entire region of charge by integrating $d\theta$ over the range 0 to 2π to obtain

$$E_z = \frac{q}{4\pi\epsilon_0}\frac{z}{(a^2 + z^2)^{3/2}}.$$

Because of the symmetry of the figure with respect to the z axis, the components of the electric intensity in a direction perpendicular to the z axis at the point P must sum to zero. For every element of the ring which contributes a component to the electric intensity parallel to the x-y plane, there is an equal element across the diameter of the ring which contributes a component in the opposite direction. Thus the electric intensity at all points on the axis of the ring is parallel to the axis of the ring and is given by the above formula.

In making the computation, we have followed the procedure outlined in the preceding paragraph, by integrating the components of the electric intensity vector. This must be done in integrating any vector quantity, for the reason that the process of integration is essentially the process of computing the limit of an algebraic sum.

23-5 Lines of Force

To visualize the direction and magnitude of the electric field in space, it is convenient to make use of the concept of *lines of force*, which was first

introduced in Section 6-16 to represent the gravitational field. *The electric lines of force are drawn so that a tangent to the line at any point will give the direction of the electric field intensity at that point. The magnitude of the electric intensity is given by the number of lines passing perpendicularly through a unit area centered at the point,* as shown in Figure 23-6. No two lines of force may cross each other, for this would infer that the force on a unit positive charge had two directions at the point of crossing.

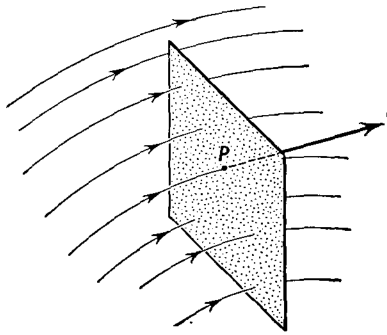


Fig. 23-6 Representation of the electric field intensity at P by the number of electric lines of force passing perpendicularly through a unit area at P.

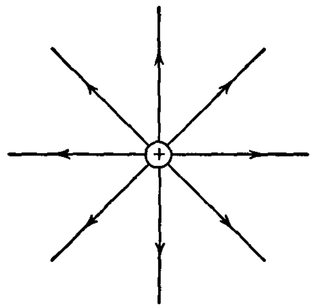


Fig. 23-7 The electric field around a small positive charge is radial and directed away from the charge.

Let us consider the appearance of the lines of force about a positive charge. Since the electric field is directed radially away from a positive charge, as shown in Figure 23-7, the lines of force are also directed away from the positive charge. The lines of force surrounding a negative charge are directed radially toward the negative charge. Thus, in vacuum, a line of force begins on positive charge and terminates on negative charge. According to Equation (23-2), the electric intensity about a charged body in vacuum is

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

The total number of lines of force passing through a sphere of radius r concentric with q is N, the product of the electric intensity, or the number of lines per unit area, by the area of the sphere, or

$$N = 4\pi r^2 E,$$

so that

$$\boxed{N = \frac{q}{\epsilon_0}.} \qquad \textbf{(23-5)}$$

Thus the total number of lines of force radiating from a positively charged body in vacuum is given by the charge of the body divided by ϵ_0. The number of lines of force terminating on a negatively charged body is given by the same quotient.

23-6 Gauss's Theorem

The lines of force radiating from a positive charge must terminate on an equal and opposite negative charge somewhere in the universe. In dealing with an isolated charged body, as shown in Figure 23-7, we think of the lines of force as terminating on negative charges at infinity. Suppose we enclose a region of charge-free space with a sphere, or with some other simple closed surface which can be reshaped into a sphere. Let us count the total number of lines of force coming out of the surface, tallying those leaving the surface as positive, and those entering the surface as negative. Since there is no charge within the closed surface, no lines of force originate or terminate within the surface. Any line of force which enters the surface at one point must leave the surface at some other point. There can be no net lines of force leaving or entering a volume that does not enclose a charge; that is, there are as many lines entering the closed surface as there are leaving it.

If there are several charges or a charge distribution within the closed surface, each coulomb of positive charge generates $1/\epsilon_0$ lines of force, and the same number of lines of force must terminate on each coulomb of negative charge. The net number N of lines of force leaving the surface must be given by

$$N = \frac{\sum q}{\epsilon_0}. \tag{23-5a}$$

To determine the relationship between the electric intensity and the number of lines of force, let us consider an element of area of a closed surface of magnitude ΔA. If the lines of force are perpendicular to this surface element, the electric intensity E is given by the number of lines ΔN leaving the volume through this surface element, divided by the area ΔA, in accordance with the convention we have chosen for representing the electric field by lines of force. In the form of an equation

$$E = \frac{\Delta N}{\Delta A}.$$

Now suppose the surface makes some angle with the lines of force, as shown in Figure 23-8. If the outward drawn normal to the surface makes an angle θ with the lines of force, the component of the element of area which

is perpendicular to the lines of force is given by $\Delta A \cos\theta$, so that

$$E = \frac{\Delta N}{\Delta A \cos\theta},$$

or

$$N = E\,\Delta A \cos\theta.$$

If we consider $\Delta \mathbf{A}$ as a vector quantity, as indicated in Section 8-2, whose magnitude is given by the area ΔA and whose direction is given by

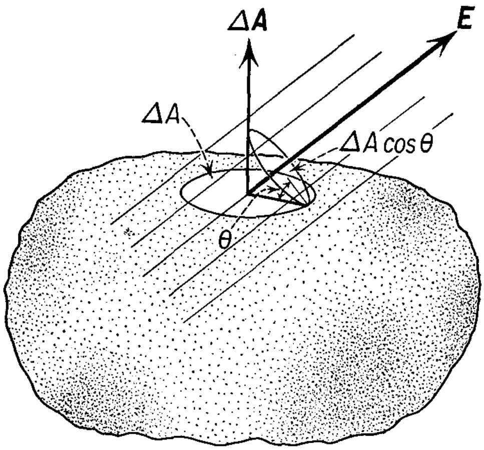


Fig. 23-8

the outward drawn normal to the surface, we can apply the definition of a scalar product given in Section 7-3, to write

$$\Delta N = \mathbf{E}\cdot\Delta\mathbf{A}.$$

To find the total number of lines of force leaving the surface, we imagine that the element of area $\Delta\mathbf{A}$ becomes very small, and in the limit we replace the symbol Δ by the symbol d, and integrate dN over the entire surface. This process may be represented symbolically as

$$N = \int dN = \int \mathbf{E}\cdot d\mathbf{A}. \tag{23-6}$$

On substituting for N its value in terms of the charge contained within the closed surface, we find

$$\boxed{\int \mathbf{E}\cdot d\mathbf{A} = \sum \frac{q}{\epsilon_0}\,.} \tag{23-7}$$

Equation (23-7) is a very important equation in electrostatics, and is known as *Gauss's theorem*. The theorem states that the *integral of the normal component of the electric field intensity over a closed surface is equal to the algebraic sum of the charge contained within that surface divided by* ϵ_0.

Gauss's theorem is of great usefulness in computing the electric field intensity E of symmetric charge distributions, in which it is possible to have some knowledge about the symmetry of E itself from observing the

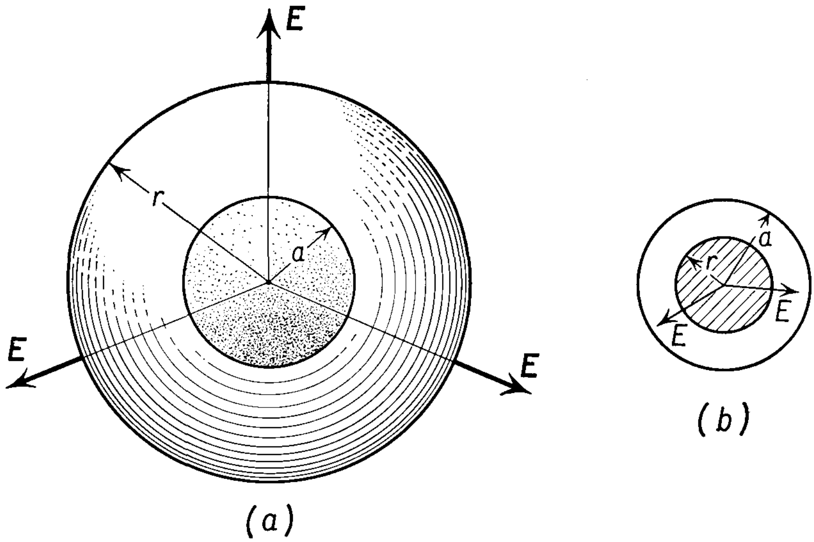


Fig. 23-9 Gaussian sphere of radius r concentric with charged sphere of radius a. (a) Gaussian sphere outside the charged sphere; (b) Gaussian sphere inside the charged sphere.

symmetry of the charge distribution. Let us consider the electric intensity associated with a sphere having a charge q distributed uniformly throughout its volume. To compute the electric intensity by subdividing the sphere into small volume elements and integrating, in the manner indicated in Section 23-4, is a tedious job. It is far simpler to observe that the value of E must be the same on all points of a second spherical surface, concentric with the first, because of the symmetry of the charge distribution. Let us draw such a surface of radius r, called a *Gaussian surface*, as shown in Figure 23-9. The radius of the uniformly charged sphere is a.

Since the direction of $\mathbf{E}$ is everywhere radial, the angle made by $\mathbf{E}$ with the normal to the surface of the Gaussian sphere is everywhere 0°. The quantity $\mathbf{E} \cdot d\mathbf{A}$ in Equation (23-7) reduces to the product of the magnitudes

of these two quantities, and we have

$$\int E\,dA = \frac{q}{\epsilon_0}.$$

Since E is everywhere constant over the surface of the Gaussian sphere, the quantity E may be taken outside the integral. The integral then represents the surface area of the Gaussian sphere, $4\pi r^2$. Thus we have

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

which leads to

$$E = \frac{q}{4\pi\epsilon_0 r^2} \tag{23-8}$$

for the magnitude of the electric field, whose direction is specified by the fact that it is everywhere radial. Notice that *the electric intensity everywhere outside a uniformly charged sphere is as though the charge were all concentrated at the center of the sphere.* Since the law of universal gravitation is of the same form as Coulomb's law, the above result is also valid for the gravitational field about a uniform spherical mass distribution. It was for this reason that we could treat the earth as though its mass were concentrated at its center. The same result obviously applies to a uniformly charged spherical shell and, in fact, to any distribution of charge having spherical symmetry. Thus any distribution of charge having spherical symmetry generates an electric field outside the charge distribution which is as though the entire charge were concentrated at the center of the sphere.

If we wish to find the electric intensity inside the charge distribution we draw a Gaussian sphere of radius $r < a$ concentric with the charged sphere, and again we observe that the electric field intensity must be radially directed and must be of equal magnitude at all points of the Gaussian sphere, from considerations of symmetry. Applying Gauss's theorem, we find

$$4\pi r^2 E = \frac{1}{\epsilon_0}\left(\frac{4}{3}\pi r^3 \rho\right),$$

where ρ is the density of charge within the sphere and is given by

$$\rho = \frac{q}{\frac{4}{3}\pi a^3} = \frac{3q}{4\pi a^3}.$$

Thus we have

$$E = \frac{\rho r}{3\epsilon_0} = \frac{qr}{4\pi a^3 \epsilon_0}. \tag{23-9}$$

Comparing the results of Equations (23-8) and (23-9), we see that both equations lead to the same result at the surface of the charged sphere where

$r = a$, namely

$$E = \frac{q}{4\pi\epsilon_0 a^2}. \tag{23-10}$$

The results of Equations (23-8) and (23-9) have been plotted in Figure 23-10. The electric intensity has a maximum value at the surface of the charged sphere and diminishes to zero at the center of the sphere and at infinity. These results are of some interest in practical problems, as, for example, in the electrical effects associated with the flight of an airplane

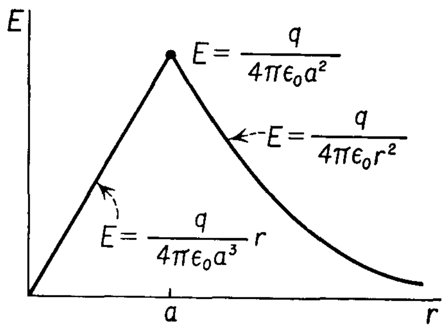


Fig. 23-10 Electric field intensity of a uniformly charged sphere of radius a as a function of the distance r from the center of the sphere.

through a thundercloud. These results are also of interest in atomic physics, where the electric intensity at the position of an outer-valence electron is made up of the field contributed by the central nucleus, which follows Coulomb's law, and the field of the inner electrons, which may be approximated by a uniformly charged sphere. The field experienced by the outer electron no longer follows an inverse square law when this electron penetrates the charged cloud, and this, in turn, serves to determine some important properties of atoms.

23-7 Conductors

By the term *electrostatics* we mean the study of the properties of electric charge *at rest.* Within the framework of electrostatics, there can be no electric field within a conductor, for by the word *conductor* we imply that electric charge is free to move. If there were an electric field within a conductor, there would be a force on the free electrons of that conductor; these electrons would be accelerated; hence they would not be at rest. Thus, simply as a matter of the consistency of our definitions, we must conclude that, for distributions of charge which are in static equilibrium, the electric intensity within a conductor is zero. In making such a statement we are speaking of an idealized conductor, for any material object is made up of

nuclei and electrons which are held together by electric forces. There are intense electric fields on a submicroscopic level of examination. In the present approximation of a conductor, we imagine the conductor to be made up of a completely homogeneous material which may be subdivided into infinitesimal parts without altering its properties. This approximation has been found experimentally valid as long as the smallest subdivision we permit ourselves to examine is one which contains hundreds of atoms. For practical purposes this is still a very small volume element.

Since the electric field within a conductor is zero, we know from Gauss's theorem that the electric charge within any portion of a conductor must be zero. There must be an equal quantity of positive and negative charge within any subvolume of the conductor. Hence, *all the charge on a charged conductor must reside on its surface.* If a hole is made within the body of a conductor, the electric intensity within that hole must be zero.

Much electrical equipment is built so that all the working parts are contained within a metallic box, generally built of sheet copper or aluminum, and called an *electrostatic shield.* Electric charges outside that box cannot produce any electric field within the box, and consequently cannot affect the operation of equipment within the box. The box therefore provides a shield against outside electrical disturbances. This effect may be demonstrated by placing an electroscope in the vicinity of an electrostatic generator or a highly charged rod. If the rod is positively charged, some negative charge is attracted to the ball of the electroscope, with the result that the leaves become positively charged and repel each other, as shown in Figure 23-11(a). If a metallic cap is placed over the electroscope, the electroscope is shielded, and no matter how great the charge on the rod,

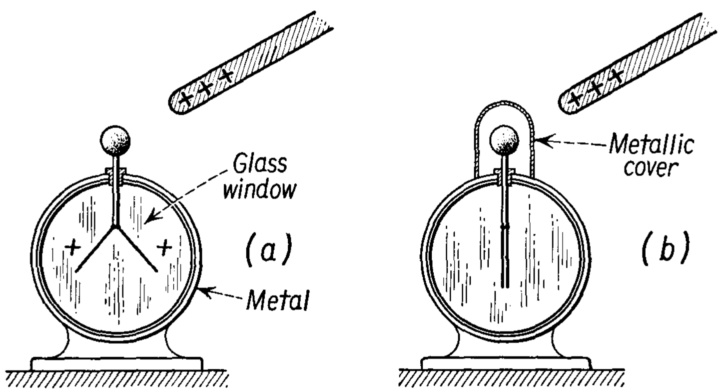


Fig. 23-11 (a) Leaves of an electroscope diverge when a charged rod is brought near it. (b) When an electroscope is shielded by a metallic cover, the charged rod does not affect the electroscope.

the leaves remain vertical, as shown in Figure 23-11(b). The metallic braid woven over the insulation of wires used in many electronic circuits is another illustration of the practical use of electrostatic shielding.

23-8 Field Outside the Surface of a Conductor

The electric field intensity immediately outside the surface of a conductor must be directed perpendicularly to the surface of the conductor. To understand this, we observe that if the electric field were oblique to the surface, it could be resolved into components parallel and perpendicular to the surface. Once again, our argument is based upon the definition of electrostatics. If there were a component of the electric intensity parallel to the conducting surface, the free electrons on the surface of the conductor would be accelerated and would no longer be at rest.

Let us suppose that there is an electric field in the vicinity of a conducting surface. Some of the lines of force associated with that field will go toward the conducting surface and terminate abruptly on it, perpendicular to the surface; other lines of force will originate from the surface and leave it perpendicularly. There are no lines within the conductor. We have already seen that lines of force begin or end on electric charges. Thus, if there is an electric field normal to the surface of the conductor, the con-

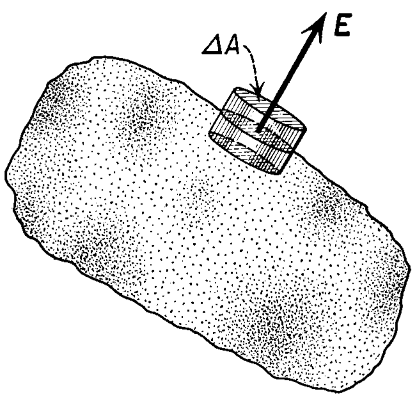


Fig. 23-12

ductor must be charged. We may apply Gauss's theorem to evaluate the relationship between the electric field at the surface of a conductor and the surface charge on that conductor.

Let us suppose that the electric charge on the surface of the conductor is of charge density σ units of positive charge per unit area. The electric field intensity normal to the surface of the conductor is E. To find the relationship between σ and E, we make use of a Gaussian surface in the shape of a *pillbox*, with cylindrical walls normal to the conducting surface, and plane top and bottom faces parallel to the conducting surface. One

face of the pillbox is imagined to be within the conducting material, while the other face is just outside the conducting surface, as shown in Figure 23-12. If the area of the face of the pillbox is ΔA, the charge contained within the pillbox is $\sigma\ \Delta A$.

The electric intensity within the conductor is zero. Furthermore, since the electric field is normal to the surface, no lines of force pass through the cylindrical walls of the pillbox. In applying Gauss's theorem in the form of Equation (23-7), the only contribution to the integral of the normal component of E is obtained from the face of the pillbox outside the conductor. Thus we have

$$E\ \Delta A = \frac{\sigma\ \Delta A}{\epsilon_0},$$

and

$$E = \frac{\sigma}{\epsilon_0}. \tag{23-11}$$

If we represent a unit vector normal to the surface of the conductor and directed outward from the conductor as $\mathbf{1}_n$, the electric intensity at the surface of the conductor is related to the surface density of charge by the equation

$$\mathbf{E} = \frac{\sigma}{\epsilon_0}\mathbf{1}_n. \tag{23-11a}$$

Thus, if the charge on the surface of the conductor is positive, $\mathbf{E}$ is parallel to $\mathbf{1}_n$; hence $\mathbf{E}$ is outwardly directed. If the charge on the surface is negative, $\mathbf{E}$ is opposite in direction to $\mathbf{1}_n$ and is inwardly directed.

It is often desired to measure the electric field at the surface of a conductor, as, for example, at the belly of an airplane in flight or at the surface of the earth. One means of doing this is to measure the charge on a unit area of the surface of the conductor and to apply Equation (23-11) to determine the electric field intensity from the surface density of charge. A small section of the conducting surface of the airplane may be insulated from the remainder of the surface and may be periodically removed from the skin of the airplane, brought within the fuselage, and connected to an electroscope to measure its charge. More practically, if an electrically isolated segment of the airplane's skin is alternately covered and uncovered by a rotating conducting plate which is electrically connected to the skin, as shown in Figure 23-13, the isolated segment may be thought of as being first on the surface of the airplane, then within the conducting shell, and so on. It becomes electrically charged when it is on the surface and is discharged through a high resistance leading to the skin of the airplane when it is covered by the rotating conductor. In such an arrangement the charge which flows to and from the isolated segment through the resistor may be

electronically amplified, and the electric field may be readily determined. From such measurements it is known that the average fair-weather electric field intensity at the surface of the earth is 100 nt/coul. The earth carries

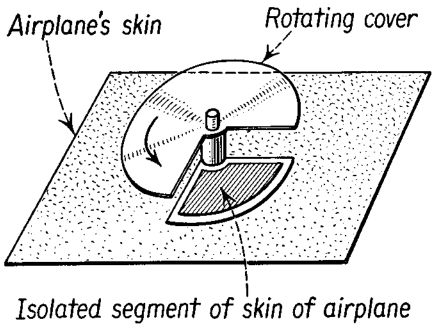


Fig. 23-13 Schematic illustration of electric field meter.

a negative charge whose total value is of the order of 500,000 coul, and whose surface density is about 0.0009 coul/km^2. The electric field intensity at the belly of an airplane in flight through a thunderstorm can be as much as 340,000 nt/coul just prior to a lightning strike.

23-9 Dielectric Strength

Let us suppose that a gas is placed in an electric field, and that one of the molecules of the gas has become ionized (or charged), say as a result of a collision with another rapidly moving molecule. The charged ion is accelerated by the electric field. The force acting upon the ion is the product of the electric field intensity E by the charge q of the ion. Suppose further that the charged particle moves an average distance l before making a collision with another molecule; l is called the *mean free path.* The work done on the particle by the electric field is given by the product Eql. This is the energy acquired by the charged particle in the interval between collisions. If the energy delivered to the ion by the field is sufficiently great so that the ion can disrupt a molecule with which it collides, at least two additional ions result from the collision process, each of which may be again accelerated by the field and may make subsequent collisions.

In this way a large number of ions may be created, and the gas then becomes electrically conducting. The electric field at which a gas becomes conducting is called the *dielectric strength* of the gas. Clearly, if there are no gas molecules present, as in vacuum, the dielectric strength is infinite, for there are no molecules present to become ionized. When only a very few molecules are present, as within a vacuum tube, an ionized particle may reach the electrodes of the tube without making a collision with another gas molecule, so that the total charge transferred is small. Again, when

the pressure is quite high, the mean free path of the molecules is reduced, so that an ion cannot move far enough in the field to acquire the energy required to disrupt another molecule. The dielectric strength of the gas is high, for a large electric field intensity is required to generate secondary ionization.

The dielectric strength of air at atmospheric pressure is about 3×10^6 nt/coul. When this is exceeded, the air becomes conducting; corona dis-

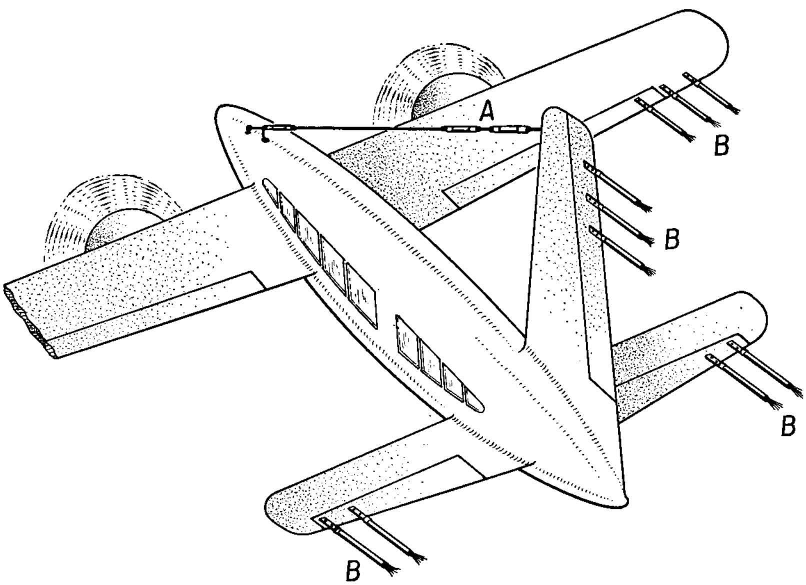


Fig. 23-14 Antistatic antennae A and static dischargers B on an airplane.

charge may be observed as a bluish glow in the region of an intense electric field, for the recombining ions give off some of their energy as light, and the sharp smell of ozone may be distinguished.

When an airplane becomes electrically charged in flight through precipitation, or when a charge is induced on the surface of an airplane when it flies near a thundercloud, corona discharge may take place from the propellers, from the wingtips, and from the radio antennae. The discharge from the antennae is particularly serious, for the erratic nature of the discharge generates radio noise called *precipitation static;* this often drowns out the signal from a radio station used by the pilot for communication and navigation purposes. Static dischargers, shown in Figure 23-14, have been placed upon the wingtips of many airplanes so that the electric charge accumulating on the airplane may be discharged noiselessly to the air. At the same time the wire radio antennae of the airplane have been insulated

with an insulating material of high dielectric strength, and specially designed antenna fittings have been used to minimize the electric field intensity at the terminations of the wires so that the corona discharge does not take place from the antenna wire itself or from its associated fittings.

The dielectric strength of insulating materials is a technically important property of insulators. Their values range from about 10^6 to about 10^8 nt/coul. It is interesting that many solid insulating materials have a dielectric strength which is not appreciably greater than air. Solid insulating material in electrical apparatus is used generally as a spacer, to keep conductors from making contact with each other rather than to improve upon the insulating properties of atmospheric air.

TABLE 23-1 PRINCIPAL EQUATIONS IN THE MKS AND CGS UNITS

Equation	MKS	CGS	
(23-1)	$\mathbf{E} = \mathbf{F}/q$	Same as mks	Electric field
(23-2)	$\mathbf{E} = \dfrac{q}{4\pi\epsilon_0 r^2}\mathbf{1}_r$	$\mathbf{E} = \dfrac{q}{r^2}\mathbf{1}_r$	Field of a point charge
(23-5)	$N = \dfrac{q}{\epsilon_0}$	$N = 4\pi q$	Lines of force
(23-7)	$\int \mathbf{E}\cdot d\mathbf{A} = \sum \dfrac{q}{\epsilon_0}$	$\int \mathbf{E}\cdot d\mathbf{A} = 4\pi q$	Gauss's theorem
(23-11a)	$\mathbf{E} = \dfrac{\sigma}{\epsilon_0}\mathbf{1}_n$	$\mathbf{E} = 4\pi\sigma\mathbf{1}_n$	Field outside a conductor

TABLE 23-2 CONVERSION FACTORS RELATING MKS AND CGS UNITS

Quantity	Symbol	MKS Unit	CGS Unit
Charge	q	1 coul	$= 3 \times 10^9$ stcoul (esu)
Electric intensity	E	$1\,\dfrac{\text{nt}}{\text{coul}}$	$= \dfrac{1}{3 \times 10^4}\,\dfrac{\text{dyne}}{\text{stcoul}}$ (esu)
Force	F	1 nt	$= 10^5$ dyne

$$\text{Permittivity of free space: } \epsilon_0 = 8.85 \times 10^{-12}\,\frac{\text{coul}^2}{\text{nt m}^2},$$

$$\epsilon_0 = \frac{1}{36\pi \times 10^9}\,\frac{\text{coul}^2}{\text{nt m}^2}.$$

Problems

23-1. What is the intensity of the electric field at a distance of 20 cm from a small sphere charged to +1,600 stcoul?

23-2. What is the intensity of the electric field at a distance of 40 cm from a small body charged with $-7{,}200$ stcoul?

23-3. An electric charge of $+15$ stcoul is placed 25 cm from one of -40 stcoul. Determine the magnitude and the direction of the electric field intensity at the mid-point of a line joining them.

23-4. In Problem 23-3 determine the electric field intensity at a point in the first quadrant 20 cm from the $+15$-stcoul charge and 15 cm from the -40-stcoul charge. The line connecting the two charges is along the x axis, and the 15-stcoul charge is at the origin. State your answer in terms of x and y components of the electric field intensity.

23-5. Two equal charges each of $+2{,}500$ stcoul are placed 24 cm apart, along the y axis. Determine the electric field intensity at a point in the right-hand half plane 15 cm from each charge. State your answer in terms of the unit vectors $\mathbf{1}_x$ and $\mathbf{1}_y$.

23-6. The electric field intensity at a point P near a charge of 144 stcoul is 9 dynes/stcoul. Where must a charge of $+324$ stcoul be placed to reduce the field intensity at P to zero? Give the position of P relative to each charge.

23-7. Two small charged bodies are placed 25 cm apart along the x axis. One has a charge of $+600$ μcoul and the other has a charge of $-1{,}800$ μcoul. Find the electric field intensity at a point 60 cm from the positive charge and 65 cm from the negative charge. State your answer in terms of unit vectors along the x and y axes.

23-8. A small charge of $+12$ stcoul is placed in a uniform electric field of 300 dynes/stcoul. Determine the force on this charge.

23-9. A body having a mass of 0.01 gm and a charge of 1 μcoul is placed in a uniform electric field. What must be the magnitude and direction of the electric field intensity if the body is to remain at rest under the influence of the electric and gravitational fields.

23-10. How many lines of force emanate from a charge of $+5$ μcoul (a) in the mks system of units? (b) In the cgs system of units?

23-11. A metallic sphere is charged to 10 μcoul, the charge being uniformly distributed over the surface of the sphere. If the sphere is 1 m in diameter, find the electric field intensity (a) at the surface of the sphere and (b) at the center of the sphere.

23-12. A large metal sphere contains a surface charge of 2 $\mu\text{coul/m}^2$. What is the electric field intensity at the surface of the sphere?

23-13. A thin spherical shell of charge of radius 1 m has a total charge of 1 coul. (a) What is the electric field intensity at the center of the shell? (b) What is the electric field intensity at a point 50 cm from the center of the shell? (c) What is the electric field intensity 2 m from the center of the shell?

23-14. Derive a formula for the electric field intensity at distance r from the center of a long uniformly charged cylinder of charge of radius a, of charge density ρ per unit volume (a) when r is less than a and (b) when r is greater than a. Check your results for these two cases by comparing them when r is equal to a. Use a Gaussian surface composed of a cylinder of radius r concentric with the charged cylinder and apply the consequences of cylindrical symmetry to Gauss's theorem.

24

Potential

24-1 Potential Difference

A positive charge q situated at some point A in an electric field where the intensity is $\mathbf{E}$ will experience a force $\mathbf{F}$ given by Equation (23-1a) as

$$\mathbf{F} = \mathbf{E}q.$$

In general, if this charge q is moved to some other point B in the electric field, an amount of work $\Delta\mathscr{W}$ will have to be performed. *The ratio of the work done* $\Delta\mathscr{W}$ *to charge* q *transferred from point* A *to point* B *is called the difference of potential* ΔV *between these points;* thus

$$\boxed{\Delta V = V_B - V_A = \frac{\Delta\mathscr{W}}{q},} \tag{24-1}$$

where V_A is the *potential* at A, and V_B is the potential at B.

The quantity of charge q should be so small that it does not disturb the distribution of the charges which produce the electric field E. Or we may imagine that successively smaller charges are used, and that the work done in moving each such charge is determined. The limit of the ratio of the work done to the charge transferred as the charge gets progressively smaller is the potential difference between the two points A and B. Difference of potential is thus the work per unit charge that would be done in transferring charge from one point to another. A difference of potential can exist between two points even though no charge is actually transferred between them.

Potential difference is a scalar quantity, since both work and charge are scalar quantities. If a positive charge q is transferred from A to B, and if the work is done by some outside agency against the forces of the electric field, then point B is said to be at a higher potential than A; if the work is

done by the electric field in moving a positive charge from A to B, then the potential at A is higher than that at B.

The unit of potential difference in the mks system is the *volt*. From Equation (24-1) we see that one volt is equal to one joule per coulomb.

In the cgs system of units, the unit of potential difference is the statvolt, which is equal to one erg per statcoulomb. We may find the relationship between the volt and the statvolt by following the usual unit conversion procedure. Thus

$$1 \text{ volt} = \frac{1 \text{ joule}}{1 \text{ coul}} \times \frac{10^7 \text{ ergs}}{1 \text{ joule}} \times \frac{1 \text{ coul}}{3 \times 10^9 \text{ stcoul}}$$

$$= \frac{10^7 \text{ ergs}}{3 \times 10^9 \text{ stcoul}} = \frac{1}{300} \text{ statvolt.}$$

More exactly,

$$1 \text{ statvolt} = 299.6 \text{ volts.}$$

In the above discussion we have used the term *potential at a point*, while the definition was in terms of the difference of potential between two points. The term "potential at a point" can have meaning if we decide upon some reference point as a point of zero potential. In practical work this reference point or zero level of electrical potential is usually taken as the earth or the ground, and the potential at any other point is measured with respect to it. Electrical equipment is practically always connected to earth or to *ground* at some point, and other potentials are spoken of as being so many volts above or below ground potential. In many calculations in physics, in dealing with the properties of finite charge distributions, without reference to their position with respect to the earth, it is convenient to refer potentials to the potential of a *point at infinity*. In such calculations a point infinitely distant from the charge distribution is considered as the zero of potential.

The assignment of a zero of electrical potential is thus somewhat arbitrary and is analogous to the assignment of the position of zero potential energy in dealing with a particle in the earth's gravitational field. Although we did not develop the concept of gravitational potential in mechanics, the gravitational potential difference may be defined in terms of the work per unit mass in moving a mass between two points in the field. Since there is no work done in moving a mass along a frictionless level surface in the gravitational field, all points on a level surface have the same gravitational potential. The work done in raising a mass m through a height h in a field of gravitational intensity g is mgh, and, dividing by m, we see that the gravitational potential difference is gh. Thus altitude is a measure of gravitational potential, which is customarily referred to an arbitrary zero of altitude at sea level.

24-2 Potential Due to a Point Charge in Vacuum

An isolated point charge in vacuum generates an electric field which is given by Equation (23-2) as

$$\mathbf{E} = \frac{q}{4\pi\epsilon_0 r^2}\,\mathbf{1}_r.$$

We recall that in this equation r is the distance from the charge q to the point where the field is being evaluated, and the unit vector $\mathbf{1}_r$ is directed from the charge to the field point. Let us place the charge q at the origin of coordinates and calculate the work which must be done by some outside agency in moving a positive charge q' from a point A in the electric field at a distance r_a from the origin, to a point B in the field at a distance r_b from the origin, as shown in Figure 24-1.

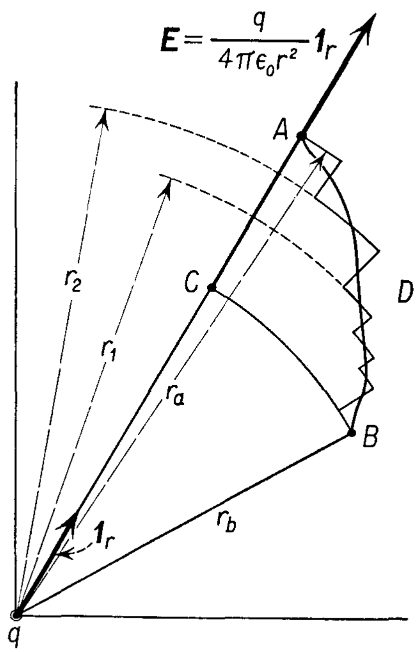


Fig. 24-1

Let us first show that the work done does not depend upon the path by which the charge q' is moved from A to B. To do this we shall consider two alternate routes. In the first of these we shall move q' radially along the line AC, and then along the arc of a circle CB. In this path, work is done only along the radial portion of the path, for here the force which must be exerted is equal and opposite to the force experienced by q' due to the electric field of q. No work is done by the agency displacing q' along the circular portion of the path, for here the force exerted is radial and is perpendicular to the displacement, which is tangential.

We may approximate the second route ADB as closely as we please by a succession of radial and circular displacements. Again we see that work is done only during a radial displacement, for in each of the circular displacements the displacement is perpendicular to the force. Since the magnitude of the electric intensity depends only upon the distance from the origin and not upon the angular position, we see that the force exerted by the outside agency, and therefore the work done in a displacement between radial coordinates r_2 and r_1, is the same whether this radial dis-

placement takes place along the path ACB or along the path ADB, or along any other path between A and B.

The *work done in carrying the charge* q′ *between two points in the field of a point charge* q *is therefore independent of the path.* The potential difference between the two points depends only on their position with respect to the charge q. To simplify the calculation of the potential difference between the points A and B, we choose the path ACB of Figure 24-1, where, as we have already seen, it is only necessary to calculate the work done along the radial portion of the path, AC. The force **F** which must be exerted by an external agency is equal and opposite to the force exerted by the electric field on this charge. The force on a positive charge q' is given by

$$\mathbf{F} = -\mathbf{E}q' = -\frac{qq'}{4\pi\epsilon_0 r^2}\mathbf{1}_r. \tag{24-2}$$

The mechanical work $\Delta\mathscr{W}$ done by the force **F** which is exerted on the charge q' in displacing it radially through a distance Δr toward the point charge q is given by

$$\Delta\mathscr{W} = F\,\Delta r = -\frac{qq'}{4\pi\epsilon_0 r^2}\Delta r. \tag{24-3}$$

To find the potential difference between the points A and B, we must sum up the work done in transporting the charge q' over all the increments of path. In the limit of small increments of displacement, the work done $\Delta\mathscr{W}$ is

$$\Delta\mathscr{W} = -\int_{r_a}^{r_b}\frac{qq'}{4\pi\epsilon_0 r^2}\,dr;$$

remembering that $$\int\frac{dr}{r^2} = -\frac{1}{r} + \text{const},$$

we find $$\Delta\mathscr{W} = \frac{qq'}{4\pi\epsilon_0}\left(\frac{1}{r_b} - \frac{1}{r_a}\right). \tag{24-4}$$

From the definition of the potential difference as the work done in transporting a test charge q' divided by the magnitude of that test charge, we have

$$\Delta V = \frac{\Delta\mathscr{W}}{q'} = \frac{q}{4\pi\epsilon_0 r_b} - \frac{q}{4\pi\epsilon_0 r_a}. \tag{24-5}$$

From Equation (24-1)

$$\Delta V = V_b - V_a.$$

If the initial point A is taken at $r_a = \infty$, it will be convenient to assign the

value zero to its potential. The potential at point B will then be given by the equation

$$V_b = \frac{q}{4\pi\epsilon_0 r_b}.$$

Dropping the subscript b, we find that the potential V at a point located at a distance r from a point charge q in vacuum is given by

$$V = \frac{q}{4\pi\epsilon_0 r}. \tag{24-6}$$

Note that in this expression V is an algebraic scalar quantity, which may be either positive or negative. The distance r is always a positive number, while the charge q must be replaced by a positive number for a positive charge and by a negative number if the charge is negative.

24-3 Potential Due to a Distribution of Charge in Vacuum

The potential at a point P due to a single point charge is a scalar quantity which represents the work per unit positive charge done in transporting charge from infinity to the field point P. Let us suppose that the electric field is generated by several point charges q_1, q_2, . . ., and so on. In transporting the test charge from infinity, work must be done against the electric field contributed by each of the charges q_1, q_2, and so on. The work done against the field of each of these charges separately is given by Equation (24–6). Since work, and therefore potential, is a scalar quantity, the total work done may be computed by finding the work done against the field due to charge q_1, the work done against the field of charge q_2, and so on, and then adding these algebraically. If the potential at P due to q_1 alone is V_1, the potential at P due to q_2 alone is V_2, and so on, and the potential V at P due to the entire charge distribution is

$$V = V_1 + V_2 + \cdots.$$

Thus we have, for the potential of a collection of point charges,

$$V = \sum \frac{q_i}{4\pi\epsilon_0 r_i}, \tag{24-7}$$

where r_i is the distance from the i'th charge q_i to the field point P at which the potential is being evaluated, and the summation is to be extended to all the charges in the distribution.

Illustrative Example. Two point charges, $q_1 = 5$ μcoul and $q_2 = -5$ μcoul, are separated by a distance of 8 cm, as shown in Figure 24-2. Find the potential at points a and b of that figure.

Since only two charges generate the field, the summation of Equation (24-7)

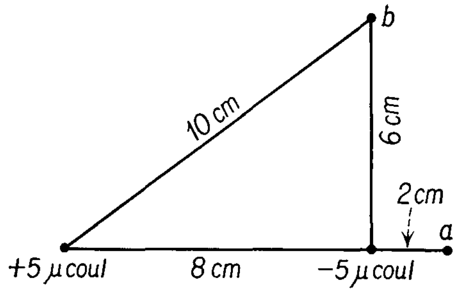
reduces to a sum of two terms. At point a, $r_1 = 0.1$ m and $r_2 = 0.02$ m. The potential V_a at point a is

$$V_a = \frac{5 \times 10^{-6}\ \text{coul}}{4\pi \times 8.85 \times 10^{-12}\ \dfrac{\text{coul}^2}{\text{nt m}^2} \times 0.1\ \text{m}} + \frac{-5 \times 10^{-6}\ \text{coul}}{4\pi \times 8.85 \times 10^{-12}\ \dfrac{\text{coul}^2}{\text{nt m}^2} \times 0.02\ \text{m}}$$

$$= 4.5 \times 10^5\ \frac{\text{nt m}}{\text{coul}} - 22.5 \times 10^5\ \frac{\text{nt m}}{\text{coul}}$$

$$= -18 \times 10^5\ \text{volts.}$$

Fig. 24-2



At the point b we see from the figure that $r_1 = 0.1$ m, while $r_2 = 0.06$ m. Substituting into Equation (24-7), we find

$$V_b = \frac{5 \times 10^{-6}\ \text{coul}}{4\pi \times 8.85 \times 10^{-12}\ \dfrac{\text{coul}^2}{\text{nt m}^2} \times 0.1\ \text{m}} + \frac{-5 \times 10^{-6}\ \text{coul}}{4\pi \times 8.85 \times 10^{-12}\ \dfrac{\text{coul}^2}{\text{nt m}^2} \times 0.06\ \text{m}},$$

$$V_b = 4.5 \times 10^5\ \text{volts} - 7.5 \times 10^5\ \text{volts}$$

$$= -3 \times 10^5\ \text{volts.}$$

In comparing this example to the corresponding illustrative example of Section 23-3, we see that the calculation of the potential is far simpler than the calculation of the electric intensity. This follows from the scalar nature of the potential and the vector nature of the electric intensity.

In the event that we have a continuous distribution of charge rather than a collection of discrete point charges, we may find the potential by integration rather than by summation. Equation (24-7) becomes

$$V = \int \frac{dq}{4\pi\epsilon_0 r}\,. \tag{24-8}$$

Illustrative Example. Calculate the potential at a point P on the axis of the uniformly charged narrow ring of Section 23-4, illustrated in Figure 23-5.

Every element of charge dq is at the same distance

$$r = (a^2 + z^2)^{1/2}$$

from the point P and contributes the same amount

$$dV = \frac{dq}{4\pi\epsilon_0 r}$$

to the potential at this point. Hence the potential at point P due to the entire ring of charge q is

$$\begin{aligned} V &= \frac{q}{4\pi\epsilon_0 r} \\ &= \frac{1}{4\pi\epsilon_0} \frac{q}{(a^2 + z^2)^{1/2}}. \end{aligned}$$

24-4 Equipotential Surfaces

An equipotential surface is a surface along which a charged body may be displaced without any work having been required in the process. The equipotential surface is defined as a locus of points in space at a common potential. In consequence, lines of force must intersect an equipotential surface at right angles to the surface, for if there were any component of the electric intensity parallel to an equipotential surface, work would be required to displace a charged body along that surface, in contradiction with our definition of an equipotential surface.

The surface of a conductor is an equipotential surface. The equipotential surfaces surrounding a point charge are concentric spherical surfaces centered at that charge; the potential of each such surface is given by Equation (24-6). The equipotential surfaces surrounding a uniformly charged cylinder are coaxial cylinders, with their common axis as the axis of the charged cylinder. In general, equipotential surfaces are drawn so that equal increments in potential separate each pair of surfaces. A few of the equipotential surfaces surrounding a point charge are shown in Figure 24-3, and some of the equipotential surfaces surrounding a charged metallic sphere are shown in Figure 24-4.

The knowledge that equipotential surfaces and lines of force intersect perpendicularly everywhere enables us to solve graphically many problems in electrostatics to a good approximation, even when these problems are too difficult for mathematical solution. Furthermore, any equipotential surface may be replaced by a metallic surface which is maintained at the appropriate potential without altering the electric field outside the conductor. Thus, for example, we may replace the appropriate imaginary equipotential surface of Figure 24-3 by a real metallic sphere which is maintained at the potential of that equipotential surface. If the charge inside the spherical shell is removed, there can be no lines of force inside

the conducting shell, and the inside of the shell is an *equipotential volume.* The lines of force and the equipotential surfaces outside the shell are unaltered. We may find the complete set of equipotential surfaces and lines

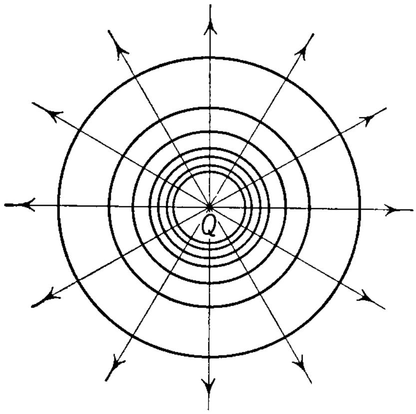


Fig. 24-3 The equipotential surfaces around a point charge are concentric spherical surfaces with the point charge at the center.

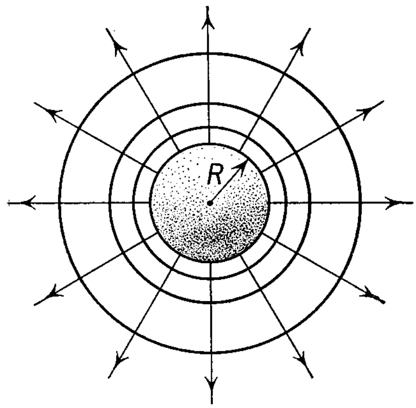


Fig. 24-4 The equipotential surfaces outside a charged metallic sphere are spherical surfaces concentric with the charged sphere.

of force of the sphere simply by erasing the lines of force and the equipotential surfaces within the conducting shell, as in Figure 24-4.

Two equal and opposite point charges $+q$ and $-q$ separated by a distance s constitute a dipole. The set of equipotential surfaces of a dipole intersect the plane of the diagram in the dotted lines shown in Figure 24-5, while the lines of force are shown as solid lines. At all points along a plane perpendicular to the line joining the two charges at the mid-point of the line, the potential is zero. At a field point P located in this plane the potential is

$$V = \frac{-q}{4\pi\epsilon_0 r} + \frac{q}{4\pi\epsilon_0 r} = 0.$$

This imaginary equipotential plane at zero potential may be replaced by a conducting plane at zero potential (obtained by connecting the plane to ground) without altering the field distribution. Thus if we are interested in obtaining the potential distribution of a charge q located a distance $s/2$ to the right of a grounded conducting plane, we may compute this field by finding the field and potential due to the dipole, and then erasing the lines of force and equipotential surfaces which appear to the left of the conducting plane.

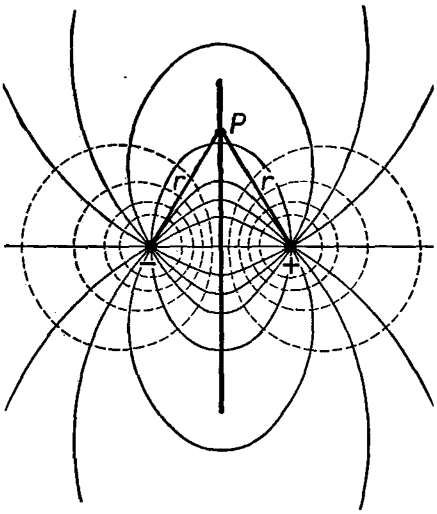


Fig. 24-5 Lines of force (solid) and equipotentials (dotted) about a dipole.

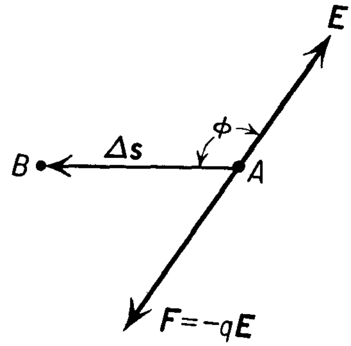


Fig. 24-6

24-5 Potential Gradient

When a charge q is displaced an amount $\Delta\mathbf{s}$ from point A to an adjacent point B, as shown in Figure 24-6, the force which must be exerted on the charge by an external agency is oppositely directed to the electric field and is given by

$$\mathbf{F} = -q\mathbf{E},$$

and the work done by this force is

$$\Delta\mathcal{W} = \mathbf{F}\cdot\Delta\mathbf{s}.$$

Substituting for $\mathbf{F}$ its value from the above equation, we find

$$\Delta\mathcal{W} = -q\mathbf{E}\cdot\Delta\mathbf{s}. \tag{24-9}$$

If we divide Equation (24-9) by the charge q, the quantity on the left-hand side of the equation is equal to the potential difference ΔV between the initial and final points of the displacement, giving

$$\Delta V = -\mathbf{E}\cdot\Delta\mathbf{s} = -E\,\Delta s\cos\phi. \tag{24-10}$$

In the limit of small displacements, we may find the potential from the electric field by integration. Symbolically, we write

$$\Delta V = \int_a^b -\mathbf{E}\cdot d\mathbf{s} = \int_a^b -E\,ds\cos\phi, \tag{24-11}$$

where ΔV represents the potential difference $V_B - V_A$ between the points A and B, and ϕ is the angle between $d\mathbf{s}$ and $\mathbf{E}$.

If we divide Equation (24-10) by the magnitude of the displacement Δs, we obtain the result that

$$\frac{\Delta V}{\Delta s} = -E \cdot \left(\frac{\Delta \mathbf{s}}{\Delta s}\right) = -\mathbf{E} \cdot \mathbf{1}_s, \tag{24-12}$$

where $\mathbf{1}_s$ is a unit vector in the direction of the displacement. Thus the rate of change of potential with distance in any direction, as specified by the direction of the unit vector $\mathbf{1}_s$, is equal to the negative of the component of the electric field intensity in that direction. If the direction of the unit vector is along the line of force, the rate of change of potential is greatest. At a given point the rate of change of the potential in the direction of most rapid change is called the *potential gradient* at that point. When the unit vector is directed along a line of force the angle ϕ between the unit vector $\mathbf{1}_s$ and the electric intensity $\mathbf{E}$ is zero, and we may write

$$E = -\frac{\Delta V}{\Delta s},$$

which may be rewritten as

$$E = -\frac{dV}{ds} \tag{24-13}$$

in the limit of small displacements.

The units of electric field intensity are therefore the same as the units of potential gradient. In the mks system of units, we may use either newtons per coulomb or volts per meter to represent either electric intensity or potential gradient. As we have seen in Section 23-9, the dielectric strength of air is approximately 3×10^6 nt/coul, or 3×10^6 volts/m. After walking across a carpeted room in the wintertime when the air of the room is quite dry, sparks as long as 5 cm may be observed to jump from one's knuckles to a doorknob. This implies that a difference of potential of approximately 1.5×10^5 volts exists between the doorknob and the knuckle.

If the potential is known as a function of the coordinates, we may find the components of the electric field parallel to any of the coordinate axes by imagining the displacement $\Delta \mathbf{s}$ to be parallel to that axis. Thus the component of the electric field intensity in the x direction, E_x, is given by the negative of the derivative of V with respect to x. In finding E_x by this means, we would examine the variation in V with respect to x holding the other coordinates constant. In the usual notation this is called a *partial derivative* and is represented by the symbol ∂ rather than the symbol d.

Following this convention, we write

$$E_x = -\frac{\partial V}{\partial x}; \qquad \text{(24-14a)}$$

similarly,

$$E_y = -\frac{\partial V}{\partial y}, \qquad \text{(24-14b)}$$

and

$$E_z = -\frac{\partial V}{\partial z}. \qquad \text{(24-14c)}$$

The electric field intensity **E** may then be expressed in terms of its components and the unit vectors in the coordinate directions; thus

$$\mathbf{E} = E_x\mathbf{1}_x + E_y\mathbf{1}_y + E_z\mathbf{1}_z. \qquad \text{(24-15)}$$

Illustrative Example. From the second example in Section 24-3, we have shown that the potential generated by a uniform ring of charge at a point along its axis is

$$V = \frac{1}{4\pi\epsilon_0}\frac{q}{(a^2 + z^2)^{1/2}}.$$

Find the electric intensity at a point on the axis of the ring by application of Equations (24-14).

First we note that the coordinates x and y do not appear in the expression for the potential. Thus the derivative of V with respect to x is zero, and the derivative of V with respect to y is zero. Hence there is no component of the electric field in the x or the y direction. To find the component of the field in the z direction, we apply Equation (24-14c).

$$E_z = -\frac{\partial V}{\partial z} = -\frac{\partial}{\partial z}\left[\frac{1}{4\pi\epsilon_0}\frac{q}{(a^2 + z^2)^{1/2}}\right].$$

Carrying out the indicated differentiation, we find

$$E_z = \frac{1}{4\pi\epsilon_0}\frac{qz}{(a^2 + z^2)^{3/2}}.$$

This result is identical with the formula obtained by direct integration in the illustrative example of Section 23-4.

In general, the electric field intensity may be obtained by the methods of Section 23-4. This requires three separate integrations to be performed, one for each of the components of the electric intensity. It is often much simpler to integrate once to find the potential, since this is a scalar quantity, according to the procedure of Section 24-3, and then to differentiate this result, following Equations (24-14) to find the electric field intensity.

TABLE 24-1 PRINCIPAL EQUATIONS IN MKS AND CGS UNITS

Equation	MKS	CGS	
(24-1)	$\Delta V = \dfrac{\Delta \mathscr{W}}{q}$	Same as mks	Potential difference
(24-6)	$V = \dfrac{q}{4\pi\epsilon_0 r}$	$V = \dfrac{q}{r}$	Point charge in vacuum
(24-12)	$\dfrac{\Delta V}{\Delta s} = -\mathbf{E}\cdot\mathbf{1}_s$	Same as mks	Potential gradient
(24-14a)	$E_x = -\dfrac{\partial V}{\partial x}$	Same as mks	Potential gradient

TABLE 24-2 CONVERSION FACTORS RELATING MKS AND CGS UNITS

Quantity	Symbol	MKS Unit	CGS Unit
Potential	V	1 volt	$= \dfrac{1}{300}$ statvolt (esu)
Electric intensity	E	$1\,\dfrac{\text{nt}}{\text{coul}} = 1\,\dfrac{\text{volt}}{\text{m}}$	$= \dfrac{1}{3\times10^4}\,\dfrac{\text{dyne}}{\text{stcoul}} = \dfrac{1}{3\times10^4}\,\dfrac{\text{statvolt}}{\text{cm}}$ (esu)
Charge	q	1 coul	$= 3\times10^9$ stcoul (esu)
Work	$\mathscr{W}$	1 joule	$= 10^7$ ergs

Permittivity of free space: $\epsilon_0 = 8.85 \times 10^{-12}\,\dfrac{\text{coul}^2}{\text{joule m}}$.

Problems

24-1. A small charge of +12 stcoul is placed in a uniform electric field whose intensity is 5,000 dynes/stcoul. (a) What is the force acting on this charge? (b) How much work is done by the electric field in moving this charge a distance of 4 cm in the direction of the field? (c) What is the difference in potential between its initial and final positions?

24-2. A small body carrying a charge of 72 μcoul is placed 0.60 m from another small body fixed in position, carrying a charge of 180 μcoul. If the 72-μcoul body moves to a place 0.90 m from the 180-μcoul body, what will be its kinetic energy?

24-3. A small charged body of 1 μcoul is released from rest in a region of space where the electric field intensity is 100 volts/m. What will its kinetic energy be when it has been displaced a distance of 150 cm?

24-4. Two plane metallic plates are located a distance of 1 cm apart. If the electric field between them is uniform, what must the potential difference between these plates be if the force on a 5-μcoul charge between the plates is to be 10^{-3} nt?

24-5. Two equal charges, each of +250 stcoul, are placed 24 cm apart on the x axis. Determine the potential at a point 15 cm from each charge.

24-6. An electric charge of +15 stcoul is located at the origin, and a charge of −40 stcoul is located at a point whose coordinates are (0, 20 cm). Find the potential at the following points: (a) (0, −5 cm), (b) (15 cm, 0), (c) (−15 cm, 0).

24-7. A charge of −13 μcoul is located at a point whose coordinates are (−5 m, 0) and a second charge of +30 μcoul is located at a point whose coordinates are (+9 m, 0). (a) Find the potential at the origin and (b) at a point whose coordinates are (0, +12 m). (c) How much work must be done by an external agency to move a 5-μcoul charge from the origin to the point (0, +12 m)?

24-8. An isolated conducting hollow sphere of radius 50 cm is charged to a potential of 100 statvolts. (a) What is the potential of the center of the sphere? (b) What charge placed at the center of the sphere would give an identical electric field distribution outside the sphere, if the conducting shell were removed? (c) What is the charge on the conducting sphere?

24-9. A uniformly charged sphere of radius a and charge density ρ coul/m^3 gives rise to an electric field outside the sphere of charge which is identical to the field generated by a point charge at the center of the sphere whose charge is $q = \frac{4}{3}\pi a^3 \rho$. Inside the charged sphere the electric field intensity is given by $E = (\rho/3\epsilon_0)r$. Find a formula for the potential at a point inside the charged sphere a distance r_0 from the center of the sphere.

24-10. The electric field intensity from a uniformly charged rod is directed radially and is given by the formula $E = \lambda/2\pi\epsilon_0 r$, where λ is the charge per unit length and r is the distance from the center of the cylinder. Find the potential difference between two points whose radial coordinates are r_a and r_b.

24-11. An electron volt (ev) is a unit of energy used in atomic and nuclear physics. It represents the energy acquired by an electron in falling through a potential difference of 1 volt. How many electron volts are there in 1 erg? The charge of the electron is 1.60×10^{-19} coul.

24-12. Two horizontal metallic plates are placed 1.5 cm apart, and a potential difference of 3,000 volts is applied between them so that the electric field is uniform and directed vertically. A small oil drop containing a charge of 32×10^{-19} coul and a mass of 10^{-10} gm is between the plates. (a) Determine the electrical force on the oil drop. (b) Determine the net force on the oil drop. (c) What potential difference must be applied to the plates for the oil drop to be in equilibrium under the action of both electrical and gravitational forces?

24-13. The binding energy of a hydrogen atom is 13.6 ev. What energy, in calories, would be required to separate the electron and proton to infinite distance from each other?

24-14. The heat of formation of water vapor is 57.8 kilocal/mole. What is the energy, in electron volts, which must be added to a molecule of water to dissociate it into hydrogen and oxygen?

24-15. An electron is liberated from the filament of a vacuum tube and is accelerated to the plate which is maintained at a potential of 300 volts above the filament. With what speed is the electron moving when it strikes the plate? The mass of the electron is 9.11×10^{-31} kg.

25

Capacitance and Dielectrics

25-1 Capacitance of an Isolated Sphere in Vacuum

When an isolated charged conducting sphere bears a charge Q, the potential of the sphere may be computed from the results of Section 23-6 by considering that the electric intensity outside the sphere is as though the entire charge of the sphere were concentrated at its center. The potential at a distance r from the center of the sphere is given by the formula

$$V = \frac{Q}{4\pi\epsilon_0 r}, \tag{24-6}$$

as long as r is greater than or equal to the radius of the sphere. At the surface of the sphere of radius a, the potential is

$$V = \frac{Q}{4\pi\epsilon_0 a}.$$

We may define the capacitance C of a conductor as the quotient of its charge Q divided by its potential V. That is

$$\boxed{C = \frac{Q}{V}.} \tag{25-1}$$

The capacitance is always a positive number, for if the charge on an isolated conductor is positive, its potential is positive; if the charge on an isolated conductor is negative, its potential is negative. Following the example of an isolated conductor, the capacitance for more complex cases is always given as a positive quantity.

From Equation (25-1) we find the capacitance of an isolated conducting sphere in vacuum to be

$$C = 4\pi\epsilon_0 a. \tag{25-2}$$

Thus the capacitance of a sphere is proportional to its radius. The mks unit of capacitance is called the *farad* in honor of Michael Faraday (1791–1867). The quantity ϵ_0 is often stated in units of farads per meter, as

$$\epsilon_0 = 8.85 \times 10^{-12} \text{ farad/m},$$

for consistency with Equation (25-2).

The farad is an impractically large unit of capacitance, being the capacitance of an enormous sphere of radius 9×10^9 m. The distance from the planet Mercury to the sun is 58×10^9 m. It is common practice to utilize units of microfarads (1 microfarad = 10^{-6} farad), abbreviated μfd, or units of micromicrofarads (1 micromicrofarad = 10^{-12} farad), abbreviated $\mu\mu$fd, to describe capacitance.

In the cgs electrostatic system of units, the potential of a charged sphere in vacuum is given by

$$V = \frac{Q}{a},$$

and the capacitance of a sphere is numerically equal to its radius in centimeters, for we find

$$C = a.$$

The cgs unit of capacitance is the capacitance of a sphere of radius one centimeter in vacuum, and is referred to as one *statfarad* (stfd). As a matter of general practice, the esu of capacitance is rarely used. From the preceding discussion

$$1 \text{ farad} = 9 \times 10^{11} \text{ stfd},$$

so that 1 $\mu\mu$fd is approximately equal to 1 stfd.

The capacitance of an isolated conductor of arbitrary size and shape may be bracketed between the capacitances of two spheres which just fit over the body and just fit inside the body. Thus the capacitance of an airplane is less than the capacitance of the smallest sphere which will enclose the airplane, and is greater than the capacitance of the largest sphere which will just fit inside the airplane.

25-2 Capacitors

To obtain the capacitance of an isolated conducting sphere, we applied the formula for the potential of the sphere in relation to its charge, which was based upon the assumption that the lines of force emanating from the conductor terminated at infinity. The potential of the conductor was established by assigning the potential zero to a point at infinity.

In most practical problems associated with electrical apparatus, the lines of force emanating from a conductor terminate upon other conductors

in the vicinity, and one must define the capacitance between any pair of conductors in a rather complex way. In the event that we are interested in two conductors which are close together and which are relatively far from other conductors in the vicinity, the lines of force emanating from one conductor terminate upon its neighbor rather than upon any distant conductor or upon an infinitely distant charge. This means that if a charge $+Q$ is placed on one conductor, a charge $-Q$ must be placed on the neighboring conductor.

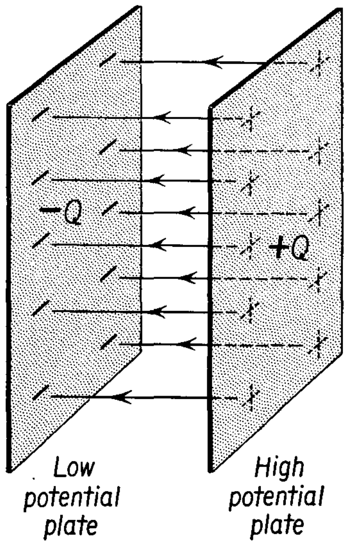


Fig. 25-1 The lines of force in the parallel plate capacitor are directed from the positive (high potential) plate to the negative (low potential) plate. The electric field is in the $-x$ direction in the figure while the potential increases in the $+x$ direction.

When the two conductors are close together, Equation (25-1) must be reinterpreted so that V is the potential difference between the two conductors, and Q is the charge on either conductor without regard to sign. A *capacitor* then consists of a pair of conductors adjacent to each other. The conductors may be in vacuum, or separated by air, or, more generally, separated by an insulating material which is referred to as a *dielectric.*

Let us consider the capacitance of a parallel-plate capacitor in vacuum. To a good approximation the results obtained will also be true when the conductors are in air, but they will have to be modified when a liquid or solid dielectric fills the space between the conductors.

When a pair of plane parallel conducting plates of area A are separated by a distance s, as shown in Figure 25-1, we may compute their capacitance by assuming that one plate has a charge of $+Q$ while the other plate has a charge of $-Q$. All lines of force beginning on the first plate will terminate on the second plate. We shall make the approximation that the electric field intensity is uniform between the plates. While this is strictly true

only for plates of infinite extent, it is very nearly true for plates whose dimensions are large compared to their separation.

The field outside the surface of a conductor is related to the surface density of charge on the conductor through the equation

$$E = \frac{\sigma}{\epsilon_0}. \tag{23-11}$$

The surface density of charge σ on the plates of the parallel-plate capacitor is given by

$$\sigma = \frac{Q}{A},$$

so that the electric intensity between the plates of the capacitor is

$$E = \frac{Q}{\epsilon_0 A}.$$

Since the field is uniform throughout the region between the capacitor plates, the work which must be done by an outside agency in moving a unit positive charge from the negative to the positive plate, through the distance s, is the potential difference V between the two plates, given by

$$V = Es = \frac{Qs}{\epsilon_0 A}.$$

From Equation (25-1) the capacitance is given by

$$C = \frac{Q}{V} = \frac{\epsilon_0 A}{s}. \tag{25-3}$$

The geometric properties of the capacitor, the area and the separation between the plates, determine the capacitance of a parallel-plate capacitor, just as the capacitance of a conducting sphere is determined by its radius.

Illustrative Example. Determine the capacitance of a huge parallel-plate capacitor whose plates are 1 km² in area, and which are 1 mm apart.

The area of each plate is $A = 10^6\ \mathrm{m}^2$; the separation of the plates is $s = 10^{-3}\ \mathrm{m}$; the value of ϵ_0 is 8.85×10^{-12} farad/m. Substituting these numerical values into Equation (25-3), we find

$$C = \frac{8.85 \times 10^{-12}\ \mathrm{farad/m} \times 10^6\ \mathrm{m}^2}{10^{-3}\ \mathrm{m}} = 8.85 \times 10^{-3}\ \mathrm{farad}.$$

In schematic circuit diagrams the symbol used to represent a fixed capacitor is shown in Figure 25-2, while the symbols used to represent variable capacitors are shown in Figure 25-3.

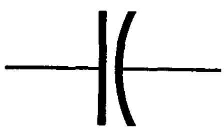


Fig. 25-2 Symbol used for representing a fixed capacitor. The terminal represented by the curved line is generally nearer to ground potential.

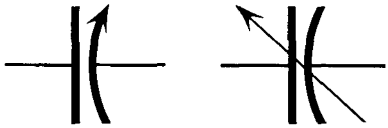


Fig. 25-3 Symbols used for representing variable capacitors.

25-3 Energy of a Charged Capacitor

When a capacitor is being charged, there is initially no charge on either plate, so that the electric field intensity between the plates is zero. To move the first increment of charge from one plate to the other requires no work. Later increments of charge transferred from one plate to the other must have work done upon them against the electric field of the charged plates. If the potential difference between the plates is V, the amount of work $\Delta \mathscr{W}$ done in transferring charge Δq from the plate at low potential to the plate at higher potential will be given by

$$\Delta \mathscr{W} = V\,\Delta q.$$

The relationship between the potential difference and the charge q on the plates is given by Equation (25-1) as

$$q = CV.$$

Thus the work $\Delta \mathscr{W}$ done in transferring charge Δq from the plate at low potential to the plate at higher potential is

$$\Delta \mathscr{W} = \frac{1}{C}\,q\,\Delta q.$$

To find the energy stored in the capacitor when it has been charged to its final potential difference V by transferring charge Q from one plate to the other, we must add all the increments in energy, and in the notation of the calculus

$$\mathscr{W} = \int d\mathscr{W} = \int_0^Q \frac{1}{C}\,q\,dq,$$

from which

$$\mathscr{W} = \frac{1}{2}\frac{Q^2}{C}. \tag{25-4a}$$

By substituting the value of Q from Equation (25-1) into Equation (25-4a), we may find alternate expressions for the energy of a charged capacitor.

We obtain

$$\mathscr{W} = \tfrac{1}{2}CV^2, \tag{25-4b}$$

and also

$$\mathscr{W} = \tfrac{1}{2}QV. \tag{25-4c}$$

From Equation (25-4c) the energy of the charged capacitor may be seen to be the energy associated with the transfer of the total charge Q through the average potential difference $V/2$. When mks units of coulombs, farads, and volts are substituted for the appropriate quantities in Equations (25-4), the energy $\mathscr{W}$ will be in joules.

Equations (25-4) illustrate one application of capacitors, that is, to store electrical energy, much as a stressed spring can be used to store mechanical energy. The ability of a capacitor to store electrical energy is utilized in electronic power supplies which convert alternating to direct current, and in high-energy accelerators, such as the *betatron*, which require so much energy in a short time that power-generating stations cannot conveniently supply it.

From another point of view, the energy of a charged capacitor may be said to reside in the electric field between its plates. The energy per unit volume in an electric field can readily be calculated with the aid of the equations just derived. The energy of a parallel-plate capacitor may be expressed, from Equations (25-4b) and (25-3), as

$$\mathscr{W} = \frac{1}{2}CV^2 = \frac{1}{2}\frac{\epsilon_0 A}{s}V^2.$$

Dividing this equation by the product As, the volume of the space between the capacitor plates, we find

$$\frac{\mathscr{W}}{As} = \frac{1}{2}\epsilon_0\frac{V^2}{s^2}.$$

The quantity V/s is the magnitude of the potential gradient, which is equal to the electric intensity E in the space between the plates of the capacitor. Let us designate the energy per unit volume in the space between the plates by $\mathscr{W}_v$.

Thus

$$\mathscr{W}_v = \frac{\mathscr{W}}{As},$$

and

$$E = \frac{V}{s},$$

so that

$$\mathscr{W}_v = \frac{\epsilon_0 E^2}{2}. \tag{25-5}$$

Thus the energy per unit volume in an electric field is proportional to the square of the intensity of the electric field.

Illustrative Example. Find the energy per unit volume in an electric field whose intensity is 100 nt/coul.

From Equation (25-5) we have

$$\mathscr{W}_v = \tfrac{1}{2} \times 8.85 \times 10^{-12} \frac{\text{coul}^2}{\text{nt m}^2} \times 10^4 \frac{\text{nt}^2}{\text{coul}^2}$$

$$= 4.42 \times 10^{-8} \frac{\text{nt}}{\text{m}^2}$$

$$= 4.42 \times 10^{-8} \frac{\text{joule}}{\text{m}^3}.$$

25-4 Capacitors in Series and Parallel

In many circuit applications several capacitors are electrically connected to produce the desired results. The simplest types of electrical connections are known as the *series* connection, shown in Figure 25-4(a), and the *parallel* connection, shown in Figure 25-5(a). We often wish to determine the effective capacitance of the combination; that is, we wish to determine the capacitance of that single capacitor which has the same effect in the circuit as the combination of capacitors.

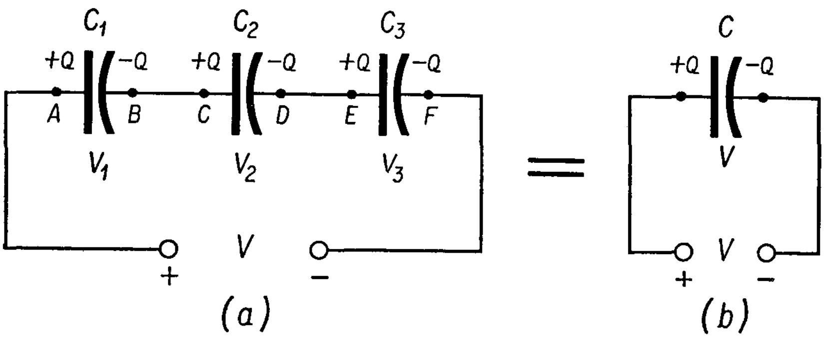


Fig. 25-4

Let us first consider the series connection of three capacitors C_1, C_2, C_3, as shown in Figure 25-4(a). When this combination of capacitors is connected to a source of electrical energy, such as a battery, electric charge flows from the source of energy to the plates of the capacitors until the plate A is at the potential of the positive terminal of the energy source, and the plate F is at the potential of the negative terminal of the energy source. The plate A bears a charge $+Q$, while the plate F bears a charge $-Q$. Lines of force emanating from plate A terminate on plate B. Thus if plate A

bears a positive charge $+Q$, an equal negative charge $-Q$ must appear on plate B, attracted by the electrical forces resulting from the charge on plate A.

The plate B, the plate C, and the wire BC connecting them may be considered as a single electrical conductor. This conductor was uncharged before the capacitors were connected to the source of electrical energy.

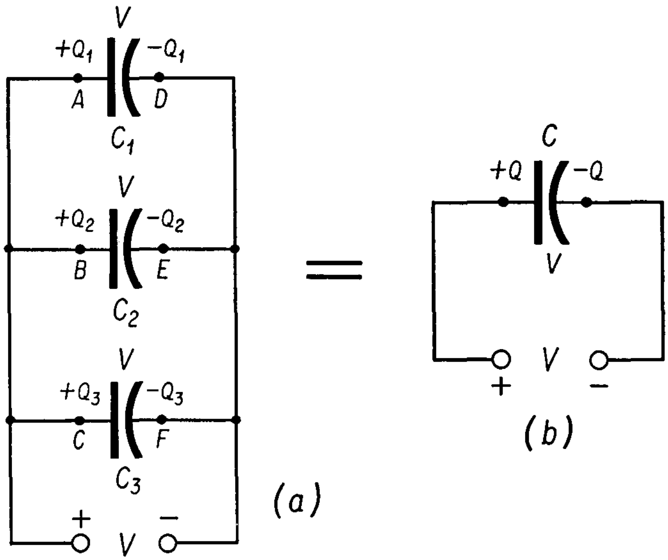


Fig. 25-5

Furthermore, this conductor is completely insulated, so that no electric charge can flow to or from the conductor. The only way a charge $-Q$ can appear on plate B is for an equal and opposite charge $+Q$ to appear on plate C. Continuing this argument for each of the capacitors of the series combination, we see that the charge on each capacitor is the same and is equal to Q. If Q_1 is the charge on capacitor C_1, Q_2 the charge on C_2, and Q_3 is the charge on C_3, we have

$$Q = Q_1 = Q_2 = Q_3.$$

The potential difference between the plates of each of the capacitors of the assembly can be found from Equation (25-1). Thus

$$V_1 = \frac{Q_1}{C_1} = \frac{Q}{C_1},$$

$$V_2 = \frac{Q}{C_2},$$

and
$$V_3 = \frac{Q}{C_3}.$$

The potential difference between the points A and F is equal to the work done in transporting a unit positive test charge from terminal A to terminal F against the electric forces. In carrying a test charge q from A to B, a quantity of work V_1q is done on the charge. Since there is electrostatic equilibrium, no work is done in moving the test charge along the wire BC, for this wire is a conductor and is an equipotential region. In transporting the test charge from plate C to plate D, the work done is V_2q, and in moving the charge from E to F the work done is V_3q. The total work done in moving the test charge from A to F may be represented by Vq, where V is the potential difference between A and F. Thus we have

$$Vq = V_1q + V_2q + V_3q,$$

or
$$V = V_1 + V_2 + V_3.$$

If we are to replace the series combination by a single capacitor of capacitance C having the same effect in the circuit as the series combination, as shown in Figure 25-4(b), a charge Q must flow to the plates of that capacitor when the potential between its plates is V. From Equation (25-1)

$$\begin{aligned}\frac{1}{C} &= \frac{V}{Q}\\ &= \frac{V_1 + V_2 + V_3}{Q}\\ &= \frac{Q/C_1 + Q/C_2 + Q/C_3}{Q}.\end{aligned}$$

Thus we find

$$\frac{1}{C} = \frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3}. \tag{25-6}$$

When any number of capacitors is connected in series, the reciprocal of the effective capacitance is equal to the sum of the reciprocals of the individual capacitances. The charge on the plates of each capacitor of the combination is the same, but different potential differences appear between the plates of the individual capacitors. Since practical capacitors are rated for maximum potential difference as well as for capacitance, it is necessary to exercise caution when connecting capacitors in series so as not to exceed the rated value of the potential difference of any one of the capacitors.

When capacitors are connected in parallel, as shown in Figure 25-5(a), one terminal of each capacitor is connected to one terminal of the source

of electrical energy. The other terminal of each capacitor is connected to the other terminal of the source of electrical energy. Thus the plates A, B, and C are all at one potential, while the plates D, E, and F are all at a different common potential. The potential difference between the plates of each capacitor is the same and is equal to the potential difference between the terminals of the source of electrical energy. Thus we have

$$V = V_1 = V_2 = V_3.$$

Unlike the series case, the charge on the plates of the individual capacitors is different, and we have

$$Q_1 = C_1V_1 = C_1V,$$

$$Q_2 = C_2V,$$

and

$$Q_3 = C_3V.$$

If we wish to replace the parallel combination of capacitors by a single capacitor of capacitance C, whose effect in the circuit is the same as the parallel combination, as shown in Figure 25-5(b), the plates of that capacitor should bear a charge equal to the charge which flows from the source of electrical energy to the parallel combination when the potential difference between the plates is V. Thus the charge on the equivalent capacitor of Figure 25-5(b) must be Q where

$$Q = Q_1 + Q_2 + Q_3.$$

Thus we have

$$CV = C_1V + C_2V + C_3V,$$

so that

$$C = C_1 + C_2 + C_3. \tag{25-7}$$

The effective capacitance of a combination of several capacitors connected in parallel is equal to the sum of the individual capacitances. The potential difference between the plates of each of the capacitors is the same, but different quantities of charge appear on the plates of the individual capacitors.

Illustrative Example. (a) Find the effective capacitance of the series-parallel combination of capacitors shown in Figure 25-6(a). (b) Find the charge on the 5-μfd capacitor when the potential difference between the points a and b is 300 volts.

(a) As a first step, let us find the equivalent capacitance of the parallel combination of the 3-μfd and 5-μfd capacitors. From Equation (25-7) we find that the effective capacitance of the combination is 8μfd. Let us imagine that the parallel combination is replaced by a single 8-μfd capacitor, as shown in Figure 25-6(b). The equivalent capacitance of the series combination of Figure 25-6(b) is given by Equation (25-6) as

$$\frac{1}{C} = \frac{1}{4\ \mu\text{fd}} + \frac{1}{8\ \mu\text{fd}},$$

or

$$C = \frac{8}{3}\,\mu\text{fd}.$$

Thus the entire series-parallel combination may be replaced by a single capacitor of $\frac{8}{3}$ μfd capacitance, as shown in Figure 25-6(c).

(b) The charge Q which appears on the single capacitor of Figure 25-6(c), when a potential difference of 300 volts exists between points a and b, is

$$Q = CV = \tfrac{8}{3}\ \mu\text{fd} \times 300\ \text{volts};$$

$$Q = 800\ \mu\text{coul}.$$

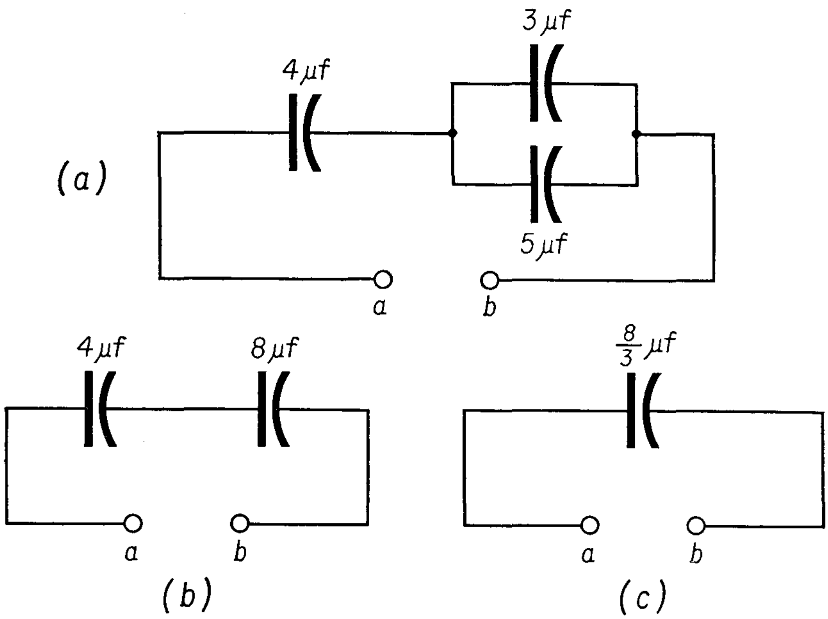


Fig. 25-6

The circuit of Figure 25-6(c) may now be replaced by the circuit of Figure 25-6(b), where a charge of 800 μcoul must appear on the plates of each of the two capacitors in series. The potential difference between the plates of the 8-μfd capacitor may be evaluated as

$$V_1 = \frac{800\ \mu\text{coul}}{8\ \mu\text{fd}} = 100\ \text{volts}.$$

The circuit of Figure 25-6(b) may now be replaced by the original circuit of Figure 25-6(a). A potential difference of 100 volts appears between the plates of each of the capacitors of the parallel combination. The potential difference between the plates of the 5-μfd capacitor is 100 volts, so that the charge on the plates of the 5-μfd capacitor is

$$Q_1 = C_1V_1 = 5\ \mu\text{fd} \times 100\ \text{volts};$$

so that

$$Q_1 = 500\ \mu\text{coul}.$$

It is now easily seen that the charge on the 3-μfd capacitor is 300 μcoul, and that the potential difference across the 4-μfd capacitor is 200 volts.

25-5 Dielectric Constant

Capacitors are commonly built with a solid or liquid insulating material, called a *dielectric,* placed between their plates. To understand the effect of a dielectric, suppose that the plates of a parallel-plate capacitor, with a vacuum between the plates, are charged to a potential difference V_{vac} and then disconnected from the source of charge; the charge on each plate will be Q. If a sheet of dielectric material is inserted between the plates, the potential difference will be found to decrease to some value V_{die}. Since the energy of a charge capacitor is $\frac{1}{2}QV$, the energy of the capacitor is decreased when a dielectric material is put between the plates of a charged capacitor. When the dielectric is removed, the potential difference returns to its initial value. This implies that there is an attractive force on the dielectric which draws it into the region of more intense electric field. Work is done by the electric field on the dielectric. When work is done by an external agency in removing the dielectric, the electrical energy of the capacitor is restored to its initial value.

From Equation (25-1) we see that the capacitance of a vacuum capacitor is less than the capacitance of the same pair of plates when a dielectric is inserted between them. If the capacitance in vacuum is C_{vac}, and the capacitance with the dielectric inserted in C_{die}, we have

$$C_{\text{vac}} = \frac{Q}{V_{\text{vac}}},$$

and

$$C_{\text{die}} = \frac{Q}{V_{\text{die}}}.$$

Let us suppose that the voltage V_{vac} is some number κ_e times the voltage of the capacitor with the dielectric between its plates. We may divide the second of these equations by the first to find

$$\frac{C_{\text{die}}}{C_{\text{vac}}} = \frac{V_{\text{vac}}}{V_{\text{die}}} = \kappa_e. \tag{25-8}$$

The ratio of $C_{\text{die}}/C_{\text{vac}}$, κ_e, is called the *dielectric constant.* It is also sometimes called the *dielectric coefficient,* the *relative permittivity,* or the *specific inductive capacity.* Experiment shows that κ_e is not truly a constant but varies somewhat with temperature, with the state of internal stress in the dielectric, and with other factors as well. By definition, the value of the dielectric constant is unity for vacuum. For air and most gases, the value deviates but slightly from unity, being 1.0006 for air at standard conditions of temperature and pressure. For most solids and liquids the dielectric constant ranges in value from about 1 to 100, although crystalline materials have been found with dielectric constants greater than 1,000.

We must be careful to distinguish between the *dielectric constant* and the *dielectric strength.* The dielectric constant affects the capacitance of a capacitor. The dielectric strength is the property which determines the maximum potential difference which can be imposed upon the plates of a capacitor without destroying its insulating properties. The dielectric constant and dielectric strength of several insulating materials are given in Table 25-1.

TABLE 25-1 PROPERTIES OF SOME DIELECTRIC MATERIALS

Material	Dielectric Constant (Relative Permittivity)	Dielectric Strength (kv/cm)
Vacuum	1	∞
Air (1 atm)	1.0006	30
Air (100 atm)	1.055	
Polystyrene	2.6	200–1,400
Polyethylene	2.3	160–240
Teflon	2.0	170–190
Paraffin wax	2.0–2.5	100
Porcelain	6.0–8.0	16–1,600
Glass	5–10	200–400
Rubber	3–6	160–480
Transformer oil	2.2	50–150
Water	55–88	
Barium titanate	$\sim$ 10,000	

The capacitance of a parallel-plate capacitor whose plates are separated by a medium of dielectric constant κ_e is given by the equation

$$C = \frac{\kappa_e \epsilon_0 A}{s} = \frac{\epsilon A}{s}. \tag{25-9}$$

In the mks system of units, the product $\kappa_e \epsilon_0$ is called the *permittivity* of the medium and is represented by ϵ; thus

$$\epsilon = \kappa_e \epsilon_0. \tag{25-10}$$

ϵ_0 is called the permittivity of free space, and κ_e is referred to as the relative permittivity.

Illustrative Example. A parallel-plate capacitor for use in a high-voltage circuit is made of sheets of tin foil plated on a large glass sheet whose dimensions are 2 m $\times$ 1.8 m $\times$ 0.02 m. Assuming the relative permittivity of the glass to be 8 and the dielectric strength of the glass to be 300 kv/cm, determine (a) the capacitance of this capacitor and (b) the maximum potential difference which may be applied to its plates.

(a) From Equation (25-9) we have

$$C = \frac{\kappa_e \epsilon_0 A}{s} = \frac{8 \times 8.85 \times 10^{-12}\ \text{farad/m} \times 3.6\ \text{m}^2}{0.02\ \text{m}},$$

$$C = 127 \times 10^{-10}\ \text{farad}.$$

(b) The dielectric strength of the glass is 300 kv/cm or 3×10^7 volts/m. The thickness of the glass is 0.02 m. The maximum potential difference which can be applied to the plates of the capacitor without destroying the insulating properties of the glass is given by

$$\frac{V}{s} = 3 \times 10^7 \frac{\text{volts}}{\text{m}},$$

so that

$$V = 3 \times 10^7 \frac{\text{volts}}{\text{m}} \times 0.02\ \text{m},$$

$$V = 6 \times 10^5\ \text{volts}.$$

25-6 The Material Medium; Polarization

The chemist describes the molecules of a substance by the term *polar* if the center of positive electricity within the molecule does not coincide with the center of negative electricity. If these two centers coincide, the molecule is called *nonpolar*. An ionic crystal such as sodium chloride is composed of positive sodium ions and negative chlorine ions and is highly polar. There are points within the substance where there is a high concentration of positive charge, while at other points there is a high concentration of negative charge. Molecular crystals, in which nonpolar molecules form the basic building blocks of the crystal, are generally only slightly polar or nonpolar.

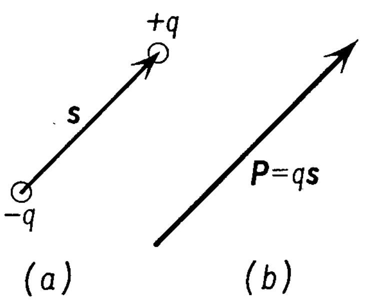


Fig. 25-7 (a) Electric dipole, and (b) dipole moment vector **P**.

The polarity of a molecule may be described by its *dipole moment*. If the displacement vector directed from the center of negative charge $-q$ to the center of positive charge $+q$ is **s**, as shown in Figure 25-7 the dipole moment **p** is given by the product of the charge by the displacement. Thus

$$\boxed{\mathbf{p} = q\mathbf{s}.} \tag{25-11}$$

The dipole moment is a vector quantity whose magnitude is the product

of the magnitude of the charge q by the distance s between the two charges, while the direction of the dipole-moment vector is from the negative to the positive charge.

A polar substance is made up of many dipoles. The substance is electrically neutral, for any macroscopic volume element, containing of the order of a few hundred atoms or molecules, contains an equal amount of positive and negative charge. When a polar substance is placed in an electric field, the positive charge experiences a force in the direction of the field, while the negative charge experiences a force in the opposite direction. The resultant torque tends to align the dipole parallel to the field, as shown in Figure 25-8. At room temperature the thermal energy of the molecules tends to disorient the dipole, while the elastic binding forces tend to restrain this rotation. Nevertheless, in an electric field a partial alignment does occur for many substances. In these the component of the dipole moment in the direction of the field, averaged over many molecules, is proportional to the electric field intensity.

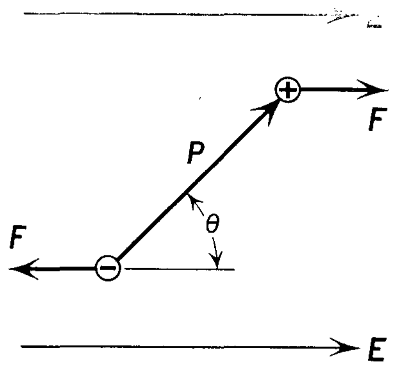


Fig. 25-8

Nonpolar substances do not possess a dipole moment in the absence of electric fields, but in the presence of an electric field the positive charges of a molecule experience forces in the direction of the field, while the negative electrons experience forces in the opposite direction so that a dipole moment may be induced. The *induced dipole moment* is parallel to the applied field. In many substances the magnitude of the induced dipole moment is proportional to the applied electric field. Thus in both polar and nonpolar substances the effect of an applied electric field is to increase the component of the polarization vector in the direction of the electric field. Rather than speaking of individual atoms and molecules, it is convenient to think of the aggregate effect produced by the electric field. We may speak of the total dipole moment per unit volume of the material. This is the vector sum of all the moments of all the elementary dipoles in a unit volume. *The dipole moment per unit volume is called the polarization*, designated by the symbol **P**. In the absence of an electric field, the dipoles of a substance are randomly oriented, so that the polarization is zero, but when the field is applied, the polarization is proportional to the electric intensity within the dielectric. The constant of proportionality relating the polarization **P** to the electric intensity **E** is called the *electric susceptibility* χ_e (chi sub e). In the form of an equation,

$$\mathbf{P} = \chi_e \mathbf{E}. \tag{25-12}$$

Let us analyze, in terms of the concept of polarization, the effect of placing a dielectric between the plates of a parallel-plate capacitor. Let us suppose that the plates of a parallel-plate capacitor have been charged in vacuum to a charge Q, as shown in Figure 25-9. When a dielectric is inserted between the plates of the capacitor, the polarization induced in the dielectric is parallel to the field. The polarization vector is directed from

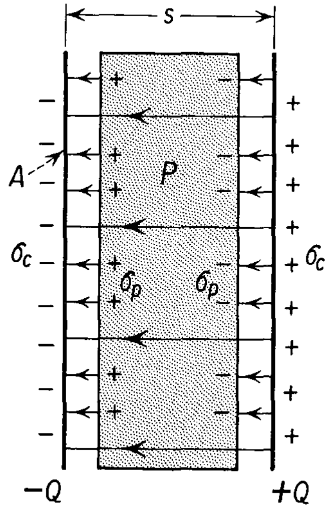


Fig. 25-9 The effect of placing a dielectric between the charged plates of a capacitor, thus polarizing the dielectric, may be most easily understood by imagining the dielectric to be replaced by two sheets of polarization charge located at the surfaces of the dielectric.

the negative to the positive charge of the dipoles induced in the dielectric. Since the positive and negative charges of a dipole are displaced through a comparatively small distance, the net effect is as though only the surfaces of the dielectric are charged with charges of opposite sign. It will therefore be convenient to imagine the dielectric to be replaced by two oppositely charged sheets of charge located at the surface of the dielectric. The imaginary sheet of negative charge is adjacent to the positively charged plate of the capacitor, while the imaginary sheet of positive charge is adjacent to the negatively charged plate of the capacitor. Let us call these imaginary sheets of charge the *polarization charge.*

Some of the lines of force emanating from the positively charged plate of the capacitor terminate on the adjacent sheet of negative polarization charge. The number of lines of force passing from the positive to the negative plate of the capacitor is therefore smaller in the presence of the dielectric than in vacuum. Since the number of lines of force per unit area is a measure of the electric field intensity, the electric field intensity within the dielectric is smaller than the electric field intensity in vacuum. The potential difference between the plates is therefore diminished, while the charge on the plates has remained the same. Thus the capacitance of the

parallel-plate structure has been increased by the insertion of the dielectric between the capacitor plates.

Let us represent the surface density of the imaginary polarization charge by the symbol σ_p, while the surface density of the *free charge* or *conduction charge* on the plates is represented by σ_c. The area of the capacitor plates is A, and their separation is s. The total polarization charge in one sheet is the product of the surface density of the polarization charge σ_p by the area of the plate A. The two sheets of polarization charge are separated by the distance s, so that the dipole moment due to the polarization charge is $\sigma_p As$. Let us now replace the polarization generated by the sheets of polarization charge by the induced polarization in the dielectric. The total dipole moment of the dielectric in which the dipole moment per unit volume is P is simply PAs; equating this to the dipole moment of the imaginary sheets of polarization charge, we get

$$PAS = \sigma_p As.$$

This yields $\quad P = \sigma_p. \qquad$ **(25-13)**

Hence the magnitude of the polarization or dipole moment per unit volume turns out to be equal to the magnitude of the surface density of the imaginary polarization charge. Substituting this value of P into Equation (25-12), we find

$$\sigma_p = \chi_e E. \qquad \textbf{(25-14)}$$

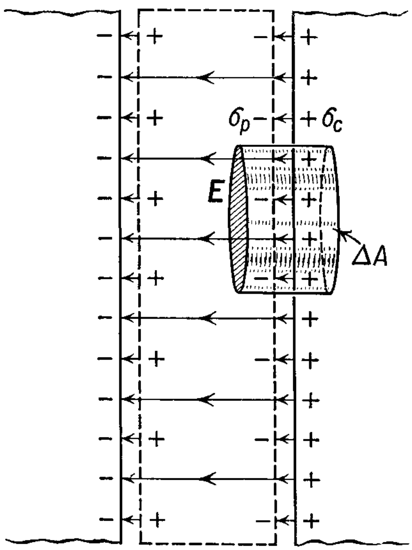


Fig. 25-10

To compute the electric intensity between the plates of the capacitor when there is a dielectric material between the plates, let us construct a Gaussian pillbox of unit surface area and of sufficient depth to penetrate the dielectric, as shown in Figure 25-10. The electric field intensity is given by Equation (23-11) as

$$E = \frac{\sigma}{\epsilon_0}\cdot$$

The total charge now consists of the conduction charge and the polarization charge. At the positive plate of the capacitor σ_c is positive, but the polarization charge σ_p is negative; hence

$$\sigma = \sigma_c - \sigma_p = \sigma_c - \chi_e E,$$

so that
$$E = \frac{(\sigma_c - \chi_e E)}{\epsilon_0}.$$

When this equation is solved for E, we obtain

$$E = \frac{1}{\epsilon_0}\frac{\sigma_c}{(1 + \chi_e/\epsilon_0)}. \tag{25-15}$$

Equation (25-15) is the expression for the electric field intensity between the plates of the capacitor containing a dielectric; hence it is the electric field intensity within the dielectric. Let us designate it by the symbol E_{die} and rewrite the equation as

$$E_{\text{die}} = \frac{\sigma_c/\epsilon_0}{1 + \chi_e/\epsilon_0}.$$

Let us compare this value with the electric field intensity when there is a vacuum between the plates. In the latter case, $\sigma_p = 0$, so that $\sigma = \sigma_c$; calling the electric field intensity now E_{vac}, we get

$$E_{\text{vac}} = \frac{\sigma_c}{\epsilon_0}.$$

Hence
$$\frac{E_{\text{die}}}{E_{\text{vac}}} = \frac{1}{1 + \chi_e/\epsilon_0}. \tag{25-16a}$$

We have already shown that

$$\frac{V_{\text{die}}}{V_{\text{vac}}} = \frac{1}{\kappa_e}. \tag{25-8}$$

Since $V = Es$ for a parallel-plate capacitor, we can write

$$\frac{E_{\text{die}}}{E_{\text{vac}}} = \frac{V_{\text{die}}}{V_{\text{vac}}} = \frac{1}{\kappa_e}. \tag{25-16b}$$

Equating the denominators of Equations (25-16a) and (25-16b) we get

$$\kappa_e = 1 + \frac{\chi_e}{\epsilon_0}. \tag{25-17}$$

Multiplying Equation (25-17) by ϵ_0, we find

$$\kappa_e\epsilon_0 = \epsilon_0 + \chi_e,$$

and from Equation (25-10)

$$\epsilon = \epsilon_0 + \chi_e. \tag{25-18}$$

The permittivity of the dielectric is made up of the sum of ϵ_0, the permittivity of free space, and χ_e, the susceptibility of the medium.

There must be atoms or molecules present if the dielectric constant is to differ from unity, or if ϵ is to differ from ϵ_0. If matter is in the gaseous phase between the plates of a capacitor, the number of molecules per unit volume is quite small, and even though the induced dipole moment of each molecule is comparatively large, the total dipole moment per unit volume, the polarization, must be small. Thus the dielectric constant of a gas must be close to 1. If the pressure of a gas is increased, or if the gas is liquefied, the number of molecules per unit volume increases by a large factor, perhaps even a factor of 1,000. The polarization is then increased, and the dielectric constant may now substantially differ from unity.

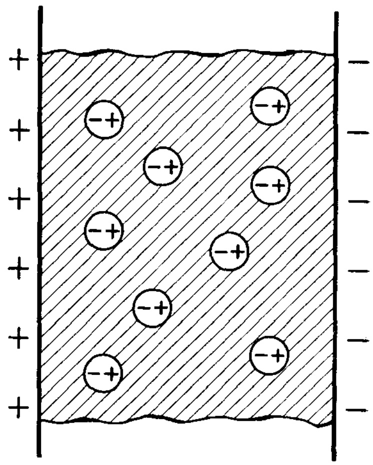


Fig. 25-11

It is possible to manufacture artificial dielectrics by distributing a large number of small conducting spheres in an insulator, as shown in Figure 25-11. A dipole is induced in each sphere by an electric field. If there are N such spheres in a unit volume, each sphere of radius a, it may be shown that the susceptibility of such a distribution is Na^3. Dispersions of conducting spheres in an insulating material have been used to produce artificial dielectrics for microwave lenses employed in high-frequency radio communication.

In the event that the space between a pair of capacitor plates is filled with a conducting substance, the electric field is reduced to zero. From Equations (25-15) and (25-17) we see that the electric field intensity between the plates of a capacitor is inversely proportional to the dielectric constant. Thus the dielectric constant of a perfect conductor is infinite.

25-7 Electric Field of a Point Charge in an Infinite Dielectric

When a point charge is placed in an infinite homogeneous dielectric, we can expect the electric field produced to be radial, from considerations of symmetry. To compute the effect of the dielectric on the field, let us imagine that the dielectric has a small spherical hole of radius a, centered on the point charge, as shown in Figure 25-12. Just as in the case of the parallel-plate capacitor, the electric field generated by the point charge will induce polarization in the dielectric whose effect can be calculated by imagining the dielectric to be replaced by a layer of polarization charge on

the surface of the spherical hole of charge density σ_p given by

$$\sigma_p = \chi_e E. \tag{25-14}$$

The total polarization charge q_p induced on the surface of the spherical hole is the product of the charge density and the surface area of the sphere and is negative in the neighborhood of a positive point charge, as given by

$$q_p = -4\pi a^2 \sigma_p = -4\pi a^2 \chi_e E_a. \tag{25-19}$$

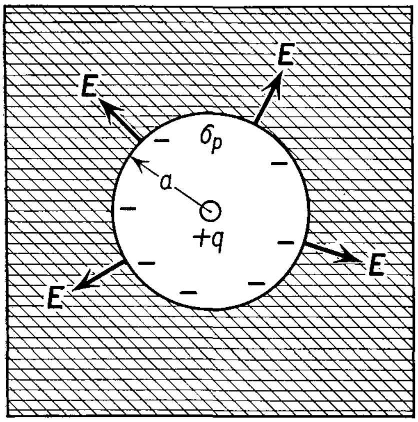


Fig. 25-12

Just as in the case of the parallel-plate capacitor, we may compute the electric field within the dielectric by considering that this electric field is generated by the original point charge and the uniformly charged shell of polarization charge. From Gauss's theorem we know that the field generated by a uniformly charged shell may be calculated as though the entire charge of the shell were concentrated at its center, for field points on or outside the shell. The electric field at the surface of the shell E_a is made up of the field generated by the original point conduction charge q_c and the field of the polarization charge q_p. Thus we have

$$E_a = \frac{q_c + q_p}{4\pi\epsilon_0 a^2}. \tag{25-20}$$

Substituting from Equation (25-20) into Equation (25-19), we find

$$q_p = -4\pi a^2 \chi_e \frac{q_c + q_p}{4\pi\epsilon_0 a^2},$$

so that

$$q_p = -q_c \frac{\chi_e/\epsilon_0}{1 + \chi_e/\epsilon_0}. \tag{25-21}$$

In Equation (25-21) we see that the value of the polarization charge q_p does not depend upon the radius of the spherical hole. We may therefore shrink the cavity down to infinitesimal radius and imagine the polarization charge to be located at the position of the point charge.

The results of the preceding discussion are such as to indicate that the electric field within the dielectric may be found by replacing the dielectric by a point polarization charge q_p, of magnitude given by Equation (25-21),

located at the position of the free charge q_c. The electric field intensity within the dielectric is therefore

$$E = \frac{q_c + q_p}{4\pi\epsilon_0 r^2},$$

$$E = \frac{q_c}{4\pi\epsilon_0 r^2}\left(1 - \frac{\chi_e/\epsilon_0}{1 + \chi_e/\epsilon_0}\right),$$

$$E = \frac{q_c}{4\pi\epsilon_0(1 + \chi_e/\epsilon_0)r^2}\cdot$$

We recall from Equation (25-17) that the quantity in the parentheses in the above equation is equal to the dielectric constant κ_e, so that the electric field of a point charge embedded in an infinite dielectric may be expressed as

$$E = \frac{q_c}{4\pi\epsilon r^2}\cdot \tag{25-22}$$

Since q_c is simply the original charge placed in the dielectric, we may drop the subscript c from q_c, for there is no longer any need to distinguish it from the polarization charge. Rewriting Equation (25-22) in vector form, recalling that the field generated by a point charge is radial, we obtain

$$\mathbf{E} = \frac{q}{4\pi\epsilon r^2}\,\mathbf{1}_r. \tag{25-22a}$$

If two point charges are embedded in an infinite dielectric, the force exerted by the charge q_1 on the charge q_2 may be obtained from Equation (25-22a) as

$$\boxed{\mathbf{F} = \frac{q_1 q_2}{4\pi\epsilon r^2}\,\mathbf{1}_r.} \tag{25-23}$$

This is the form of Coulomb's law applicable to charges placed in an infinite dielectric medium of permittivity ϵ.

25-8 The Electric Displacement

In dealing with problems associated with electric fields in dielectrics, it is convenient to introduce an auxiliary vector called the *electric displacement*, represented by the symbol $\mathbf{D}$, a concept introduced into the study of electricity by James Clerk Maxwell (1831–1879). We may define $\mathbf{D}$ by the equation

$$\boxed{\mathbf{D} = \epsilon_0\mathbf{E} + \mathbf{P}.} \tag{25-24}$$

The concept of lines of force was very useful for describing the electric field in vacuum. From this concept we were able to develop Gauss's theorem and to obtain considerable insight into the structure of electric fields. When we came to face the problem of the dielectric medium, the calculation of electric fields became much more difficult, and we were forced to invent the construct of a polarization charge.

The displacement vector is used in electrostatics for the purpose of recovering the convenience of lines of force when dielectric media are involved. Instead of lines of force, it is possible to describe the electric field by *lines of electric displacement.*

If we take account of the polarization charge q_p on the surface of a dielectric which is enclosed within a Gaussian surface, Gauss's law becomes

$$\int \mathbf{E}\cdot d\mathbf{A} = \frac{q_p}{\epsilon_0} + \frac{q}{\epsilon_0},$$

where q is the original conduction charge enclosed within the Gaussian surface. The total polarization charge q_p on the surface of the dielectric may be represented as

$$q_p = -\int \mathbf{P}\cdot d\mathbf{A},$$

for the polarization $\mathbf{P}$ is a vector quantity. Substituting into the above equation, transposing, and multiplying through by ϵ_0, we obtain

$$\int (\epsilon_0\mathbf{E} + \mathbf{P})\cdot d\mathbf{A} = q.$$

From the definition of $\mathbf{D}$ given in Equation (25-24), Gauss's law takes the form

$$\boxed{\int \mathbf{D}\cdot d\mathbf{A} = q.} \tag{25-25}$$

Thus in the mks system of units we may assert that one D line emerges from each real, positive, conduction charge of 1 coul. In this system of units, D is expressed in coulombs per square meter, for consistency with Equation (25-25).

In the case of the parallel-plate capacitor with a dielectric we saw how some of the lines of force, or E lines, originating on the free or conduction charges, terminated on polarization charges on the surface of the dielectric. If we add to the E lines within the dielectric additional lines associated with the polarization, as indicated by Equation (25-24), the resultant lines are lines of electric displacement, or D lines. We find that the D lines are continuous from one plate of the capacitor, through the dielectric, to the

other plate. Thus the D lines originate on positive conduction charges and terminate on negative conduction charges, and are continuous on passing through a dielectric. They do not terminate abruptly on passing into a dielectric.

Having obtained a solution to the distribution of D lines in space for a particular charge distribution, we must still transcribe that solution so that it is stated in terms of the electric field intensity E. The transcription from **D** to **E** is especially simple for *homogeneous isotropic dielectrics*, such as we have been discussing. If we substitute from Equation (25-12) into Equation (25-24), we find

$$\mathbf{D} = \epsilon_0 \mathbf{E} + \mathbf{P} = (\epsilon_0 + \chi_e)\mathbf{E}.$$

From Equation (25-18)

$$\epsilon_0 + \chi_e = \epsilon,$$

so that

$$\boxed{\mathbf{D} = \epsilon \mathbf{E}.} \tag{25-26}$$

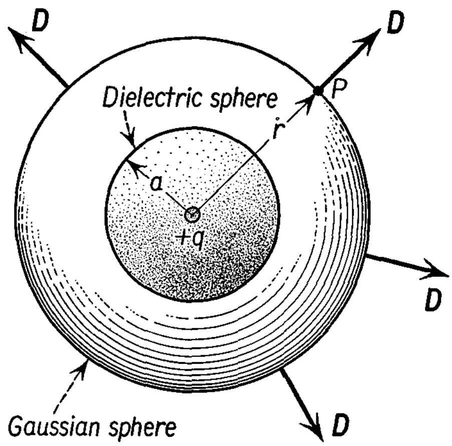


Fig. 25-13

Illustrative Example. Find the electric intensity at a distance r from a point charge q embedded at the center of a dielectric sphere of radius a, as shown in Figure 25-13.

From symmetry, the D lines are radial and continuous. If we draw a Gaussian sphere of radius r whose center is at q, we may apply Gauss's theorem in the form of Equation (25-25) to find

$$4\pi r^2 D = q,$$

so that

$$D = \frac{q}{4\pi r^2}.$$

TABLE 25-2 PRINCIPAL EQUATIONS IN MKS AND CGS UNITS

Equation	MKS	CGS	
(25-1)	$C = Q/V$	Same as mks	Capacitance
(25-2)	$C = 4\pi\epsilon_0 a$	$C = a$	Sphere
(25-3)	$C = \dfrac{\epsilon_0 A}{s}$	$C = \dfrac{A}{4\pi s}$	Parallel-plate capacitor in vacuum
(25-4a)	$\mathscr{W} = \dfrac{1}{2}\dfrac{Q^2}{C}$	Same as mks	Energy of a charged capacitor
(25-5)	$\mathscr{W}_v = \dfrac{\epsilon_0 E^2}{2}$	$\mathscr{W}_v = \dfrac{E^2}{8\pi}$	Energy density of electric field in vacuum
(25-6)	$\dfrac{1}{C} = \sum \dfrac{1}{C_i}$	Same as mks	Series capacitors
(25-7)	$C = \sum C_i$	Same as mks	Parallel capacitors
(25-9)	$C = \dfrac{\epsilon A}{s}$	$C = \dfrac{\kappa_e A}{4\pi s}$	Parallel-plate capacitor with dielectric
(25-11)	$\mathbf{p} = q\mathbf{s}$	Same as mks	Dipole moment
(25-12)	$\mathbf{P} = \chi_e \mathbf{E}$	Same as mks	Susceptibility
(25-17)	$\kappa_e = 1 + \chi_e/\epsilon_0$	$\kappa_e = 1 + 4\pi\chi_e$	Dielectric constant
(25-18)	$\epsilon = \epsilon_0 + \chi_e$	No analogous equation	
(25-22a)	$\mathbf{E} = \dfrac{q}{4\pi\epsilon r^2}\mathbf{1}_r$	$\mathbf{E} = \dfrac{q}{\kappa_e r^2}\mathbf{1}_r$	Point charge in an infinite dielectric
(25-24)	$\mathbf{D} = \epsilon_0 \mathbf{E} + \mathbf{P}$	$\mathbf{D} = \mathbf{E} + 4\pi\mathbf{P}$	Displacement
(25-25)	$\int \mathbf{D} \cdot d\mathbf{A} = q$	$\int \mathbf{D} \cdot d\mathbf{A} = 4\pi q$	Gauss's theorem
(25-26)	$\mathbf{D} = \epsilon \mathbf{E}$	$\mathbf{D} = \kappa_e \mathbf{E}$	Displacement

When r is greater than a, the field point is in vacuum, so that

$$D = \epsilon_0 E,$$

and

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

When r is less than a, the field point is within the dielectric, so that

$$D = \epsilon E,$$

and

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

The distribution of lines of electric displacement for the case of a

dielectric sphere placed in a uniform electric field is shown in Figure 25-14. Problems of this sort may be solved by mathematical methods beyond the

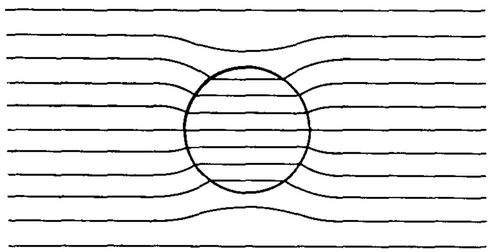


Fig. 25-14 Lines of electric displacement D for a dielectric sphere in a uniform electric field.

scope of this book through the application of the basic concepts developed here. The electric displacement is an important concept for describing the electric field in any case involving the use of insulating materials, and therefore in all practical engineering design. We shall see in subsequent chapters that the concepts developed here for the treatment of the electric field in the material medium will be of value in our study of the magnetic field.

TABLE 25-3 CONVERSION FACTORS RELATING MKS AND CGS UNITS

Quantity	Symbol	MKS Unit	CGS Unit	
Capacitance	C	1 farad	$= 9 \times 10^{11}$ stfd	(esu)
Displacement	D	1 coul/m^2	$= 3 \times 10^{5}$ stcoul/cm^2	(esu)
Dipole moment	p	1 coul m	$= 3 \times 10^{11}$ stcoul cm	(esu)
Polarization	P	1 coul/m^2	$= 3 \times 10^{5}$ stcoul/cm^2	(esu)
Susceptibility	χ_e	$1\,\dfrac{\text{coul}^2}{\text{joule m}}$	$= 9 \times 10^{9}\left(\dfrac{\text{stcoul}^2}{\text{erg cm}}\right)^*$	(esu)
Potential	V	1 volt	$= 300$ statvolts	(esu)
Charge	Q	1 coul	$= 3 \times 10^{9}$ stcoul	(esu)
Electric intensity	E	1 volt/m	$= \dfrac{1}{3 \times 10^4}$ statvolt/m	(esu)

Permittivity of free space: $\epsilon_0 = 8.85 \times 10^{-12}\,\dfrac{\text{farad}}{\text{m}} = 8.85 \times 10^{-12}\,\dfrac{\text{coul}^2}{\text{joule m}}$.

*The electric susceptibility is dimensionless in the cgs electrostatic system of units. The dimensions in the parentheses cancel; they are included in the table to facilitate conversion between the two systems of units.

Problems

25-1. What is the capacitance in microfarads of a sphere 150 cm in radius (a) in air? (b) Immersed in an infinite bath of oil of dielectric constant 1.5?

25-2. What are the upper and lower limits for the capacitance of a charges conducting cube 1 m on an edge, in vacuum?

25-3. Show that the units of ϵ_0 in farads per meter are equivalent to unitd of $\text{coul}^2/\text{nt m}^2$.

25-4 A parallel-plate capacitor whose plates are 10 cm^2 in area has a charge of 10^{-9} coul. The potential difference between the plates is 100 volts. What is the separation between the plates of the capacitor?

25-5. A capacitor of 2 μfd capacitance is charged until the difference of potential between its plates is 120 volts. (a) Determine the charge on the capacitor. (b) Determine the amount of work done in charging it.

25-6. Two parallel-plate capacitors of identical dimensions differ only in that one has air between the plates and the other has oil of dielectric constant 2 between the plates. The capacitance of the air capacitor is 175 μfd. (a) What is the capacitance of the oil capacitor? (b) Each capacitor is charged to a potential difference of 30 volts. Determine the charge on each capacitor. (c) Determine the energy of each capacitor.

25-7. Each of the two plates of a parallel-plate capacitor has an area of 400 cm^2. The plates are 2 mm apart in vacuum. (a) Determine the capacitance in statfarads. (b) If a potential difference of 125 statvolts is maintained between the plates, determine the charge on each plate. (c) Determine the energy of this charged capacitor.

25-8. The plates of a parallel-plate capacitor are arranged so that the distance between them can be varied. When the distance between them is d, the capacitor is charged until the difference of potential is V. The plates are then separated until the distance between them is $2d$. Assuming that the charge Q on the plates is unchanged, determine (a) the difference of potential between the plates of this capacitor, (b) the change in energy due to the increase in the distance between the plates, and (c) the work done in separating the plates.

25-9. Derive a formula for the capacitance between a pair of concentric conducting spheres of radii r_1 and r_2, in vacuum.

25-10. Derive a formula for the capacitance per unit of length of a pair of coaxial cylinders in vacuum. (See Problem 23-14.)

25-11. Two capacitors, one of 3 μfd capacitance and the other of 5 μfd capacitance, are connected in parallel and charged until the potential difference is 100 volts. Determine (a) the charge on each capacitor, (b) the equivalent capacitance of the system, and (c) the energy of this system.

25-12. Two capacitors, one of 4 μfd capacitance and the other of 6 μfd capacitance, are connected in series and charged to a difference of potential of 120 volts. Determine (a) the equivalent capacitance of the combination, (b) the charge on each capacitor, (c) the potential difference across each capacitor, and (d) the energy of the system.

25-13. A capacitor of 4 μfd capacitance has a charge of 40 μcoul, and a capacitor of 3 μfd capacitance has a charge of 10 μcoul. The negative plate of each one is connected to the positive plate of the other. Determine (a) the charge on each capacitor and (b) the potential difference across each one.

25-14. If the charged capacitors of Problem 25-13 are connected so that plates of like charge are connected together, determine (a) the initial potential

difference across each capacitor, (b) the final potential difference across each capacitor, (c) the initial energy of each capacitor, (d) the final energy of the combination, and (e) the energy lost in connecting them together.

25-15. Two capacitors of 3 and 4 μfd are connected in parallel, and the combination is connected in series to a third capacitor of 5 μfd. Determine (a) the effective capacitance of the combination, (b) the potential difference across the 3-μfd capacitor when 100 volts are applied across the entire combination, and (c) the charge on the 4-μfd capacitor under these circumstances.

25-16. It is desired to construct a parallel-plate capacitor of aluminum foil and polystyrene sheeting 5 mils thick (1 mil = 0.001 in.). The dielectric strength of polystyrene is 2,500 volts/mil, and the dielectric constant of this material is 2.5. The capacitor is to be able to withstand a maximum voltage of 100,000 volts. How large an area of plate will be required to yield a capacitance of 10 μfd?

25-17. What is the dipole moment of a pair of charges of opposite sign of 5 stcoul separated by a distance of 2 cm? State the units as well as the magnitude.

25-18. A parallel-plate capacitor, whose plates are 0.5 m^2 in area and are separated by 0.01 m, has a potential difference of 100 volts across its plates. The plates are separated by a dielectric whose dielectric constant is 3. Determine (a) the susceptibility of this dielectric, (b) the polarization of the dielectric, and (c) the polarization charge density at the surface of the dielectric.

25-19. Two point charges of 5 stcoul and -10 stcoul lie along the x axis separated by a distance of 5 cm. The charges are immersed in insulating oil of dielectric constant 2.5. (a) Find the attractive force between them. (b) Find the electric field intensity at a point 3 cm from the 5-stcoul charge and 4 cm from the -10-stcoul charge.

25-20. A parallel-plate capacitor whose plates are separated by a distance s has a sheet of dielectric constant 2.5 and thickness $0.95s$ inserted between its plates, the remainder of the space between the plates being filled by air of dielectric constant 1. A potential difference V is applied to the plates. (a) What is the electric field intensity in the air? (b) What is the electric field intensity in the dielectric? (c) How does this compare with the electric field intensity in an identical capacitor in which the space between the plates is completely filled by air or by dielectric? (Solve by use of the displacement **D**.)

25-21. By use of a parallel-plate capacitor filled with a dielectric, find the energy per unit volume of the electric field in a medium of dielectric constant κ_e.

26

The Electric Current

26-1 Sources of Electric Energy

Electricity is the basis of our present highly technical civilization. It is the means whereby energy from various sources is delivered to the consumer in a form suitable for its conversion into the many kinds of energy demanded by him. It is through the intermediary of the electric circuit that energy is transmitted electrically from the primary source, no matter where it is situated, to the ultimate consumer wherever he desires it. A complete electric circuit contains some form of electric generator, which is essentially a device for converting other forms of energy to electrical energy, a set of conductors for transmitting the electrical energy, and some device for converting this electrical energy into the desired form of energy.

There are various types of electric generators. We have already mentioned the electrostatic induction machine, of which the electrophorus is a crude example. More common is the chemical generator, of which there are two general types: (a) the primary chemical cell, in which a potential difference is developed by means of chemical action among some of the substances composing the cell (the so-called "dry cell" is a common form of primary chemical cell) and (b) the secondary chemical cell or *storage cell*, which must first be charged by sending electricity through it from some other generator, after which it acts just like a primary cell. The storage *battery* used in cars and on farms consists of several storage cells connected together. As a result of the chemical action which takes place in each of these cells, chemical energy is converted to electrical energy.

For the generation of large amounts of electric energy, the *dynamo* is used. The dynamo is driven by some kind of engine such as a steam engine, gasoline engine, or water turbine. Essentially, the dynamo converts mechanical energy into electric energy.

The chemical cell and the dynamo, as well as other forms of electric generators, will be discussed in greater detail in later chapters. In the present chapter we shall make use of electric generators, particularly the

d-c or *direct-current* generator, to study the properties of the electric circuit, but we shall reserve for a later chapter the discussion of how the generator operates. A d-c generator has two terminals, one of which is at a higher potential than the other. The terminal at the higher potential is called the *positive terminal*, and the other is called the *negative terminal.*

26-2 Current

Electric charge in motion constitutes an *electric current.* In the steady flow of charge in a wire of cross-sectional area A, as shown in Figure 26-1,

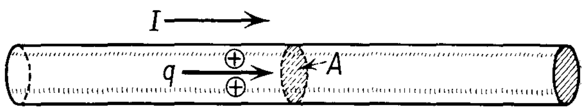


Fig. 26-1 A current in a conductor consists of the flow of charges through any cross-sectional area A.

the total charge passing through this area in unit time is defined to be the electric current I at this place. If a total charge q flows through this area in time t, the current I is given by

$$\boxed{I = \frac{q}{t}\,.} \tag{26-1}$$

If q is expressed in statcoulombs and t in seconds, the current I is in cgs electrostatic units called *statamperes.* If q is expressed in coulombs and t in seconds, the current I is expressed in mks units called *amperes* (abbreviated amp). Thus an ampere is a coulomb per second, and 1 amp is approximately equal to 3×10^9 statamperes. The ampere is named for André Marie Ampère (1775–1836), a pioneer electrical scientist.

The direction of the current is defined as the direction in which a positive charge would move. The nature of the charges whose motion constitutes the current depends upon the nature of the conducting substance. If the conductor is a metal, the current consists of the motion of *free electrons.* In a gas the charges which are set in motion are *positive* and *negative ions,* and, under conditions of low pressure, there may be electrons as well as the ions. In nonmetallic liquid conductors, such as electrolytes, the current consists of the motion of positive and negative ions. The positive charges move in the direction of the current, while the negative ions move in the opposite direction, as shown in Figure 26-2. In a conductor, flow of negative charge to the left is equivalent to a flow of positive charge to the

right, so that a current flowing to the right in the figure may be made up of positive charges flowing to the right and of negative charges flowing to the left.

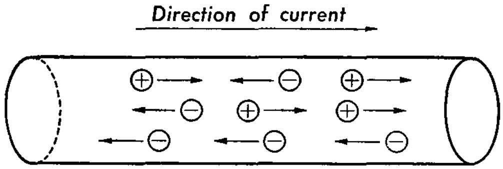


Fig. 26-2 The direction of the current is the direction of motion of the positive charges. Negative charges move in a direction opposite to that of the current.

Unless specifically stated to the contrary, the equations dealing with electric currents are based upon the convention that the direction of the current is the direction of flow of positive charge.

When the flow of charge is not uniform, we may define the instantaneous current in terms of the rate of flow of charge. Thus we have

$$I = \frac{dq}{dt}. \tag{26-2}$$

26-3 Current Density

In dealing with the flow of electricity in a continuous medium, it is convenient to speak of the *current density* **J**. *The current density is the quantity of charge passing through a unit area, perpendicular to the direction of motion of the charge, in a unit of time.* The direction of the current-density vector **J** is the direction of motion of positive charge and hence of the current, and is opposite to the direction of motion of negative charge. In the case of the current in a wire such as that of Figure 26-1, the magnitude of the average current density is given by

$$J = \frac{I}{A}, \tag{26-3}$$

and the direction of the current density is along the wire in the direction of current.

We may represent the steady flow of current through a conductor of variable cross section by flow lines similar to the streamlines of fluid flow. Such a line of electric flow would then represent the path taken by charged particles, and the number of flow lines passing through a surface of unit area perpendicular to the direction of flow would represent the current density. As in the case of the streamline flow of a fluid, electric flow lines never cross. In the steady state the current passing any point in a conductor is constant. Thus the current density in a conductor of nonuniform

cross section, as shown in Figure 26-3, is inversely proportional to the cross-sectional area. At the points 1 and 2 in the conductor we have

$$I_1 = I_2,$$

so that

$$J_1A_1 = J_2A_2.$$

Thus we have

$$\frac{J_1}{J_2} = \frac{A_2}{A_1}. \tag{26-4}$$

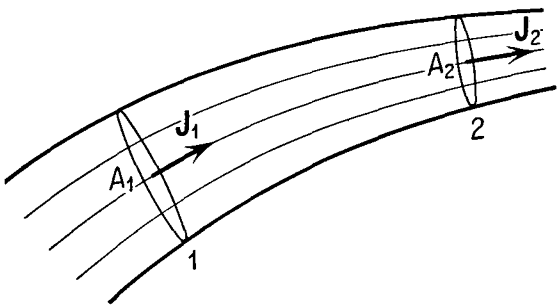


Fig. 26-3

This analogy between electric flow lines and streamlines in a fluid is of considerable practical use in the solution of aerodynamic flow problems through the use of *electrolytic plotting tanks.*

Let us suppose that a positively charged cloud containing n particles per unit volume, each of charge q, is moving with velocity v, as shown in

Fig. 26-4

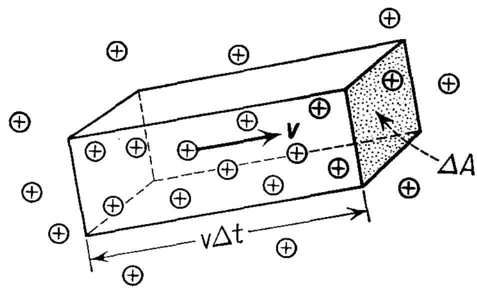


Figure 26-4. All of the charged particles contained in a parallelpiped of base ΔA and of altitude $v\ \Delta t$ will pass through an element of area ΔA perpendicular to the direction of flow in the time Δt. Thus the current density J is given by

$$J = \frac{(nq)(v\ \Delta t)(\Delta A)}{\Delta t\ \Delta A},$$

or

$$J = nqv. \tag{26-5}$$

Since the direction of the current density is the direction of motion of a positively charged particle, we may rewrite Equation (26-5) in vector form as

$$\mathbf{J} = nq\mathbf{v}. \tag{26-5a}$$

From Equation (26-5a) we see that if the charged particles are negatively charges, a negative number must be substituted for q, in which case $\mathbf{J}$ is in the direction of $-\mathbf{v}$.

26-4 Ohm's Law

The current in a conductor of electricity is proportional to the difference of potential between the ends of the conductor. This experimental observation was made by Georg Simon Ohm (1787–1854) in 1826 and is today known as Ohm's law. The constant of proportionality relating the current I to the potential difference V is called the *conductance* G, so that we may write Ohm's law in the form of an equation as

$$I = GV. \tag{26-6}$$

The same proportionality is often expressed in terms of a second constant called the resistance R as

$$\boxed{V = IR.} \tag{26-7}$$

When the current is expressed in amperes and the potential difference in volts, the resistance is expressed in *ohms.* From Equations (26-6) and (26-7) the conductance is the reciprocal of the resistance; that is,

$$G = \frac{1}{R}. \tag{26-8}$$

The units of conductance are reciprocal ohms or *mhos.* The cgs system of units is rarely used in connection with the current through a conductor; cgs units of conductance and resistance are called statmhos and statohms, respectively.

Ohm's law in the form of Equation (26-7) is simultaneously a very concise statement about the properties of a great many conductors and a definition of the resistance of a conductor. In a strict sense the term "resistance" should not be used for a conductor unless there is a fixed constant value of R for that conductor, when it is maintained at a fixed temperature, for all values of the potential difference across it. For metals, Ohm's law is true over an extremely large range of values of V separated by factors of 10^{12} or more. The term "resistance" has sometimes been extended to mean the quotient of the voltage across the terminals of a

conductor divided by the current through the conductor without regard to whether this quotient is constant for a particular temperature.

Resistance elements, called *resistors*, are constructed for use in electrical circuits. These are schematically represented by the symbols shown in Figure 26-5. The symbol commonly used for abbreviating the word "ohm"

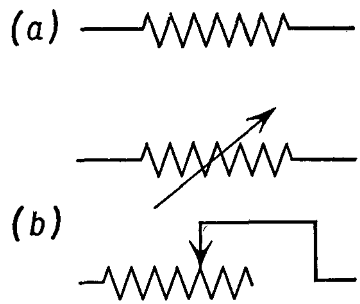


Fig. 26-5 Schematic symbols for (a) fixed resistor and (b) variable resistor.

is the greek capital omega Ω. Following the usual system of notation for multiples in the metric system, large resistances are generally represented in *megohms,* while small resistances are represented as *microhms.*

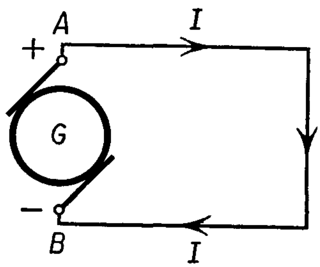


Fig. 26-6 Direction of the current in a circuit.

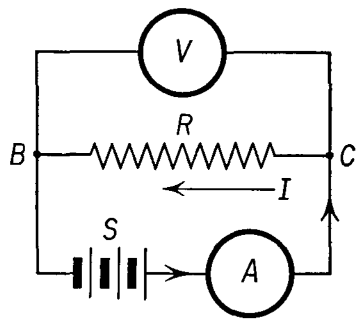


Fig. 26-7 Voltmeter-ammeter method for measuring the resistance of a conductor.

We have previously shown that when a positive charge moves under the influence of an electric field, its direction of motion is from a point of higher potential to a point of lower potential. If two such points, say the terminals A and B of a generator, are connected by a conductor, as shown in Figure 26-6, the direction of the current in the conductor will be from A, the point at the higher potential, to B, the point at the lower potential.

A simple way of measuring the resistance of a wire is illustrated in Figure 26-7. The wire BC whose resistance R is to be measured is connected to a storage battery S and an ammeter A. The voltmeter V is

connected across the terminals of the wire BC. In the illustration the various elements of the electric circuit are represented schematically in a conventional manner. The longer of the two strokes representing each of the cells of the storage battery represents the positive terminal. A conductor having negligible resistance is represented by a straight line. The reading of the ammeter gives the current I in the wire BC, assuming that the current through the voltmeter is negligible, while the voltmeter reading gives the difference of potential V between B and C. This is sometimes referred to as the *voltage* across BC. The resistance R of the wire is then calculated from Ohm's law

$$R = \frac{V}{I}.$$

26-5 Electrical Energy

When a particle of charge q is displaced through a difference of potential V, an amount of work $\mathscr{W}$ is performed such that

$$\mathscr{W} = qV. \tag{26-9}$$

When the work is done by the electric field on the charged particle, that work may appear as a difference in kinetic energy of the charged particle, or, in the event that the charged particle moves with constant speed as a result of the action of some type of resisting force opposing the motion, the work $\mathscr{W}$ may appear as heat. Equation (26-9) may be expressed in terms of the current I instead of the charge q. For the case of a constant current, we get from Equation (26-1)

$$q = It;$$

hence

$$\mathscr{W} = VIt. \tag{26-10}$$

Dividing both sides of Equation (26-10) by the time, we find

$$\frac{\mathscr{W}}{t} = VI,$$

and since the power $\mathscr{P}$ is equal to the work divided by the time, we have

$$\mathscr{P} = VI. \tag{26-11}$$

It is essential that consistent units be used in these equations; when mks units of electrical quantities are used, the mks units of work, the joule, and of power, the watt, must be used. For example, when a current of 5 amp flows between two points whose potential difference is 110 volts, the electrical power supplied is 550 watts.

Comparing Equations (26-7) and (26-11), we obtain the power $\mathscr{P}$ supplied to a conductor of resistance R, through which a current I flows

when a potential difference V is applied across its terminals as

$$\mathcal{P} = I^2R, \tag{26-12a}$$

or

$$\mathcal{P} = \frac{V^2}{R}. \tag{26-12b}$$

The power supplied to an electrical conductor when current flows through it appears in the form of heat and possibly light. Among the many experiments which Joule performed to determine the mechanical equivalent of heat, there were some which involved the transformation of electrical

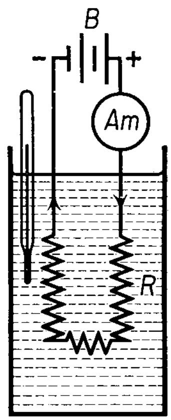


Fig. 26-8 Transformation of electric energy into heat; heating a quantity of water by means of an electric current in a wire of resistance R immersed in it. Am is an ammeter, B is a battery.

energy into heat. A coil of wire was immersed in water of known mass in a calorimeter, as shown in Figure 26-8. Joule found that the rate at which electrical energy was converted into heat in a conductor was proportional to the square of the current flowing through the conductor, a result which was important in establishing the universality of the concept of energy. We have used the energy concept as a means of deriving Joule's experimental result from Ohm's law in Equation (26-12a), which is sometimes called *Joule's law*.

Illustrative Example. The heating coil of an electric iron operating on a 110-volt line has a current of 5 amp in it. Determine (a) the resistance of the coil, (b) the power supplied to it, and (c) the amount of heat flowing out of the coil in 4 min.

(a) The resistance of the coil is, from Equation (26-7),

$$R = \frac{V}{I} = \frac{110 \text{ volts}}{5 \text{ amp}} = 22\Omega.$$

(b) The power supplied to the coil is, from Equation (26-11),

$$\mathcal{P} = VI = 110 \text{ volts} \times 5 \text{ amp} = 550 \text{ watts},$$

or, from Equation (26-12a),

$$\mathcal{P} = I^2R = 25\ \text{amp}^2 \times 22\Omega = 550\ \text{watts}.$$

(c) Since all of the electrical energy $\mathcal{W}$ is converted into heat, we can write

$$\mathcal{W} = \mathcal{P}t = 550\ \text{watts} \times 240\ \text{sec}$$
$$= 132{,}000\ \text{joules},$$

or, using the mechanical equivalent of heat

$$4.2\ \text{joules} = 1\ \text{cal},$$

we get

$$\mathcal{W} = \frac{132{,}000}{4.2}\ \text{cal} = 31{,}400\ \text{cal}.$$

26-6 Resistivity and Conductivity

To describe the properties of a continuous medium, we may rewrite Ohm's law in terms of the conductivity σ and the resistivity ρ rather than in terms of the conductance G and the resistance R which were used for discrete conductors. If the electric field intensity within a conductor is E and the current density is J, we may write Ohm's law for a continuous medium as

$$J = \sigma E \tag{26-13a}$$

or as

$$E = \rho J, \tag{26-13b}$$

where σ is the reciprocal of ρ; thus

$$\sigma = \frac{1}{\rho}\cdot \tag{26-14}$$

Although both Equations (26-13a) and (26-13b) have been written in scalar form, we recognize that the vector $\mathbf{J}$ is parallel to the vector $\mathbf{E}$ and that σ and ρ are scalar constants relating the current density to the electric field intensity, in a homogeneous isotropic conductor. From Equations (26-13) and (26-14) the units of σ in mks units are *mhos per meter*, while the units of ρ are *ohms meter*.

To find the relation between conductance and conductivity, let us consider a uniform wire of cross-sectional area A and length s in which there is a current I, as shown in Figure 26-9. From Equation (26-3) the current density J in the wire is given by

$$J = \frac{I}{A}\cdot$$

If the potential difference between the ends of the wire is V, the end a being at the higher potential, the electric field intensity within the wire is given by

$$E = \frac{V}{s}$$

TABLE 26-1 RESISTIVITY OF CONDUCTING MATERIALS

Substance	Resistivity in ohm-cm at 20°C	Temperature Coefficient per °C at 20° C
Aluminum	2.83×10^{-6}	0.0039
Carbon	3.5×10^{-3}	−0.0005
Copper	1.72×10^{-6}	0.0038
Iron	10×10^{-6}	0.0050
Manganin	44×10^{-6}	0.00001
Nichrome	100×10^{-6}	0.0004
Platinum	10×10^{-6}	0.0038
Silver	1.63×10^{-6}	0.0038
Tungsten	5.51×10^{-6}	0.0045

and is in the direction of the current, from the point a to the point b parallel to the wire. Substituting these values into Equation (26-13a), we have

$$\frac{I}{A} = \sigma \frac{V}{s},$$

$$I = \frac{\sigma A}{s} V.$$

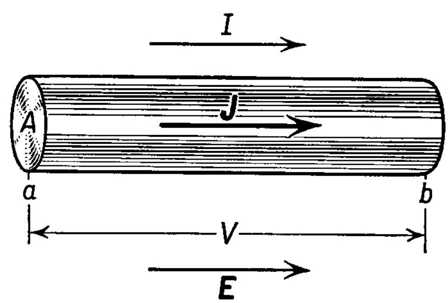


Fig. 26-9

By comparing this result with Equation (26-6), we see that the conductance G of a wire may be expressed in terms of its cross-sectional area A, its length s, and the conductivity σ of the material of which it is made as

$$G = \frac{\sigma A}{s}. \tag{26-15}$$

Similarly, the resistance R of a wire may be expressed in terms of its resistivity ρ as

$$R = \frac{\rho s}{A}. \tag{26-16}$$

From Equation (26-15) it may be seen that the conductivity σ is the conductance of a unit cube whose area A is 1 and whose length is also 1; similarly it may be observed from Equation (26-16) that the resistivity ρ is the resistance of a unit cube. In handbooks the resistivity is often expressed in units of ohm centimeter rather than ohm meter. This number may be used directly in Equation (26-16) by using the length of the wire in centimeters and its cross-sectional area in square centimeters, or by converting ohm centimeters to ohm meters according to

$$1 \text{ ohm cm} = 1 \text{ ohm cm} \times \frac{1 \text{ m}}{100 \text{ cm}} = \frac{1}{100} \text{ ohm m},$$

and using the appropriate mks units of length and area.

A variety of factors influence the resistivity of substances. For example, the resistivity of a conductor is affected by the state of internal strain. This is utilized in the *electrical strain gauge* in which a wire is cemented to a structural member and the strain of the member is determined by measuring the change in electrical resistance of the wire when a load is applied to the member.

The resistivity of bismuth is affected by a magnetic field, thus providing a basis for the measurement of the magnetic field with a bismuth probe. The resistivity of selenium is altered by incident light, and thin films of selenium may be used to measure light intensity. At very low temperatures, in the vicinity of a few degrees Kelvin, some metals have extremely low resistivity. The transition from the normal state to this *superconductive* state takes place in lead at a temperature of 7.3°K. If currents are started in a lead ring in the superconductive state, the current does not diminish by 0.1 per cent per hour. These currents are started and detected by magnetic means, to be described in a subsequent chapter.

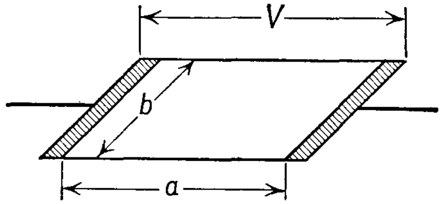
Insulating materials permit the passage of very weak currents which are usually negligible. These materials do not generally obey Ohm's law; typical values of the resistivity obtained at a particular value of the electric field intensity are usually quoted in handbooks for design purposes.

Insulating materials of high resistivity are used in the construction of electrical apparatus, and there is often the problem of current flow along the surface of the insulation rather than through the volume of the insulation. To minimize these surface currents a long surface path is often provided by corrugating the surface of an insulator. If we consider the factors affecting the electrical resistance of a conducting surface, we see that we would expect the resistance to increase directly with the separation of the electrodes a and to decrease with the length of the electrodes b, as shown in Figure 26-10. In the form of an equation we write

$$R = \rho_s \frac{a}{b} \cdot \tag{26-17}$$

Since both a and b are lengths, the quantity ρ_s, the surface resistivity, has the dimensions of resistance. The surface resistivity is often stated in units of ohms, or ohms per square, for in a square the lengths a and b are the same. The surface resistivity is the resistance between opposite sides of any square

Fig. 26-10



on the surface of the material. The volume resistivity and the surface resistivity for several insulating materials are given in Table 26-2. The surface resistivity of an insulator is affected by humidity. This effect has been utilized to determine the humidity of the atmosphere by employing glass surfaces with specially prepared surface coatings.

TABLE 26-2 VOLUME RESISTIVITY AND SURFACE RESISTIVITY OF SOME INSULATING MATERIALS

Material	Volume resistivity (ohm cm)	Surface resistivity (90% humidity) (ohms per square)
Bakelite	2×10^{11}	2×10^{8}
Ceresin wax	5×10^{18}	1×10^{17}
Glass	2×10^{13}	2×10^{6}
Hard rubber	1×10^{18}	2×10^{9}
Porcelain	3×10^{14}	5×10^{6}
Shellac	1×10^{16}	6×10^{9}
Wood (maple)	3×10^{10}	2×10^{9}

26-7 Temperature and Resistivity

The resistivity of a conducting substance generally depends upon its temperature. Just as in the case of the thermal expansion of solids discussed in Chapter 14, it is possible to represent the resistivity as a function of temperature by means of a mathematical series. If ρ_0 is the resistivity of the substance at some reference temperature t_0, the resistivity ρ_t at temperature t may be expressed as

$$\rho_t = \rho_0[1 + \alpha\,(\Delta t) + \alpha'\,(\Delta t)^2 + \cdots],$$

where α, α', and so on, are constant coefficients to be evaluated at the

reference temperature t_0, and

$$\Delta t = t - t_0.$$

It is often sufficient to approximate the above expression by the first two terms, so that we obtain

$$\rho_t = \rho_0 (1 + \alpha \Delta t), \tag{26-18}$$

where α is called the *temperature coefficient of resistivity* at the temperature t_0. Values of α for some of the common conducting materials are listed in Table 26-1. For metallic conductors α is always positive, which means that the resistance of a metallic conductor increases with temperature. Many nonmetallic conductors, such as carbon, have negative temperature coefficients, so that the resistivity decreases as the temperature increases. Other substances, such as the semiconductors germanium and silicon, exhibit positive coefficients at low temperatures and negative coefficients at high temperatures. In the particular case of semiconductors, the temperature coefficient of resistance is particularly sensitive to the impurity content. Circuit elements which exhibit large temperature coefficients of resistance are useful in many applications, either to compensate for the temperature variations in other parts of the circuit or to act as temperature indicators. Such elements are called *thermistors*. Some alloys, such as manganin, have been developed which have negligible temperature coefficients of resistance. These are useful for making coils whose resistance does not vary with temperature to any appreciable extent.

Some substances which are very good insulators at ordinary temperatures may become good conductors at high temperatures. The resistivity of glass at ordinary temperatures is of the order of 10^{14} ohm cm, but if a glass rod is raised to a temperature of about 400°C, it becomes a good conductor. If a glass rod is connected to the terminals of a 110-volt source and is heated to about 400°C, an appreciable current begins to flow in the rod. If the flame is removed, the current will continue to flow and even to increase in value until the glass melts, for the glass has a negative coefficient of resistivity.

The change with temperature in the resistance of a wire provides the basis for the *resistance thermometer*. The temperature-measuring element consists of a small coil of fine wire wound on an insulating quartz rod and surrounded by a protective quartz cylinder. The platinum resistance thermometer has been accepted as a standard means of measuring temperature in the range −190°C to +660°C. To achieve sufficient precision in the measurement of temperature with a platinum resistance thermometer the second- and third- order terms must be included in the formula for the variation of resistivity with temperature.

26-8 Resistance Measurements: Wheatstone Bridge

We have already discussed one method for measuring a resistance which involves the use of an ammeter and voltmeter. A much more accurate method for determining the resistance of a resistor, known as the *Wheatstone bridge method,* named for Charles Wheatstone (1802–1875), involves the use of three other resistors and a very sensitive current-measuring instrument known as a *galvanometer*. The circuit used is shown in Figure 26-11(a) in

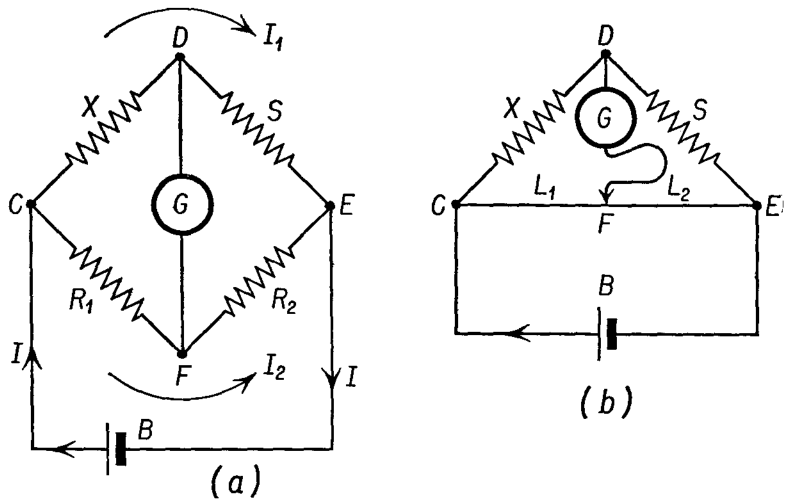


Fig. 26-11 (a) The Wheatstone bridge. (b) The slide wire form of Wheatstone bridge. The currents in the bridge are shown when the bridge is balanced; that is, when there is no current through the galvanometer.

which X is the resistance to be measured, R_1, R_2, and S are three other resistors, G is the galvanometer, and B is a battery. Let us suppose that the values of the three resistances R_1, R_2, and S have been so chosen that the galvanometer reads zero; that is, no current flows through it, which means that the difference of potential between D and F is zero, or that D and F are at the same potential. The bridge is then said to be balanced. At balance the potential difference between C and D is equal to the potential difference between C and F, or

$$V_{CD} = V_{CF},$$

and, similarly,

$$V_{DE} = V_{FE}.$$

Since there is no current from D to F, the current in DE is the same as that in CD. For the same reason the current in FE is the same as that in CF.

Applying Ohm's law to each of the resistors and substituting in the

above equations, we get

$$I_1X = I_2R_1,$$

and

$$I_1S = I_2R_2,$$

in which I_1 is the current in X and S, and I_2 is the current in R_1 and R_2.

From the above equations, we get

$$X = \frac{R_1}{R_2} S. \tag{26-19}$$

Thus if S is a known resistance, and the ratio of the two resistances R_1 and R_2 is known, the resistance X can be easily determined.

In laboratory work, S is either a single standard resistance whose value is known accurately, or else it consists of a series of coils of known resistances permitting a choice of a variety of values for S. R_1 and R_2 may consist of a single wire of known length and a slider making contact at some point F, Figure 26-11(b). Usually, when contact is first made at some point on this wire, a current will be started in the galvanometer. The slider is then moved along the wire until a point F is reached which is at the same potential as D, so that there is no current through the galvanometer. If the length of the wire from C to F is L_1, and from F to E is L_2, then, since the resistance of a uniform wire is proportional to its length, we can write

$$\frac{R_1}{R_2} = \frac{L_1}{L_2},$$

and Equation (26-19) becomes

$$X = \frac{L_1}{L_2} S. \tag{26-20}$$

Thus the measurement of a resistance is reduced to the measurement of the ratio of two lengths. This form of the Wheatstone bridge is usually called a *slide-wire* bridge.

Problems

26-1. What is the current when a charge of 6 coul passes through a wire in 2 sec?

26-2. How many electrons pass through the cross-sectional area of a wire in 1 sec if the wire is carrying a current of 1 amp. The charge of an electron is -1.60×10^{-19} coul.

26-3. A charged cloud containing 10^{15} electrons/cm^3 is moving in the positive x direction with a velocity of 150 m/sec. What is the current density (magnitude and direction) to be associated with the motion of the charged cloud?

26-4. A resistor carries a current of 6 amp when the voltage across it is 120

volts. (a) Determine its resistance. (b) Determine its conductance. (c) How much power is supplied to the resistor?

26-5. A 120-volt, 40-watt lamp is connected to a 120-volt line. (a) How much current does it require under normal operating conditions? (b) What is its resistance under these conditions?

26-6. An electric furnace powered by a 1,000-watt heating element contains a small opening 10 cm^2 in area in one end. If the only heat loss from the furnace is due to radiation from the opening, what will be the maximum temperature of the furnace?

26-7. (a) Calculate the resistance of a copper wire 1 m long and 2 mm in diameter at 0°C. (b) What is the resistance of a piece of Nichrome wire of the same size at the same temperature?

26-8. A Nichrome wire 10 m long and 2 mm in diameter is connected to a 6-volt storage battery. (a) What is the resistance of the wire? (b) What will be the current in the wire? (c) What will be the difference in potential between two points on the wire 1 m apart?

26-9. The resistance of a platinum wire used in a resistance thermometer is 4.85 ohms at 0°C. When used to measure the temperature of a liquid, the resistance is found to be 5.97 ohms. Determine the temperature of the liquid.

26-10. The resistance of a platinum wire is 6.25 ohms at 20°C. What is its resistance at 100°C?

26-11. A pair of plates 1 m^2 in area is separated by a distance of 10 cm. The plates are immersed in a brine solution. Assuming that the current density in the brine solution is uniform, determine the conductivity of the brine solution if there is a current of 10 amp between the plates when the difference in potential is 24 volts.

26-12. The conductivity of a normal solution of KCl at 15°C is 0.093 mho/cm. (a) What is the conductance between the plates 1 m^2 in area immersed in this solution if the separation between the plates is 25 cm? (b) What potential difference must be maintained between the plates if a current of 5 amp is to flow through the tank? (c) What will be the current density, assuming that it is uniform?

26-13. In a printed circuit two conductors 10 cm long are printed onto a dielectric. The conductors are separated by a distance of 3 mm. If the surface resistivity of the dielectric is 2×10^8 ohms/square, what will be the resistance between the conductors?

26-14. Two conductors in the form of concentric cylindrical rings are imbedded in a sheet of Bakelite 1 cm thick of volume resistivity 2×10^{11} ohm cm. The diameter of the inner conducting ring is 2 cm, while the diameter of the outer conducting ring is 5 cm. Find the resistance of the Bakelite between the two rings. (Use calculus method.)

26-15. In Problem 26-14 compute the surface resistance between the conducting rings, taking both surfaces of the Bakelite into account. The surface resistivity of the Bakelite is 2×10^8 ohms/square.

26-16. The precision to which commercial electrical panel meters are made is 2 per cent. What is the largest percentage error which can be expected in measuring a resistor by the use of a voltmeter and ammeter?

27

Direct-Current Circuits

27-1 Electromotive Force

When a charged particle traverses a closed path in an electrostatic field in space, the total work done on the particle is zero. The electric field is *conservative*. Associated with each point in the field, there is a fixed value of the electrical potential.

In a simple circuit consisting of a chemical cell and a resistor, as shown in Figure 27-1, we have seen that the current flows through the resistor from the positive terminal of the cell to the negative terminal of the cell. The positive terminal is at the higher potential, so that the current flows from the higher to the lower potential outside the cell, but inside the cell the direction of current flow is from the lower to the higher potential. The cell must do work upon the charge in order to raise its potential.

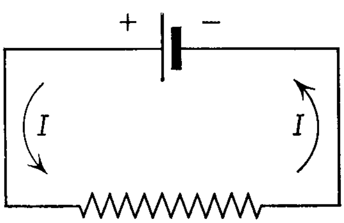


Fig. 27-1

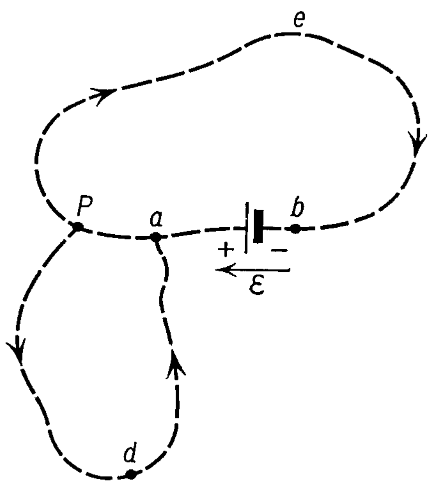


Fig. 27-2

Let us consider the effect of moving a positive probe charge through a closed path in an electric field in which such a cell is located. As long as the path of the charged particle does not pass through the cell, the work

done in traversing a closed path is zero, as in the path *PdaP* in Figure 27-2. When the path traverses the cell, the particle may acquire energy as a result of the conversion of chemical energy to electrical energy, as in the path *PebaP*, or the particle may lose energy as the result of the conversion of electrical energy to chemical energy if this same path is traversed in the direction opposite to the original direction, as in the path *PabeP*. If one takes into account only the mechanical work done by the agency moving the probe charge, the cell appears to be a place where energy may be gained or lost, depending on the direction in which the passage through the cell is effected. A similar result would be obtained by passage of the probe charge through any electric generator in which the conversion of energy from some other form to electrical energy is reversible.

The concept of *electromotive force* has been introduced to describe the energy relations associated with electric circuits which incorporate chemical cells or other electric generators. If a net quantity of work $\mathscr{W}$ is done in carrying a charge q around a closed path in which no current is flowing, as in a circuit when the switch is open, the total electromotive force $\mathscr{E}$ in that path is defined as

$$\mathscr{E} = \frac{\mathscr{W}}{q}\cdot \tag{27-1}$$

In this definition of electromotive force $\mathscr{E}$, abbreviated *emf*, it is important to specify that the charge may traverse the path with arbitrary slowness. In other words, if the work done in traversing the path depends upon the speed of the particle, we will imagine that the particle is carried around the path with near zero speed. Since the work done when a charge traverses a resistor depends upon the current, which may be related to the speed of the charged particle, the work done in very slowly moving a probe charge through a resistor will approach zero, as long as the potential difference between the ends of the resistor is zero.

From Equation (27-1) we see that the units of electromotive force are the same as the units of potential difference. In the *mks* system of units, the unit of emf is the volt. The emf is not a force but an energy per unit charge. If the potential difference across the terminals of a cell is 2 volts when no current is flowing, we say that the cell has an emf of 2 volts.

Since the work done on a probe charge q in traversing the path *PadP* is zero, the emf along this path is zero. In traversing the path *PebaP*, if the potential difference between the terminals of the cell is V, the work done by the cell on the charge is Vq, and the emf is

$$\mathscr{E} = \frac{Vq}{q} = V,$$

for the particle has more energy when it has returned to P than it had

initially. Positive work has been done on the particle. In traversing the same path in the opposite direction $PabeP$, the emf is

$$\mathcal{E} = \frac{-Vq}{q} = -V,$$

for the particle has done work in traversing the cell. In other words, negative work has been done on the particle, so that its energy is less upon its return to P than it was initially. The *emf of a cell* or of a generator thus has a sense, and *is directed from the negative terminal of the cell to its positive terminal* inside the cell. When a positive charge passes through a cell in the direction of the emf, it gains in potential energy. When a positive charge passes through the cell in a direction opposite to the emf, it loses potential energy. In both cases the change in electrical potential is equal to the magnitude of the emf $\mathcal{E}$ of the cell.

27-2 Series and Parallel Connections

An electric circuit may be very simple and consist of one or two electrical devices connected to a source of power, or it may be very complex and consist of many different elements connected in a variety of ways. In

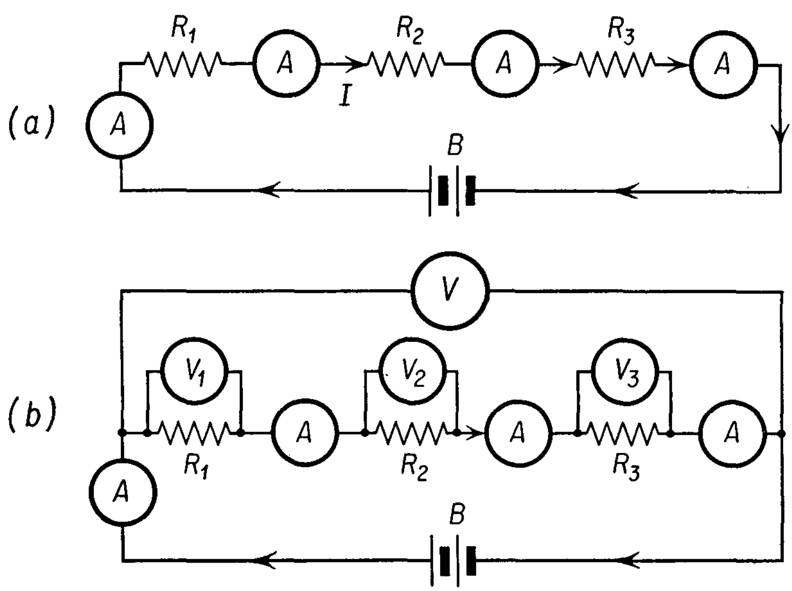


Fig. 27-3 Resistors connected in series. (a) The current is the same in each resistor. (b) The difference of potential across all the resistors in series is equal to the sum of the differences of potential across each of the resistors.

practical applications it is important to be able to determine the *equivalent resistance* of any circuit or any section of the circuit in terms of the resistances of the individual elements of a circuit. Suppose we have a circuit consisting of a battery, three resistors, and four ammeters, connected as shown in Figure 27-3(a). This is called a *series* circuit, and the various elements are said to be connected in *series.* In a series circuit *the current is the same in all parts of the circuit.* The same current flows in each of the resistors, each of the ammeters, and the battery.

Another type of connection is shown in Figure 27-4 in which the three resistors are connected in *parallel.* Since the difference of potential between two points, such as C and D, can have only one value, the difference of potential across each resistor is the same. A voltmeter connected across C and D will give the difference of potential across each resistor and, in this case, will also give the difference of potential across the terminals of the battery. The voltmeter is always connected in parallel with that portion of the circuit whose voltage is to be measured, while the ammeter is always connected in series in that portion of the circuit in which the current is to be measured.

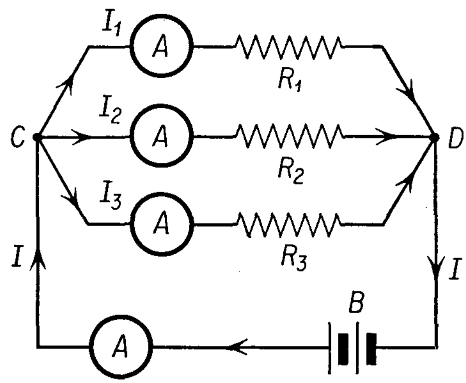


Fig. 27-4 Resistors in parallel. The difference of potential is the same across each resistor.

In Figure 27-3(b), if three voltmeters are connected across the three resistors in series, R_1, R_2, and R_3, and a fourth voltmeter is connected across all three resistors, it will be observed that the difference of potential V across all three resistors in series is equal to the sum of the differences of potential across each of the resistors, or

$$V = V_1 + V_2 + V_3. \tag{27-2}$$

Applying Ohm's law to each resistor and remembering that the current I is the same in each one, we get

$$V_1 = IR_1, \qquad V_2 = IR_2, \qquad \text{and} \qquad V_3 = IR_3,$$

which yields, upon substitution in Equation (27-2),

$$V = IR_1 + IR_2 + IR_3. \tag{27-3}$$

The equivalent resistance R of this circuit is one which would have the same

current I flowing in it when the same potential difference V is applied to it; that is,

$$V = IR. \tag{27-4}$$

Equating these two values of V, we get

$$R = R_1 + R_2 + R_3. \tag{27-5}$$

The equivalent resistance of a group of resistors connected in series is the sum of their individual resistances.

If ammeters are inserted into the circuit containing three resistors in parallel, as shown in Figure 27-4, it will be observed that the current I which leaves the battery divides at C in such a way that it is equal to the sum of the currents in the individual resistors, or

$$I = I_1 + I_2 + I_3, \tag{27-6}$$

where I_1 is the current in R_1, I_2 the current in R_2 and I_3 the current in R_3. These currents recombine at D and flow back to the battery. Applying Ohm's law to each resistor and remembering that the potential difference V is the same across each one, we get

$$V = I_1R_1, \qquad V = I_2R_2, \qquad \text{and} \qquad V = I_3R_3.$$

We may therefore write Equation (27-6) as

$$I = \frac{V}{R_1} + \frac{V}{R_2} + \frac{V}{R_3};$$

and if R is the equivalent resistance of these three parallel resistors, then

$$I = \frac{V}{R} = \frac{V}{R_1} + \frac{V}{R_2} + \frac{V}{R_3},$$

from which

$$\frac{1}{R} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3}; \tag{27-7}$$

this is the relationship between the equivalent resistance R and the individual resistances of three resistors in parallel. The reciprocal of the resistance is called the *conductance* of the resistor. Equation (27-7) may be read as follows; the equivalent conductance of a parallel circuit is the sum of the conductances of the individual resistors connected in parallel. Equation (27-7) also shows that the equivalent resistance of a parallel combination is less than the resistance of any one of the resistors.

Illustrative Example. Find the potential difference between the terminals of the 12-ohm resistor of Figure 27-5(a) if the battery has an emf of 36 volts and has no internal resistance.

Let us first determine the equivalent resistance of the three resistors in the circuit. The equivalent resistance of the parallel combination of the 6- and

12-ohm resistors may be found as

$$\frac{1}{R} = \frac{1}{6 \text{ ohms}} + \frac{1}{12 \text{ ohms}},$$

or

$$R = 4 \text{ ohms}.$$

We may replace the parallel combination by a single 4-ohm resistor, as shown in Figure 27-5(b). This series combination of two resistors has an effective resistance of 6 ohms. Thus the three resistors of Figure 27-5(a) may be replaced by a single 6-ohm resistor, as shown in Figure 27-5(c).

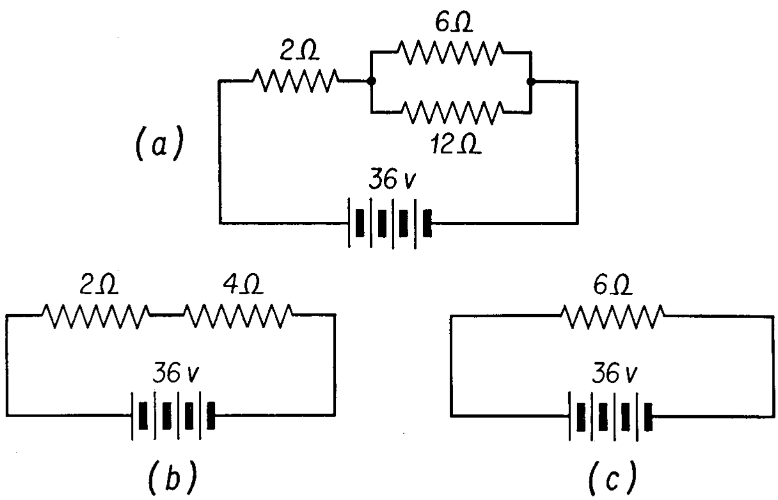


Fig. 27-5

From Ohm's law the current in the circuit of Figure 27-5(c) is

$$I = \frac{V}{R} = \frac{36 \text{ volts}}{6 \text{ ohms}}$$

$$= 6 \text{ amp}.$$

Thus a current of 6 amp flows from the battery in each of the equivalent circuits of Figure 27-5. In Figure 27-5(b) the potential difference between the terminals of the 4-ohm resistor is therefore equal to 24 volts. Replacing the 4-ohm resistor by the original parallel combination, as in Figure 27-5(a), the potential difference between the terminals of the 12-ohm resistor is 24 volts.

27-3 Terminal Voltage

In practical electrical generators the passage of current through the generator is accompanied by the evolution of heat within the generator. To

describe the evolution of heat in a quantitative way, we say that practical generators have an emf and an *internal resistance* as well. It is customary to represent the internal resistance of a practical cell by a resistor r which is in series with a resistanceless cell of emf $\mathcal{E}$. The potential difference as measured by a voltmeter connected across the terminals of the cell ab (Figure 27-6) will depend upon the current I flowing through the cell.

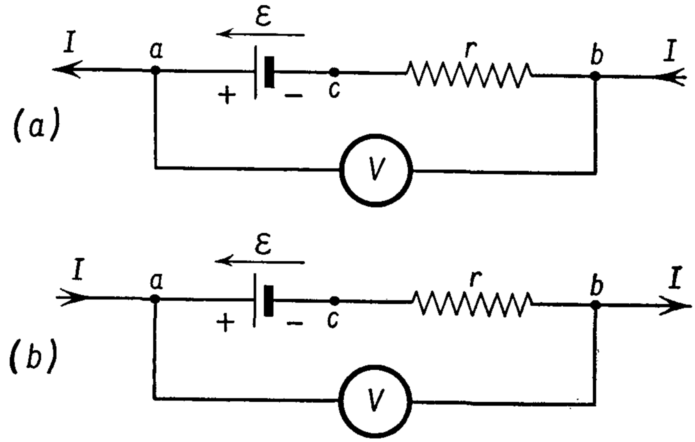


Fig. 27-6 Terminal voltage and emf of (a) cell discharging and (b) cell being charged.

Let us suppose that the current through the cell is in the direction of the emf, as shown in Figure 27-6(a). The current through the internal resistance r is in the direction bc. According to Ohm's law, the point c must be at a lower potential than the point b by an amount given by

$$V = Ir.$$

If we carry a unit positive probe charge from the point b to the point c and then to the point a, the potential falls by Ir in passing from b to c and rises by $\mathcal{E}$ on passing from c to a. The potential difference between b and a is given by

$$V_a = V_b - Ir + \mathcal{E}, \tag{27-8}$$

so that

$$V_a - V_b = \mathcal{E} - Ir. \tag{27-9}$$

Thus the difference of potential between the terminals of the cell, also called the *terminal voltage,* is less than the emf of the cell when the current is in the direction of the emf; this is the case when the cell is discharging.

When the direction of the current is opposite to that of the emf, as in Figure 27-6(b), we note that the point c must be at a higher potential than the point b by an amount Ir, so that the potential difference between points

b and a is given by

$$V_a - V_b = \mathcal{E} + Ir. \tag{27-10}$$

When a battery is being charged, the current is in a direction opposite to its emf, and its terminal voltage is greater than $\mathcal{E}$.

As a dry cell or a storage battery deteriorates with use, its emf remains substantially constant, but its internal resistance increases. The cell must be tested under conditions in which it is required to supply currents appropriate to its use, in order to determine whether the cell is "dead." When the terminal voltage is appreciably less than the emf, the battery may no longer be suitable for its intended application.

27-4 Voltmeters and Ammeters

The voltmeters and ammeters used in electrical measurements are constructed from a basic meter movement which will be described in a subsequent chapter. For our present purposes we need only know that this basic meter is a device which measures current, called a *galvanometer.* Galvanometers are usually constructed so as to have scale deflections proportional to the current passing through them. The galvanometer has an internal resistance.

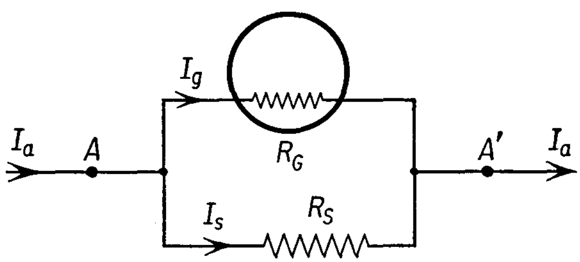


Fig. 27-7 Ammeter.

Let us suppose that we have a galvanometer which is so constructed that its internal resistance is R_g and that it is deflected to a full-scale reading whenever the current passing through it is I_g. If we wish to have an ammeter whose full-scale reading is I_a, we may construct such a meter from the galvanometer by connecting a resistor R_s across the terminals of the galvanometer. Such a resistor is called a galvanometer *shunt.* The shunt resistor is often connected inside the case of the galvanometer; the external connections to the instrument are shown in Figure 27-7 as AA'.

Suppose we wish the galvanometer to indicate a full-scale deflection when a current I_a flows through the connections AA'. Since the construc-

tion of the galvanometer is such that it reads full-scale deflection when the current through it is I_g, the current flowing through the shunt resistor I_s must be given by

$$I_a = I_g + I_s.$$

Since the galvanometer is in parallel with the shunt resistor, the potential difference across the galvanometer must be the same as the potential difference across the shunt resistor. Thus

$$I_g R_g = I_s R_s.$$

The value of the required shunt resistor may be determined by eliminating the current I_s from these two equations. We have

$$I_a = I_g \left(1 + \frac{R_g}{R_s}\right), \tag{27-11}$$

so that a knowledge of the resistance of the galvanometer, the current through the galvanometer for full-scale deflection, and the desired full-scale ammeter reading suffice to determine the required shunt resistance.

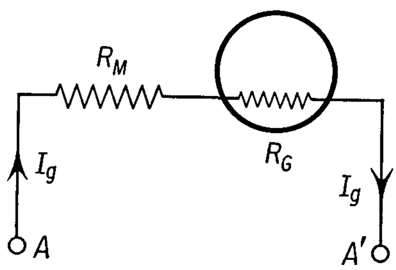


Fig. 27-8 Voltmeter.

It is impossible to make an ammeter whose full-scale deflection requires less current than is required by the galvanometer from which it was constructed. An ammeter reading large currents requires a shunt resistor of smaller resistance than does one reading small currents. Since an ammeter is connected in series with the circuit in which the current is being measured, an ammeter with very low resistance is desired in order that it may have a minimum influence upon the current in the circuit.

The same galvanometer may be used as a voltmeter by connecting a *multiplier* resistor in series with it, as shown in Figure 27-8. The terminals of the voltmeter AA' are to be connected across (in parallel with) two points in a circuit whose potential difference we wish to measure. The galvanometer reads full-scale deflection when a current I_g passes through it. The same current also passes through the multiplier resistor R_m. From Ohm's law the potential difference V, between the points A and A', is given by

$$V = I_g(R_m + R_g). \tag{27-12}$$

Thus we may construct a voltmeter with any desired full-scale reading V by inserting a resistor R_m in series with a galvanometer, if the internal

resistance R_g of the galvanometer and the current I_g required for full-scale deflection are known.

A good voltmeter should be constructed from a galvanometer of high sensitivity. The voltmeter is connected in parallel with the circuit component whose voltage is being measured. In order that the circuit be least disturbed by the presence of the voltmeter, the current through the voltmeter should be small. Thus it is important that the resistance of the voltmeter be high.

It is possible to construct multipurpose meters from a single galvanometer. A number of appropriate resistors are mounted in the case with the galvanometer, and a switching arrangement is provided which connects selected resistors in series with the galvanometer for voltage measurement, while other resistors are connected in parallel with the galvanometer for use of the instrument as an ammeter.

Illustrative Example. It is desired to convert a galvanometer, that has an internal resistance of 5 ohms and gives a full-scale deflection for a current of 10 μamp, into a voltmeter whose full-scale deflection corresponds to 300 volts. What value of multiplier resistance should be used?

Substituting numerical values into Equation (27-12), we have

$$300 = 10 \times 10^{-6}(R_m + 5),$$

$$R_m = 2{,}995 \text{ ohms}.$$

To convert the same galvanometer to an ammeter with a full-scale reading of 100 μamp, a shunt resistor is used whose resistance is given by Equation (27-11).

$$100 \times 10^{-6} = 10 \times 10^{-6}\left(1 + \frac{5}{R_s}\right)$$

$$R_s = \tfrac{5}{9} \text{ ohm}.$$

27-5 The Potentiometer

When a voltmeter is used to measure a potential difference, some error is always introduced because of the fact that the voltmeter draws current from the circuit to which it is connected. When a voltmeter is used to measure the potential difference between the terminals of a cell, the measured voltage is always less than the emf because of the voltage drop across the internal resistance of the cell. In many applications it is important to be able to measure the potential difference between the two terminals without drawing current to the measuring device. In such cases a *potentiometer* is used.

The schematic circuit of a potentiometer is shown in Figure 27-9. At the heart of the potentiometer is a *slide-wire* resistor made of a uniform

wire ac, provided with a sliding contact, shown in the figure at point b. The resistance of the portion of the wire from the fixed contact at a to the sliding contact at b is proportional to the length of the wire between these two points. A constant current I is maintained in the slide wire by means of the battery B, so that the potential difference between a and b is proportional to the length of the wire between these two points. In this way any potential difference from zero to the terminal voltage of the battery can be obtained between the points a and b simply by moving the sliding contact.

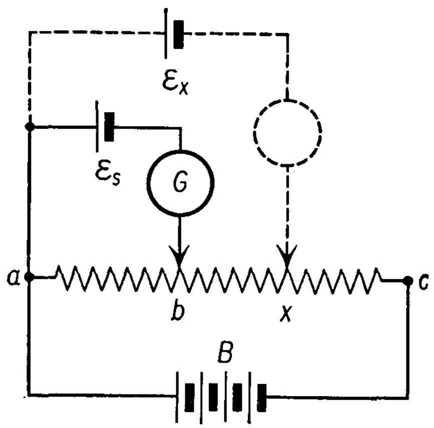


Fig. 27-9 A potentiometer.

In order to use the potentiometer, it is necessary first to calibrate it with the aid of a *standard cell* such as the Weston normal cell whose emf has been previously determined. The positive terminal of the standard cell is connected to the same point a as the positive terminal of the battery, while the negative terminal of the standard cell is connected to the point b through a sensitive galvanometer. The sliding contact b is moved along the slide wire until there is no current through the galvanometer, indicating that there is no potential difference across the terminals of the galvanometer. Thus the emf of the standard cell is equal to the potential difference V_{ab} along the slide wire between these points. If the resistance of the slide wire between the point a and the point b is R_{ab} we have,

$$\mathcal{E}_s = V_{ab} = IR_{ab}.$$

To determine the emf of an unknown cell $\mathcal{E}_x$, the standard cell is replaced by the unknown cell, and the sliding contact is shifted to a new position x to achieve a balance. We have

$$\mathcal{E}_x = V_{ax} = IR_{ax}.$$

Dividing the second of these equations by the first, we obtain

$$\mathcal{E}_x = \mathcal{E}_s\left(\frac{R_{ax}}{R_{ab}}\right).$$

If l_b is the length of the slide wire from a to b and l_x is the length from a to x,

and since the resistance of a uniform wire is proportional to its length, we have

$$\frac{R_{ax}}{R_{ab}} = \frac{l_x}{l_b},$$

so that

$$\mathcal{E}_x = \mathcal{E}_s\left(\frac{l_x}{l_b}\right). \tag{27-13}$$

The measurement of the emf of a cell is thus reduced to the measurement of two lengths along a uniform slide wire. Since the potentiometer uses a galvanometer to determine the *null condition,* the condition of no current flow, an extremely sensitive instrument may be used. Furthermore, we have no need to compound the errors of measurement by first measuring the current from the battery or the resistance of the slide wire. Any potential differences may be measured relative to a standard cell with an accuracy limited only by the uniformity of the slide wire and the sensitivity of the galvanometer.

The potentiometer may be used to calibrate voltmeters and ammeters, to measure potentials in an electrolytic plotting tank, to measure the thermal emf in a thermocouple for the measurement of temperature, and, in fact, wherever an accurate determination of potential difference is required.

27-6 Kirchhoff's Laws

Complicated circuits made up of resistors and sources of emf often cannot be readily resolved into series and parallel combinations of resistors. The procedures for solving such networks were first stated by Gustav Robert Kirchhoff (1824–1887) and are known as Kirchhoff's laws. In order to state concisely these rules for the solution of circuit problems, we shall first define two terms, a *junction* and a *loop.* A *junction,* or branch point, is a point where three or more conductors, or branches, are electrically connected. A *loop* is any closed path in an electrical circuit.

The concept of conservation of charge leads to the first of Kirchhoff's laws, which states that *the algebraic sum of the currents flowing into a junction is zero.*

$$\sum I = 0.$$

In the above equation, representing Kirchhoff's first law, current flowing into a junction is generally regarded as positive current, while current flowing out of the junction is regarded as negative. The current flowing into any junction must equal the current flowing out of that junction. If this were not true, we would have to imagine that charge could be created or

destroyed at a junction, in contradiction of the principle of conservation of electric charge.

The second of Kirchhoff's laws may be stated as *the algebraic sum of the potential differences around a closed loop is equal to zero.* Since the potential differences between the terminals of the circuit elements are composed of emf's, and IR drops across resistors, we may write Kirchhoff's second law symbolically as

$$\sum\mathcal{E} + \sum IR = 0.$$

In effect, this rule states that the potential of a point in an electric circuit is a fixed quantity. If we start at any point in the circuit and imagine that we transport a probe charge around a closed loop, the algebraic sum of the potential changes associated with passing from one junction to the next in passing around the loop is equal to zero.

To apply Kirchhoff's laws to a network, we first imagine the network to be broken up into loops, as shown in Figure 27-10, in which circuit elements have been represented as rectangles. We choose these loops to be as simple as possible, so that each circuit element is contained in at least one loop. Some circuit elements may be part of more than one loop. We assign arbitrarily a direction to the current in each loop. We indicate this direction by an arrow, and assign to the current an algebraic magnitude represented by the symbol I_1, I_2, and so on. In drawing these currents as continuous through a junction, we have automatically fulfilled Kirchhoff's first law. In the event that we have assigned the wrong direction to any of these loop currents, the numerical solution for that current will yield a negative answer.

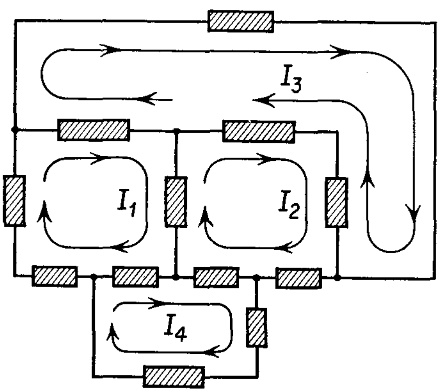


Fig. 27-10 Current loops.

To each loop we apply Kirchhoff's second law; starting at a particular point in the circuit, we imagine that we carry a unit positive probe charge from junction to junction. The probe charge may be carried around the loop in any direction, without regard to the direction of the loop current. When the probe charge is carried through a source of emf in the direction of the emf, passing from the negative to the positive terminal, the potential of the probe charge is increased by the magnitude of the emf. When the probe charge is carried through the source of emf from the positive to the

negative terminal, its potential is decreased in amount by the magnitude of the emf. When the probe charge passes through a resistor in the direction of the current, the potential of the probe charge is decreased, for the direction of the current in a resistor is from high to low potential. If the probe charge is passed through the resistor in a direction opposite to the direction of the current, the potential of the probe charge is increased. Because of the manner in which we have indicated the loop currents, some elements will have current contributions from more than one loop current. Since the net current through the resistor determines the potential difference between its terminals, we may treat the potential change due to each loop current as an independent contribution to the potential change of the probe charge.

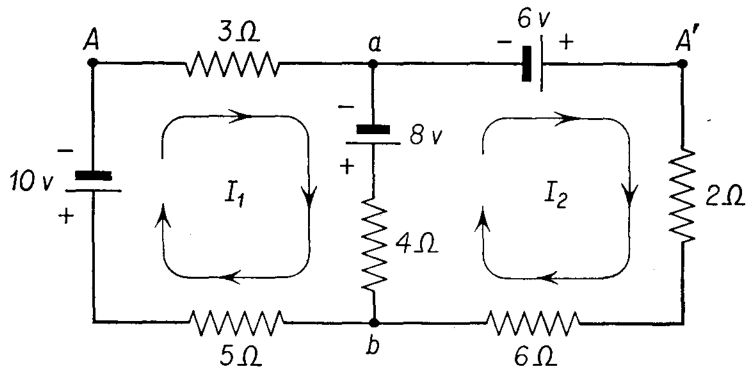


Fig. 27-11

Illustrative Example. Find the potential difference between the points a and b in the circuit of Figure 27-11.

We imagine the circuit to be broken into two closed loops, as shown in the figure, with the 8-volt cell and the 4-ohm resistor common to both loops, and represent the currents in the two loops by I_1 and I_2 which have been drawn arbitrarily in the clockwise direction. We imagine that we have a unit positive probe charge at the point A, and we carry the probe charge around the first loop in the direction of the current in this loop. We shall represent an increase in the potential of the probe charge as positive and a decrease of the potential of the probe charge as negative. The algebraic sum of the potential changes of the probe charge in passing around the loop must be equal to zero.

On passing through the 3-ohm resistor in the direction of the current I_1, the potential of the probe charge falls. The change in potential of the probe charge is $-3I_1$. The probe charge next passes from the negative to the positive terminal of the 8-volt cell, so that its potential change is $+8$ volts. There are two contributions to the potential change of the probe charge in passing through the 4-ohm resistor. Since it passes through the resistor in the direction of the current I_1, the potential of the probe charge decreases by $4I_1$. At the same time the probe charge passes through the resistor in a direction opposite to I_2, so that its

potential increases $4I_2$. The net change in potential of the probe charge is

$$-4I_1 + 4I_2.$$

Next the probe charge passes through the 5-ohm resistor where its potential changes by $-5I_1$. Finally it returns to the point A by passing through the 10-volt cell from the positive to the negative terminal, making a change in potential of -10 volts. Adding all these changes algebraically and setting the sum equal to zero, we find

$$-3I_1 + 8 - 4I_1 + 4I_2 - 5I_1 - 10 = 0,$$

from which

$$-12I_1 + 4I_2 - 2 = 0$$

for the equation resulting from the application of the loop condition to the first loop.

Starting with the point A' and traversing the second loop in the counterclockwise direction, opposite to the direction of the current I_2, we find

$$-6 + 8 - 4I_1 + 4I_2 + 6I_2 + 2I_2 = 0,$$

yielding

$$-4I_1 + 12I_2 + 2 = 0$$

as the equation resulting from the second loop.

We may solve these equations simultaneously to find

$$I_1 = I_2 = -\tfrac{1}{4} \text{ amp.}$$

The currents in both loops are therefore opposite to the directions indicated in Figure 27-11.

Having found the currents in the circuit, the remainder of the problem is quite straightforward. Since I_1 and I_2 are equal in magnitude and opposite in direction between a and b, there is no current between the junctions a and b. To find the potential difference between these two points, we imagine the unit probe charge to be at position a. At this point its potential is V_a. To move the probe charge to position b it must pass through the 8-volt cell, its potential being increased by the emf of the cell, and through the 4-ohm resistor. Since there is no current through the resistor, there is no potential change on passing through it. The probe charge is then at the potential of point b, V_b. In the form of an equation

$$V_a + 8 \text{ volts} = V_b;$$

$$V_b - V_a = 8 \text{ volts.}$$

27-7 Back EMF of a Motor

As we have seen in Section 27-1, the concept of electromotive force is a useful one in connection with the operation of generators of electricity. We may extend this concept so that it is applicable to motors as well, through the idea of a *back emf*. It is possible to construct a d-c motor so that the direction of rotation of the motor depends upon the direction of the

current through it. Unlike chemical cells or generators, which are polarized so that they convert some other form of energy to electrical energy when the charge passes in one direction and convert electrical energy back to that other form when the charge passes in the reverse direction, a motor always converts electrical energy to mechanical energy.

It is possible to describe this property of the motor electrically in terms of an emf, called a back emf, which is always opposite in direction to the current. The electrical power $\mathcal{P}$ converted by the motor to mechanical power is

$$\mathcal{P} = \mathcal{E}I,$$

where I is the current delivered to the motor, and $\mathcal{E}$ is its back emf. In general, a motor may be described by its back emf $\mathcal{E}$ and its internal resistance R when it is part of an electrical circuit.

Illustrative Example. (a) What is the back emf of an electric motor which has an internal resistance $r = 5$ ohms and draws a current $I = 2$ amp when connected to a source whose terminal voltage V is 110 volts? (b) What is the efficiency of the motor?

(a) The schematic circuit is shown in Figure 27-12 where the motor has been represented by a generator, whose emf $\mathcal{E}$ is directed opposite to the direction of the current, in series with a resistor representing the internal resistance of the motor. If we imagine that the probe charge is initially at the point A and that we carry the probe charge in a counterclockwise direction around the loop, we find, from Kirchhoff's second law,

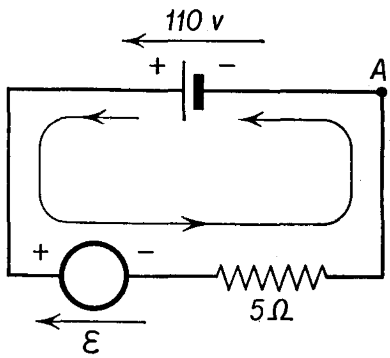


Fig. 27-12

$$V - \mathcal{E} - rI = 0.$$

Substituting numerical values, we get

$$+\,110 \text{ volts} - \mathcal{E} - 5 \text{ ohm} \times 2 \text{ amp} = 0,$$

$$\mathcal{E} = 100 \text{ volts}.$$

(b) To find the efficiency of the motor, we observe that the energy $\mathcal{W}_i$ delivered to the motor in time t is

$$\mathcal{W}_i = 110 \text{ volts} \times 2 \text{ amp} \times t = 220 \text{ watts} \times t.$$

The energy converted by the motor to mechanical energy is given by the product of the back emf by the charge which has passed through the motor in time t so that the work done by the motor $\mathcal{W}_0$ is given by

$$\mathcal{W}_0 = 100 \text{ volts} \times 2 \text{ amp} \times t \text{ sec} = 200 \text{ watts} \times t.$$

The efficiency e of the motor is

$$e = \frac{\mathscr{W}_0}{\mathscr{W}_i} = \frac{200 \text{ watts} \times t}{220 \text{ watts} \times t},$$

$$e = 0.91.$$

Problems

27-1. Two resistors of 30 and 45 ohms resistance, respectively, are connected in parallel, and the combination is connected to a 120-volt source. Determine (a) the effective resistance of this combination and (b) the current through each resistor.

27-2. Three coils of 20, 30, and 50 ohms resistance, respectively, made of uniform wire, are connected in parallel, and the group is then connected to a 110-volt source. Find (a) the resistance of the combination and (b) the current through each resistance.

27-3. Three resistors having resistances of 15, 25, and 50 ohms, respectively, are connected in series, and a difference of potential of 120 volts is maintained across the combination. (a) What is the current in each resistor? (b) What is the voltage across each resistor? (c) How much power is supplied to this combination?

27-4. In the circuit sketched in Figure 27-13, the current in the 10-ohm coil is 4.5 amp. (a) Calculate the value of the resistance R. (b) Determine the amount of heat developed in 1 min in the 10-ohm coil.

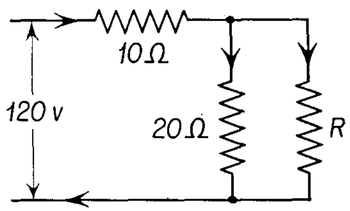


Fig. 27-13

27-5. Three resistors having resistances of 4, 8, and 12 ohms, respectively, are connected in series. A storage battery maintains a difference of potential of 12 volts across the combination. How much power is delivered to each resistor?

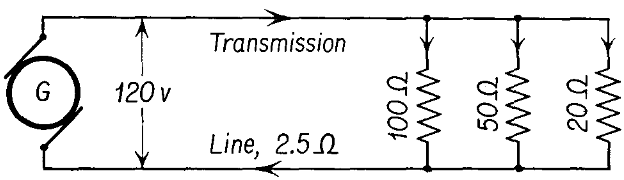


Fig. 27-14

27-6. A lamp, an electric heater, and an electric iron are connected in parallel, as shown in Figure 27-14. Their resistances, when hot, are 100, 50, and 20 ohms, respectively. If the generator produces a voltage across its terminals of 120 volts, and if the transmission line has a resistance of 2.5 ohms, find (a) the current supplied by the generator, (b) the voltage across the terminals of the lamp, (c) the current in the heater, and (d) the power consumed by the heater.

27-7. A generator which maintains a constant terminal voltage of 120 volts

supplies a current of 16 amp to a group of 10 identical lamps in parallel. The line connecting the generator to the lamps has a resistance of 0.5 ohm. (a) What is the voltage at the lamps? (b) What is the resistance of each lamp?

27-8. Figure 27-15 is a diagram of a part of an electric circuit. If the current in the 6-ohm resistor is 3 amp, find the following quantities: (a) the reading of the voltmeter connected between C and D; (b) the current in the 8-ohm resistor; (c) the reading of the ammeter placed between B and C; (d) the potential difference between points A and B; and (e) the current in the 20-ohm resistor.

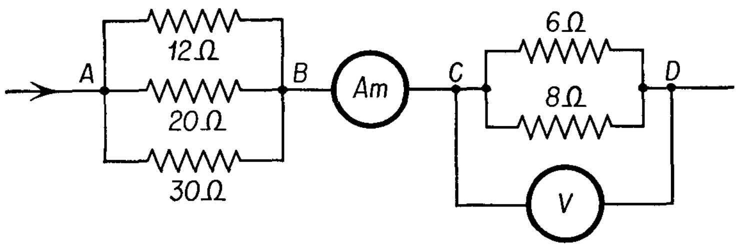


Fig. 27-15

27-9. Two lamps, each rated at 40 watts and 120 volts, are used as resistors in a circuit. If the two lamps are connected in series, (a) what is their combined resistance and (b) how much power is dissipated by these lamps when the voltage across the two is 120 volts?

27-10. A battery has an emf of 6 volts and an internal resistance of 3 ohms. What will be the current delivered to a 100-ohm resistor connected to the terminals of the battery?

27-11. When a generator is delivering a current of 50 amp, its terminal voltage is 110 volts. When the same generator delivers a current of 100 amp, its terminal voltage is 100 volts. What are the values of (a) the emf and (b) the internal resistance of this generator?

27-12. When a storage battery is delivering a current of 20 amp to a load, its terminal voltage is 5.9 volts. When the same battery is being charged by a generator at a current of 5 amp, its terminal voltage is 6.05 volts. What are the values of (a) the internal resistance and (b) the emf of the battery?

27-13. A generator with an emf of 120 volts and an internal resistance of 5 ohms delivers a current of 3 amp to a motor whose internal resistance is 2 ohms. (a) What is the back emf of the motor? (b) What is the mechanical power delivered by the motor? (c) What is the efficiency of the motor? (d) What is the over-all efficiency of the system?

27-14. An ammeter reads a full-scale deflection of 1 amp and has a shunt resistance of 0.1 ohm. The galvanometer from which it is constructed gives a full-scale deflection when a current of 100 μamp passes through it. What is the resistance of the galvanometer?

27-15. It is desired to construct a multimeter with ammeter ranges of 0.01 amp, 0.1 amp, and 1.0 amp and with voltmeter ranges of 0.1 volt, 1.0 volt, and 10 volts from a galvanometer having a full-scale deflection at a current of 100

μamp and an internal resistance of 50 ohms. Find the value of the shunt or multiplier resistance in each case.

27-16. A voltmeter having a resistance of 100,000 ohms and a full-scale deflection of 1 volt is connected across a 50,000 ohm-resistor. The voltmeter indicates a potential difference of 0.75 volt. What is the potential difference between the terminals of the resistor when the voltmeter is disconnected?

27-17. An ammeter which has a resistance of 50 ohms and a full-scale deflection of 0.1 milliampere (abbreviated ma) is connected in series with a circuit consisting of a cell and a 200-ohm resistor. The meter reads 0.08 ma. What is the current in the circuit when the milliammeter is removed?

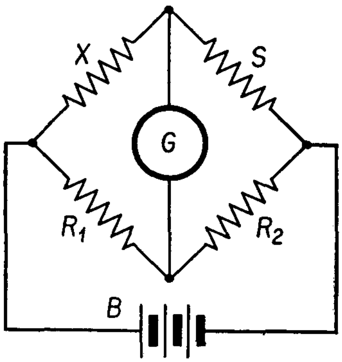


Fig. 27-16

27-18. In the Wheatstone bridge of Figure 27-16, the standard resistor S is 100 ohms, while the resistor R_1 is 150 ohms and R_2 is 50 ohms at balance. (a) What is the value of the unknown resistor X? (b) Find the current supplied by the battery whose emf is 6 volts and whose internal resistance is 3 ohms.

27-19. A battery has an internal resistance of R ohms. How large a resistor should be connected across its terminals in order that the greatest amount of heat is generated in the resistor?

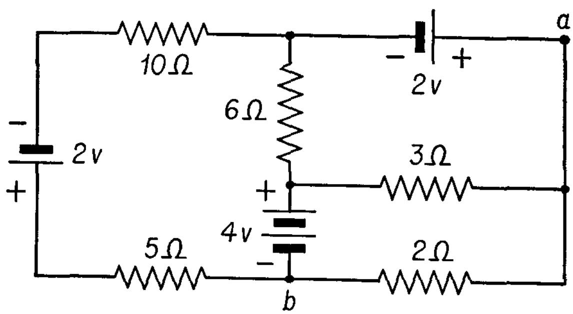


Fig. 27-17

27-20. Three 6-volt storage batteries are connected in parallel to a resistor of 100 ohms. The internal resistances of the batteries are 1 ohm, 3 ohms, and

10 ohms, respectively. (a) Find the heat generated in the 100-ohm resistor. (b) Find the current delivered by each of the storage batteries.

27-21. In the circuit of Figure 27-17, find the current through the 6-ohm resistor and the difference in potential $V_b - V_a$ between the points a and b.

27-22. A d-c generator having an emf of 110 volts and an internal resistance of 5 ohms is used to charge a 60-volt bank of storage batteries, whose internal

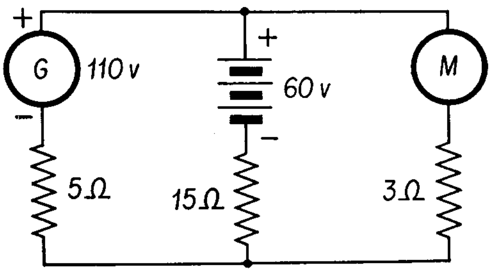


Fig. 27-18

resistance is 15 ohms, and to drive a motor whose back emf is 85 volts and whose internal resistance is 3 ohms, as shown in Figure 27-18. (a) Find the current supplied by the generator. (b) Find the charging current delivered to the batteries. (c) Find the mechanical power delivered by the motor.

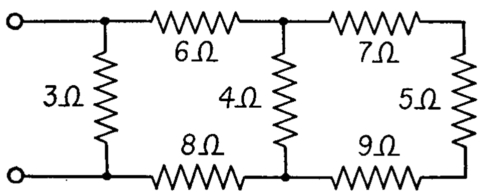


Fig. 27-19

27-23. Find the single resistance equivalent to the network of Figure 27-19.

28

Electrical Conduction in Liquids and Solids

28-1 Conductivity of Liquids. Electrolytes

We have seen that electrical conduction in solids is associated with the drift of free electrons in the solid, a process which will be examined in greater detail in Section 28-4. A similar mechanism may be used to account for the conduction of electricity in liquid metals. In other liquids electricity is conducted by the migration of positive and negative *ions* through the liquid.

When two terminals, or *electrodes*, are immersed in a liquid and a source of emf is connected to them, there will be a current through the liquid. Under the influence of the electric field established between the positive terminal, or *anode*, and the negative terminal, or *cathode*, the ions drift through the liquid. In general, if n_+ is the number of positive ions per unit volume moving with an average drift velocity v_+, and n_- is the number of negatively charged ions per unit volume moving with a drift velocity v_-, we find the current density J in the liquid to be

$$J = n_+q_+v_+ - n_-q_-v_-.$$

The drift of the positive ions is in the direction of the current, while the drift of negative ions is opposite to the direction of the current.

The electrical conductivity of a liquid depends upon the number of ions per unit volume and upon their drift velocity. The drift velocity of an ion varies with the electric field intensity, with the mass of the ion, and with other factors as well. Thus the electrical conductivity of different liquids may be expected to have widely different values.

Oily substances are very poor conductors of electricity, with conductivity of about 10^{-10} mho/m. Such oils find application as insulating oils in electrical apparatus. Pure solvents such as water or alcohol are relatively poor conductors, with a conductivity of about 10^{-4} mho/m. The conduc-

tivity of solutions of chemical salts in water is much higher, being about 10 mho/m. Such solutions are generally classed as good conductors and are called *electrolytic* solutions. The dissolved substance, or *solute*, is known as the *electrolyte.* While the conductivity of electrolytes is quite high compared to the conductivity of the pure solvent, their conductivity is small compared to a metallic conductor such as copper whose conductivity is about 10^8 mhos/m. The electrical conductivity of different liquids is shown in Table 28-1.

TABLE 28-1 ELECTRICAL CONDUCTIVITY OF LIQUIDS

Substance	Conductivity (mhos/m)	
KCl solution (0.1n)	1.05	(at 15°C)
NaCl solution (saturated)	20.14	(at 15°C)
Ethyl alcohol	3.3×10^{-4}	(at 15°C)
Paraffin oil	10^{-14}	
Distilled water	2×10^{-4}	(at 18°C)
Mercury	1.04×10^{6}	(at 20°C)

Our present understanding of the conduction of electricity by electrolytic solutions indicates that electrolytes are materials which are held together by strong electrical forces in the solid state. In crystalline sodium chloride, positively charged sodium ions and negatively charged chlorine ions occupy adjacent places in the crystal lattice. These ions are firmly anchored in position in the solid phase. Solid sodium chloride is a nonconductor of electricity. At high temperatures the ions have sufficient thermal energy to overcome the attractive forces holding them in place, and the substance melts. Molten sodium chloride is a good conductor of electricity, having a conductivity about one twentieth that of a solution of sodium chloride in water. The sodium ions and the chlorine ions in molten sodium chloride drift under the action of the electric field. The sodium ions migrate toward the cathode, and the chlorine ions migrate toward the anode, so that pure sodium and pure chlorine may be separated in an electrolytic cell. Free electrons do not traverse the cell.

When sodium chloride is placed in water, there are strong forces of attraction between the water molecules and the ions of sodium and chlorine. The water molecule is highly polar. The positive or hydrogen end of the water molecule tends to attach itself to the negative chlorine ion, while the negative or oxygen end of the water molecule tends to attach itself to the positive sodium ion, as shown in Figure 28-1. A shell of water molecules tends to surround each ion, and we say that the ions are *hydrated* in solution. This shell of water molecules tends to separate the sodium and

chlorine ions and to reduce the electrical force of attraction between them, both as a result of the physical separation and as a result of the reduction in electric field intensity of a point charge in a medium of high dielectric constant (see Section 25-7).

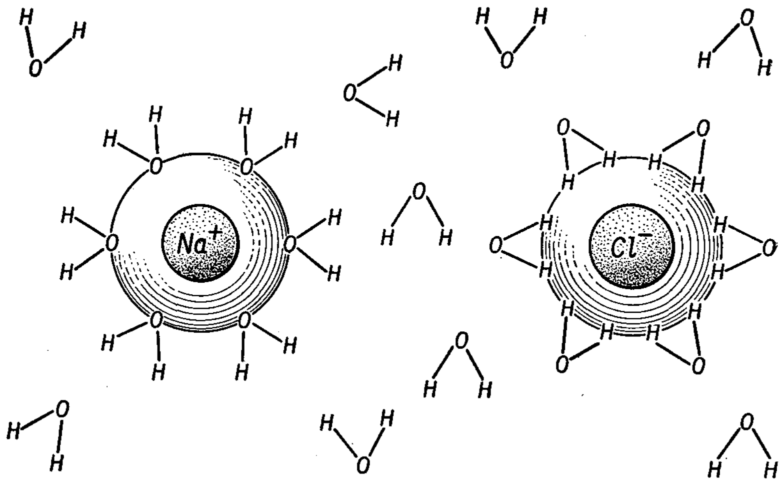


Fig. 28-1 Sodium and chlorine ions and associated water molecules.

It is interesting that ionic substances dissolve only in polar solvents. The process of solution is evidently furthered by the manner in which the molecules of the solvent are able to attach themselves to the ions of the solute. Ionic substances are relatively insoluble in nonpolar solvents, such as benzene. When electrodes are immersed in an electrolytic solution, the ion and its associated water of hydration drift as a unit through the liquid, so that the effective mass of the ion is the mass of the entire assemblage.

28-2 Faraday's Laws of Electrolysis

The flow of electricity through an electrolytic solution is associated with the transport of ions through the solution. If two silver electrodes are immersed in a solution of silver nitrate in water, it is observed that the mass of the anode diminishes as current is passed through the cell, while the mass of the cathode increases. The total mass of the two electrodes remains constant; that is, the effect of the passage of current is to transport silver from anode to cathode. Silver atoms from the anode go into solution as ions at the anode, replacing those ions which have left the solution as atoms at the cathode. Electric charges are thus transported between the electrodes. The concentration of the electrolyte in the solution remains

unchanged. When a cathode of some metal other than silver is used, this process results in *electroplating* silver onto the other metal.

The results of experiments with electrolytic cells can be summarized in terms of two laws first formulated by Faraday, which state:

(1) *The mass of any substance liberated from the solution is proportional to the total quantity of charge passing through the circuit.*

(2) *If the same quantity of electric charge passes through a series of cells, the masses of the elements deposited or liberated are in the ratio of their respective chemical equivalents.*

The *chemical equivalent* of an element is defined as the atomic weight divided by the valence. The direct implication of Faraday's laws is that the flow of electric charge through a cell is always accompanied by the movement of ions, and that no charge is transported through the solution by the motion of free electrons.

Faraday's two laws of electrolysis can be combined into the following single equation:

$$\boxed{M = Q\frac{A}{v}\frac{1}{\mathcal{F}},} \tag{28-1}$$

where M is the mass of an element deposited or liberated by the transfer of a quantity of electricity Q through an electrolytic cell, A is the atomic weight of the element, v is its valence, and $\mathcal{F}$ is a constant known as the *faraday*. The faraday is equal to the quantity of electricity transferred through the cell in depositing or liberating a gram chemical equivalent of an element. The gram chemical equivalent is the number of grams numerically equal to the chemical equivalent.

The best experimental value of the faraday is

$$\mathcal{F} = 96{,}496 \text{ coul/gm atomic wt},$$

using the chemical scale of atomic weights in which the atomic weight of oxygen is taken as 16. The faraday has a slightly different value on the physical scale of atomic weights in which the atomic weight of a particular isotope of oxygen containing 16 nucleons (8 protons and 8 neutrons) is taken as 16.

It is possible to make use of Faraday's laws of electrolysis to measure the charge transferred in a circuit by the passage of an electric current. Thus, a silver *coulombmeter*, an electrolytic cell having silver electrodes and a solution of a silver salt as electrolyte, may be placed in an electric circuit, and the charge transferred may be determined by measuring the mass of silver plated onto the cathode or liberated from the anode.

The results of experiments in electrolysis enable us to determine the

charge carried by ions. The faraday represents the total charge transferred by the ions when a gram atomic weight of an element of unit valence is liberated from the solution. Since there are N_0 atoms in a gram atomic weight of an element, where N_0 is the Avogadro number, this means that N_0 ions, each carrying a charge e equivalent to the charge of an electron, are transferred in the process. We can write

$$\mathcal{J} = N_0 e. \tag{28-2}$$

The Avogadro number N_0 has been determined by other experiments and its presently accepted value is

$$N_0 = 6.023 \times 10^{23} \text{ atoms/gm atomic wt.}$$

Equation (28-2) therefore yields the value

$$e = 1.602 \times 10^{-19} \text{ coul}$$

for the electronic charge.

The above procedure is one of the most accurate methods for the determination of the charge of the electron. The first demonstration of the unique value of the electronic charge was made by Robert Andrews Millikan, who measured the charge on small droplets of oil suspended between two horizontal brass plates, by studying the speed with which the charged droplets fell through air under the combined action of gravitation and the imposed electric field between the two plates.

28-3 Electrochemical Cells

In an electrochemical cell chemical energy is transformed into electrical energy. When electrodes of two different materials are immersed in an electrolyte and no current is permitted to flow to any external circuit, it is found that there is a potential difference between them, called the emf of the cell. When a metallic electrode is immersed in an electrolyte, there is a tendency for ions of the metal to go into solution, and, since ions formed by a metal are generally positively charged, the electrode becomes negatively charged. The negatively charged electrode tends to attract ions onto itself as a precipitate. Equilibrium is reached when the number of ions leaving the electrode to go into solution in a given time interval is equal to the number attracted to the electrode in that time interval. The charge on the electrode when equilibrium is reached depends upon the chemical activity between the electrode and the solvent. Since chemical activity varies with temperature, the charge on the electrode, and hence its potential, vary with temperature.

The potential acquired by an electrode cannot itself be measured without some reference potential. It is customary to measure the potential

difference between a particular electrode and a *hydrogen electrode.* The hydrogen electrode consists of a piece of platinum foil, coated with finely divided platinum, immersed in water. Hydrogen gas is bubbled over the platinum electrode and tends to collect at the surface of the platinum. An electrode potential is reached which is a measure of the tendency for hydrogen ions to accumulate in the solution. The potential of this reference

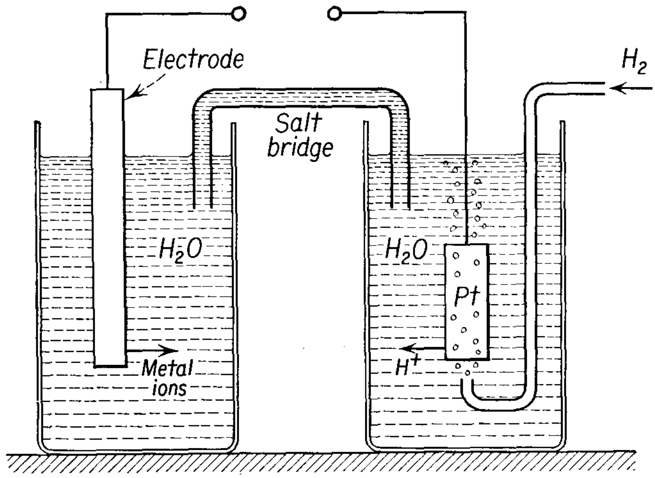


Fig. 28-2 Cell made of a metal electrode and a hydrogen electrode for measurement of electrode potentials.

electrode is arbitrarily assigned the value zero. The potentials of other electrodes are then determined by measuring the potential difference between the hydrogen electrode and another electrode when these are made part of a cell, as shown in Figure 28-2. The results of such measurements with different electrodes are shown in Table 28-2. The algebraic difference in potential between two values in the table then gives approximately the emf of a cell made up of the two electrodes.

The values of electrode potentials given in Table 28-2 represent the potential of a particular electrode, when in equilibrium with ions of that electrode, referred to the potential of a hydrogen electrode in equilibrium with hydrogen ions. The position of a substance in the table therefore represents the tendency of an electrode to go into solution. The more negative the electrode potential, the greater the tendency of the substance to go into solution. Since zinc has a more negative electrode potential than copper, we find that, if a zinc strip is placed in a copper sulphate solution, zinc ions will tend to replace copper ions in the solution, and the zinc strip will acquire a coating of copper.

We shall describe briefly only a few of the many practical cells in common use:

The *Daniell cell* consists of an amalgamated zinc electrode immersed in a solution of zinc sulphate and a copper anode immersed in a saturated solution of copper sulphate. The emf is approximately 1.06 volts, which may be compared to a value of 1.10 volts obtained from Table 28-2. The

TABLE 28-2 ELECTRODE POTENTIALS AT 25°C

Element	Reaction	Electrode Potential (volts)
Li	$e^- + Li^+ \rightarrow Li$	-3.05
Zn	$e^- + \frac{1}{2}Zn^{++} \rightarrow \frac{1}{2}Zn$	-0.76
Cd	$e^- + \frac{1}{2}Cd^{++} \rightarrow \frac{1}{2}Cd$	-0.40
H	$e^- + H^+ \rightarrow \frac{1}{2}H_2$	0.00
Cu	$e^- + \frac{1}{2}Cu^{++} \rightarrow \frac{1}{2}Cu$	$+0.34$
	$e^- + Cu^+ \rightarrow Cu$	$+0.52$
Hg	$e^- + \frac{1}{2}Hg_2{}^{++} \rightarrow Hg$	$+0.79$
Pd	$e^- + \frac{1}{2}Pd^{++} \rightarrow \frac{1}{2}Pd$	$+0.99$

zinc sulphate solution is of lower density than the copper sulphate solution, so that the two solutions remain apart by gravity, or may be separated by a porous diaphragm. The zinc electrode is negative in this cell.

One type of *dry cell* has a zinc cylindrical shell which forms the outer container of the cell and is the negative electrode. The positive electrode is a carbon rod which is inside, but which does not touch the cylinder. The dry cell is shown in Figure 28-3. The rest of the cell is filled with a paste of carbon powder, manganese dioxide, and ammonium chloride solution. The ammonium chloride solution acts as the electrolyte. The manganese dioxide reacts with the hydrogen liberated at the carbon anode when current is drawn from the cell. The top of the cell is sealed with a cement so that the cell does not leak. The emf of this cell is about 1.55 volts when new. The dry cell is capable of supplying small amounts of current, of the order of tens of milliamperes for steady use, and larger currents in intermittent use.

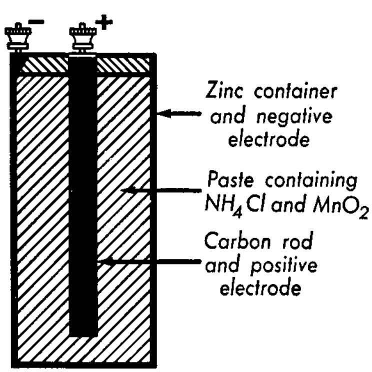


Fig. 28-3 Dry cell.

A cell which is widely used as a secondary standard of potential is the *Weston standard cell* shown in Figure 28-4. If the current drawn from the

cell is very small, the cell maintains a very stable emf. The Weston cell has an anode of mercury, in contact with a paste of mercurous sulphate. The cathode consists of a cadmium amalgam in contact with a layer of cadmium sulphate crystals. The electrolyte consists of a saturated solution of cadmium sulphate.

The *lead storage cell* has a positive electrode made of lead dioxide and a negative electrode of lead, with a dilute solution of sulphuric acid as the electrolyte. When the cell is delivering current to an external circuit, the

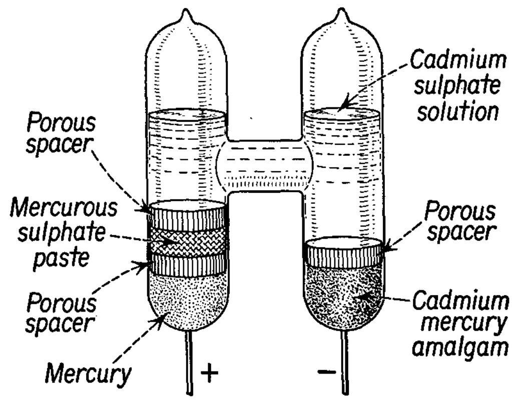


Fig. 28-4 Weston standard cell. (Courtesy of Weston Electrical Instrument Corporation.)

chemical reaction forms lead sulphate on both plates and makes the solution more dilute. The specific gravity of a fully charged lead cell is 1.285 when the emf of the cell is 2.1 volts. When the cell is discharged, the specific gravity of the electrolyte falls to 1.1, and the emf of the cell is about 1.8 volts. Thus a hydrometer may be used to measure the specific gravity of the electrolyte and to determine the state of charge of the cell. The chemical action in a storage cell is reversed when current flows through the cell in a direction opposite to the normal discharge direction. A storage cell may be repeatedly charged and discharged in normal operation.

28-4 Conduction in Solids

The most important feature of atomic structure is that there is a discrete set of energy levels in an atom. For our present purposes we may divide the electrons of an atom into two groups. We shall call one group of electrons the *core* electrons. These consist of electrons which are relatively close to the nucleus and which occupy energy levels in which they are tightly bound to the atom. The other group of electrons are called the *valence* electrons. These are the outermost electrons of the atom and are the least tightly bound to the atom. The valence electrons may be shared

by other atoms in the formation of molecules. In general, an atom may have up to 8 valence electrons. These valence electrons occupy energy levels which may be designated as *s* levels and *p* levels. The *s* levels are more tightly bound than the *p* levels. The energy separation of the levels of the different classes is of the order of several electron volts of energy.

No more than 2 electrons may be in any one *s* level and no more than 6 electrons may be in any one *p* level. An element in the first column of the periodic table has a valence of 1; this means that it has 1 electron outside a closed core. This valence electron is in an *s* level. Elements in the second column of the periodic table have 2 valence electrons. These fill the *s* levels outside the core. An element of valence greater than 2 will have 2 electrons in the *s* level and the additional electrons in the *p* level.

The energy levels of an atom are associated with the fact that the electric field experienced by the valence electrons has a well-defined center of attraction. Each valence electron is bound to its parent nucleus, as long as the atom is distant from all other atoms. When two different atoms combine to form a molecule, as in the case of molecular hydrogen H_2, the valence electrons of both atoms experience two centers of attraction, so that these electrons are shared by the two nuclei, and it is no longer possible to say which electron belongs to which nucleus. When a number of identical atoms are assembled in a periodic lattice structure to form a crystal, the valence electrons of the atoms of the crystal may no longer experience a well-defined center of attraction. In this case these electrons are sometimes free to drift through the crystal as *conduction* electrons.

We may gain additional insight into the problem of the conduction of electricity in solids by considering what happens to the energy levels of the atoms when the crystal is assembled. The presence of adjacent nuclei may alter the energy associated with the *s* levels and the *p* levels of an atom. The total number of levels in the system remains the same, but the position of the levels may be altered. The collection of *s* levels of the crystal may then be thought of as a continuous band, called the *s* band. In a similar way the *p* levels spread out into a *p* band. Depending upon the position of adjacent atoms in the crystal, these bands may remain far apart, they may be brought relatively close together, and, in some cases, they may even overlap, as shown in Figure 28-5.

If the atom is divalent, there are two *s* electrons in each atom. When the *s* band is relatively far from the *p* band in energy, all the *s* levels are occupied, and the electrons cannot move readily to the *p* band except by the addition of large amounts of energy. This band structure then describes an insulator, for the electrons are not free to alter their velocities. If the atom is monovalent, only half of the available *s* levels are occupied. The application of an electric field can influence the direction and magnitude of the velocity of the electrons, and the material is a conductor. Thus the

monovalent metals such as copper, silver, and so on, are good conductors of electricity.

In the general case of a metallic conductor, the bands overlap. Thus there are a large number of available energy levels, and the valence electrons of the crystal are free to alter their velocities in accordance with an applied electric field.

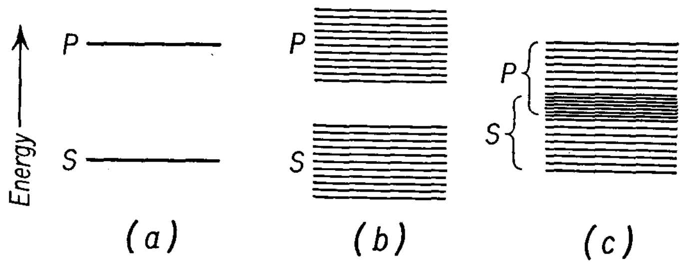


Fig. 28-5 (a) Energy levels of an atom. (b) Energy levels of a solid (nonoverlapping). (c) Overlapping energy bands of a conductor.

When the *s* band is only a few tenths of an electron volt distant from the *p* band, the energy associated with small changes in temperature may be sufficient to raise an electron from a filled *s* band to an unfilled *p* band. The resistivity of such a material is markedly affected by temperature. Such a material is an *intrinsic semiconductor*. An electron may be raised from the filled *s* band to the *p* band by absorbing energy from incident light. In this case the material is called *photoconductive*.

If there are distortions in the crystal lattice, or if there are impurities present, the structure of the bands may be altered. In the vicinity of these impurity sites, the level structure may be altered, and the material may become partially conducting, as in an *impurity semiconductor*. As an example, let us consider the effect of impurities in a germanium crystal. Atoms of germanium have 4 outer electrons each. If one atom of antimony which has 5 outer electrons replaces one atom of germanium, there will be 1 excess electron in this region. The energy-level structure is locally altered, and this electron may be set in motion by the application of a small electric field. A germanium crystal with this type of impurity is called an *n*-type crystal, the *n* standing for negative charge. On the other hand, if an atom of indium, which has only 3 outer electrons, replaces a germanium atom, there will be a deficiency of electrons, or a *hole* left in this part of the crystal. Again the energy-level structure is locally altered, and if an electric field is applied to such a crystal, electrons from other parts of the crystal flow toward the holes, leaving holes in other parts of the crystal. Thus the current in this type of crystal may be considered as the motion of holes.

A hole in a region which would be normally occupied by an electron is equivalent to the addition of a positive charge at this place. A germanium crystal with a type of impurity which produces holes is called a p-type crystal, where p stands for positive charge.

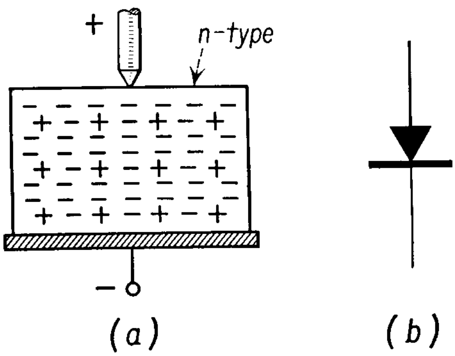


Fig. 28-6 (a) Point contact rectifier, direction of easy conductor. (b) Conventional symbol of rectifier; the arrow is in the direction of the forward current.

When a point contact is made between a metal wire and an n-type semiconductor, as shown in Figure 28-6(a), it is found that the resistance of the contact is comparatively low when the wire is made positive relative to the semiconductor, but the resistance is considerably higher when the wire is negative. In such a semiconductor there are relatively few energy levels available to the conduction electrons. When the wire is made positive, electrons from the semiconductor are attracted to the wire, so that there is a continuous supply of available levels in the vicinity of the contact for other electrons in the semiconductor to occupy. These electrons diffuse toward the contact point, establishing a current sometimes called the forward current. When the wire is made negative, electrons are injected from the wire into the available levels in the semiconductor in the vicinity of the contact point. All available levels become occupied, and additional electrons can no longer flow into the semiconductor until the excess of electrons in the vicinity of the contact point has diffused through the body of the semiconductor into the opposite terminal. Thus the resistance of the contact depends upon the direction of the current. Such a point contact may be used as a *diode* or *rectifier;* the conventional symbol is shown in Figure 28-6(b); the arrow shows the direction of the forward current.

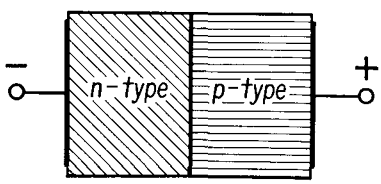


Fig. 28-7 p–n junction; direction of easy conduction.

In a similar way, when an n-type semiconductor is placed in contact with a p-type semiconductor, as shown in Figure 28-7, the direction of easy

electron flow is from n to p, that is, when the p-type semiconductor is made positive. Once again, when electrons flow from the p to the n material, the available energy levels in the n material in the vicinity of the contact become occupied, and no additional electrons may flow until these diffuse away.

28-5 Thermoelectric Effects

One thermoelectric effect with which we are already familiar is the liberation of heat due to the passage of electric current through a conductor. This is an irreversible process, for heating a conductor to uniform temperature does not generate an electric current. Besides the Joule effect, there are several reversible thermoelectric effects of interest.

The most frequently applied of these effects is called the *Seebeck effect,* after its discoverer Thomas Johann Seebeck (1770–1831). If wires of two

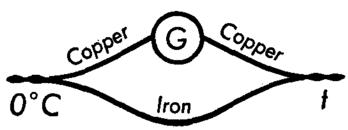


Fig. 28-8 Thermocouple circuit.

dissimilar metals are joined at their ends, and these ends are maintained at different temperatures, a current may be observed in the wires of the circuit. It is possible to analyze the current in terms of a *thermal emf* developed at the junctions of the wires and the resistance of the circuit. In general, the interposition of a third metal into the circuit does not affect the thermal emf if the third metal is maintained at a constant temperature.

The primary application of the Seebeck effect is in the measurement of temperature through a device called a *thermocouple.* In practice, one junction of a thermocouple is kept at the temperature of melting ice, while the other junction is immersed in the medium whose temperature is being measured. In some applications a galvanometer is inserted into the thermocouple circuit, as shown in Figure 28-8, but more generally the thermal emf is measured with a potentiometer, and the unknown temperature is determined by comparison with a calibration chart. This application of the potentiometer is one of its principal uses in industry, where the measurement of temperature is of great importance in metallurgy, in chemical engineering, and so on.

Figure 28-9 shows the thermoelectromotive force as a function of the temperature of the hot junction when the cold junction is kept at 0°C. The thermoelectromotive force at first increases with increasing temperature until the point N, known as the neutral point, is reached, and then decreases as the temperature of the hot junction is increased still further

until the point I, known as the inversion point, is reached, when the emf reverses direction as the temperature is increased still further. The neutral point for a copper-iron thermocouple is about 260°C; it varies with the purity of the metals.

When there is a current in a thermocouple circuit because of a difference in temperatures at the two junctions, heat is absorbed at the higher-temperature junction, and a smaller quantity of heat is emitted at the cooler junction, the difference between the two quantities of heat being converted into electrical energy.

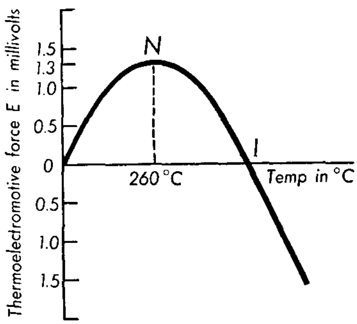


Fig. 28-9 Thermoelectromotive force of a copper-iron thermocouple as a function of the temperature of the hot junction.

Thermocouples can be made of any combination of metals. The electromotive force $\mathcal{E}$ of a thermocouple depends upon the nature of the two metals used and the temperatures of the junctions. If one junction is kept at a fixed temperature and the other at any other temperature T, the emf of the thermocouple will be given, to a very close approximation, by the equation

$$\mathcal{E} = a + bT + cT^2, \tag{28-3}$$

where a, b, and c are constants for the particular thermocouple and have to be determined experimentally from the measurements at known temperatures. Once these constants have been determined, the thermocouple may be used as a thermometer. A thermocouple thermometer, one wire of which is made of platinum and the other of an alloy of 90 per cent platinum and 10 per cent rhodium, is used for measuring temperatures on the international scale in the range from 660°C to 1063°C. Special alloys have been developed for use in thermocouples. One of the advantages of a thermocouple thermometer is that only one junction need be heated, and since this can be made very small, only very small quantities of heat are needed to produce a measurable effect. Because of its small size and mass, the thermocouple junction will follow very rapid changes in temperature.

More recently, there has been considerable research into the possibility of exploiting the Seebeck effect as a means of generating electrical energy directly from heat energy without the use of rotating machinery. One unit, developed in 1954, was capable of converting solar energy into electrical energy with an efficiency of 3.35 per cent, as compared to an efficiency of about 5.8 per cent for a conventional gasoline-powered 1-kw generator.

A second thermoelectric effect was discovered by Jean C. A. Peltier (1785–1845). The *Peltier effect* is the inverse of the Seebeck effect. When there is a current through the junction of two dissimilar metals, the temperature of the junction changes. An emf is developed at the junction as electrons from one metal diffuse into the other to fill vacant energy levels in the other metal. When the current is in the direction of the emf, the temperature of the junction falls, while if it is in the opposite direction, the temperature of the junction rises. This effect is superimposed upon the usual Joule heating of a conductor, and, in spite of the Joule heating, it is possible to cool a suitably designed junction to a temperature lower

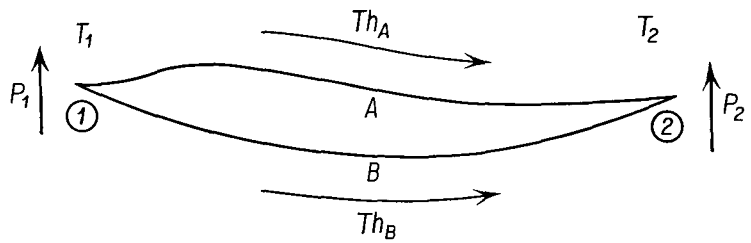


Fig. 28-10 Peltier emf's P_1 and P_2 at the junctions 1 and 2 respectively, and the Thomson emf's Th_A and Th_B in the wires A and B of a thermocouple. The junctions are maintained at temperatures T_1 and T_2 respectively.

than that of the surroundings. The Peltier effect in an arrangement of wires of two dissimilar metals similar to a thermocouple may be thought of as an electrically operated heat pump, albeit a very inefficient one at present.

A third thermoelectric effect, called the *Thomson effect,* was predicted from thermodynamic considerations by Sir William Thomson (Lord Kelvin, 1824–1907). When different parts of the same metallic conductor are maintained at different temperatures, differences of potential may be observed in the conductor. The Thomson emf's are of the order of millivolts (abbreviated mv). The Thomson effect may be interpreted on the assumption that the free, or conduction, electrons within a metal behave like the molecules of a gas. Like the molecules of a gas, the electron density is highest where the temperature is lowest. The low-temperature portions of the conductor become negatively charged and thus are at lower potentials with respect to regions of the conductor at higher temperature.

Thermoelectric effects occur in all conductors. In electrical apparatus of high sensitivity, it is desirable to avoid temperature differences in different parts of the apparatus to avoid difficulties associated with the thermal electromotive force.

The three thermal emf's discussed above are not independent of each

other. When the junctions of two wires of dissimilar metals are maintained at different temperatures T_1 and T_2, respectively, there are Thomson emf's in both wire A and wire B; these are designated as Th_A and Th_B respectively, as shown in Figure 28-10. There are Peltier emf's P_1 and P_2 at the junctions 1 and 2. The Seebeck emf $\mathcal{E}$ is the sum of the emf's around the thermocouple. Thus

$$\mathcal{E} = P_1 + Th_A - P_2 - Th_B.$$

Problems

28-1. Determine the mass of silver which is deposited on a cathode from a silver nitrate solution by a current of 2 amp in 15 min.

28-2. When a current is passed through water containing a small amount of sulphuric acid to make it conducting, the water is decomposed into hydrogen and oxygen. (a) How large an electric charge must pass through the solution to liberate 1 gm atomic wt of oxygen? (b) How much hydrogen will be liberated at the same time? The chemical formula for water is H_2O.

28-3. A dry cell having an emf of 1.55 volts and an internal resistance of 0.08 ohm supplies current to a 4-ohm resistor. (a) Determine the current in the circuit. (b) Calculate the terminal voltage of the cell.

28-4. A lead storage battery whose emf is 6.3 volts, when being charged with a current of 12 amp, has a terminal voltage of 7.2 volts. (a) Determine the internal resistance of the battery. (b) How much energy is supplied to the battery in 30 min? (c) How much of this energy is converted into heat? (d) What becomes of the rest of this energy?

28-5. Two storage cells having the same emf of 2.1 volts, but having different internal resistances of 0.2 and 0.4 ohms, respectively, are connected in parallel. The parallel combination of cells is connected to a 10-ohm resistor. Determine (a) the current through the resistor and (b) the terminal voltage across either cell.

28-6. When a pure metal is immersed in an electrolyte, ions of the metal form in the solution until equilibrium is reached and the electrode attains its equilibrium electrode potential. Explain why a commercial metal such as iron or aluminum continues to corrode when it is immersed in salt water.

28-7. The emf of a silver-platinum thermocouple when one junction is at 0°C and the other at 100°C is 0.74 mv; when the second junction is at 200°C, the emf is 1.77 mv. (a) Determine the constants of Equation (28-3). (b) Calculate the temperature of the second junction when the emf is 1.30 mv.

28-8. A Pt—Pt-10 per cent Rh thermocouple is used as a standard for the determination of temperature. The following table gives the emf in millivolts when one junction is at 0°C and the other junction is at the temperature given in the table. (a) Plot a graph of emf against temperature. (b) Determine the temperature of the second junction when the emf is 5.00 mv. (c) Determine the constants of Equation (28-3) for this range of temperatures.

emf	0.643	1.436	2.316	3.251	4.221	5.224	6.260	7.329	8.432	9.570
temp (°C)	100	200	300	400	500	600	700	800	900	1000

29

The Magnetic Field

29-1 Natural and Permanent Magnets

Natural magnets, called lodestones, have been known since ancient times. The lodestone, a magnetic oxide of iron called magnetite (Fe_3O_4), was mentioned by Thales of Miletus. By the eleventh century the magnetic compass was known to the Chinese, and in the twelfth century references to

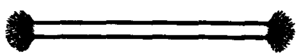


Fig. 29-1 Iron filings cling to the poles of a bar magnet.

the compass were made in Western Europe. The lodestone is capable of attracting pieces of iron and of imparting permanent magnetism to other pieces of iron so that these too could attract iron filings. If an iron bar is magnetized, as the result of being near a piece of lodestone, and is then dipped into iron filings, the filings will cling mostly to the ends of the bar, as shown in Figure 29-1. These ends are called the *poles* of the magnet.

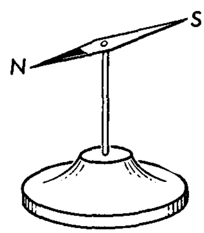


Fig. 29-2 A simple magnetic compass needle.

The use of the magnet as a compass depends upon the fact that if a permanent magnet is suspended so that it can swing freely in a horizontal plane, the magnet will set itself in nearly a north-south direction. The end which points toward the north is called the north-seeking pole, or simply the *north pole*, while the other end is called the *south pole*. A line joining the south to the north pole is called the *axis* of the magnet. A simple magnetic compass is shown in Figure 29-2. When such a compass is mounted upon a sheet of cork and is floated in water, the cork sheet remains at rest, implying that the net force on the compass is zero.

A ship's compass, a somewhat more elaborate instrument, is shown in Figure 29-3. It consists of a compass card fastened to a set of parallel magnets with their north poles facing in the same direction and mounted

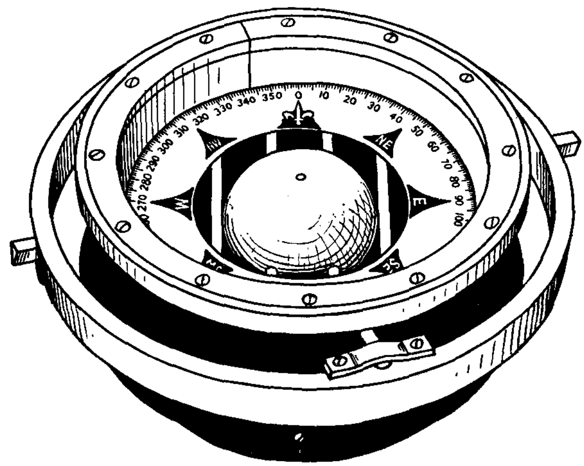


Fig. 29-3 A ship's compass. The compass card is mounted on four parallel magnets.

so that the whole assembly rotates freely in a horizontal plane. The case housing the compass is mounted in a set of bearings so that it can remain horizontal even when the ship rolls and pitches. A marker on the case housing the compass shows the direction in which the ship is headed.

29-2 Coulomb's Law of Force

Suppose that a bar magnet is suspended from its center by a string so that it can swing freely in a horizontal plane. If the north pole of a second bar magnet is brought near the north pole of the first one, there will be a force of repulsion between these poles. The suspended magnet will move so that its north pole goes away from the north pole of the approaching magnet. On the other hand, there will be a force of attraction when the north pole of one magnet approaches the south pole of the other one. In his book, *De Magnete*, William Gilbert (1540–1603) first showed that the earth could be considered as a huge spherical magnet with two poles, and that the compass needle was oriented by the forces exerted by the earth's magnetism.

About two centuries later Charles Augustin de Coulomb (1736–1806) measured the law of force between the poles of two magnets. A long thin magnet was suspended from a fine wire, and the pole of another thin magnet was brought near a like pole of the suspended magnet, as shown in Figure

29-4. Because of the force between these two poles, the suspended magnet experienced a torque which caused it to rotate about the wire as an axis, thus twisting the wire. Coulomb studied the force between the poles of two magnets as a function of the distance between them; he also studied the force between magnets of different strengths, keeping the distance between poles constant. He found that the force between two poles of two magnets varied inversely with the square of the distance between them, and varied directly with the strengths of the poles. The results of Coulomb's experiments could be summed up by stating that *the force between the two magnetic poles is proportional to the product of the strengths of the poles and inversely proportional to the square of the distance between them*, or

$$F = k_1 \frac{p_1 p_2}{r^2}, \qquad \textbf{(29-1)}$$

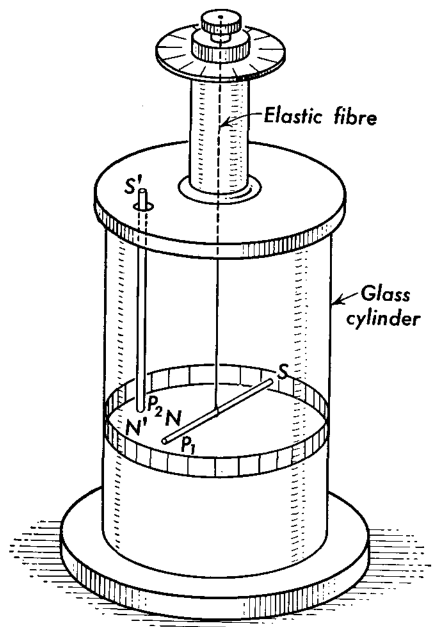


Fig. 29-4 Coulomb type torsion balance. NS, magnet of pole strength p_1, is suspended from elastic fiber. $N'S'$ is another magnet of pole strength p_2.

in which p_1 is the pole strength of the first magnet, p_2 is the pole strength of the second magnet, r is the distance between the poles, k_1 is a constant of proportionality, and F is the force between them. The force between the two poles is directed along the line joining the two poles; it is attractive between unlike poles and repulsive between like poles.

Following the analogy of Coulomb's law for electric charges, we may write this proportionality in the form of an equation, but we must first define either the constant of proportionality k_1 or the magnitude of the unit pole. As before, in the *cgs* system of units, the constant of proportionality k_1 is set equal to 1, when the medium is vacuum, and the subsequent equation is used to define a *unit pole*. The units based upon this equation are then called the *cgs electromagnetic units* (abbreviated emu). In the form of an equation, we have

cgs emu $$\mathbf{F}_2 = \frac{p_1 p_2}{r^2} \mathbf{1}_r \qquad \textbf{(29-1a)}$$

for the force in dynes between two poles in vacuum (or in air for which this equation is very nearly correct). In Equation (29-1a) $\mathbf{F}_2$ is the force on the pole p_2 in dynes, r is the distance between the poles in centimeters, and the unit vector $\mathbf{1}_r$ is directed from the pole p_1 to the pole p_2. The pole strength is expressed in unit poles, writing a north pole as positive and a south pole as negative. *A cgs unit pole is* implicitly *defined* in Equation (29-1a) *as one which, placed one centimeter from a like pole in vacuum, will repel it with a force of one dyne.*

In the mks system of units, the constant of proportionality is written as $\frac{1}{4\pi\mu_0}$, so that Coulomb's law becomes

mks units
$$\boxed{\mathbf{F}_2 = \frac{1}{4\pi\mu_0}\frac{p_1p_2}{r^2}\mathbf{1}_r.} \tag{29-1b}$$

The constant μ_0 is assigned the value $4\pi \times 10^{-7}$, for reasons which must be deferred until a subsequent chapter. Equation (29-1b) implicitly defines an *mks unit pole* called a *weber.* In Equation (29-1b) $\mathbf{F}_2$ is the force on pole p_2 expressed in newtons, p_1 and p_2 are expressed in webers, and r is expressed in meters. *An mks unit pole of strength one weber is one which, placed one meter from a like pole in vacuum, will repel it with a force of* $10^7/(4\pi)^2$ *nt.* This is a force of $10^{12}/(4\pi)^2$ dynes. From Equation (29-1a) two cgs unit poles placed 1 m apart repel each other with a force of 10^{-4} dyne. Thus we see that

$$1 \text{ weber} = 10^8/(4\pi) \text{ cgs poles.} \tag{29-2}$$

In a form appropriate to Equation (29-1b) the value of μ_0 may be given as

$$\mu_0 = 4\pi \times 10^{-7}\,\frac{\text{weber}^2}{\text{nt m}^2}\cdot \tag{29-3}$$

As the theory of magnetism is developed, other dimensions will be stated for μ_0 which are equivalent to the dimensions given above.

The properties of a magnet and of magnetic poles appear to be similar to the properties of electric charge, and yet there are extremely significant differences. For example, when a long magnetized needle is broken in two, new poles appear at the broken ends, so that each piece of the needle is observed to contain a north pole and a south pole. *An isolated magnetic pole has never been observed in nature.* Poles always occur in pairs. In calculating the forces on a magnet due to the presence of other magnets, we must always take into account the forces due to the north and the south poles of the magnets generating the force.

In a Coulomb type of experiment however, it is possible to use magnets long enough so that when two poles are placed near each other, the other poles are far enough removed so that they do not appreciably affect the result. The pole of a magnet is simply a small region of the magnet which acts as a center of force rather than the location of isolated magnetic charge. In spite of the fact that we shall find it convenient to speak of the magnetic field in terms of magnetic poles, we must always bear in mind that the smallest element of magnetism observed physically is not a magnetic pole but a magnetic dipole.

29-3 Intensity of Magnetic Field

Because of the effect that a magnet produces on magnetic materials placed anywhere in its neighborhood, we can say that a *magnetic field* exists in the neighborhood of a magnet. We can make a quantitative statement of the intensity of the magnetic field at some point in space by placing the north pole of a long, thin, needlelike magnet at this point. If the pole strength of this magnet is p, its north pole will experience a force **F**. The *magnetic field intensity* **H** at this point is defined as the force **F** divided by the pole strength p, or, in the form of an equation,

$$\boxed{\mathbf{H} = \frac{\mathbf{F}}{p}.} \tag{29-4}$$

The *magnetic field intensity* **H** is a vector quantity, whose magnitude is given by the magnitude of the force on a unit north pole, and whose direction is the direction of the force on a north pole. Algebraically, it is conventional to represent a north pole as a positive pole and a south pole as a negative pole. In cgs emu the force is stated in dynes, the pole strength is cgs unit poles, and H is given in *oersteds*. If the intensity of the magnetic field H is known at a particular point in space, the force that would be experienced by each pole of a magnet placed at this point is given by Equation (29-4). For example, if the north pole of a long thin magnet having a pole strength of 2 cgs poles is placed at a point where the intensity is 45 oersteds, the force on it is 90 dynes.

Illustrative Example. A bar magnet 24 cm long has a pole strength of 800 cgs unit poles. Calculate the intensity of the magnetic field at point A 16 cm from the magnet, measured along the line which is the perpendicular bisector of the bar.

Referring to Figure 29-5, we note that the point A is 16 cm from the center C of the bar and 20 cm from each end, N and S. For purposes of calculation, let us imagine a long thin magnet of unit pole strength with its north pole at A,

and apply Coulomb's law to calculate the force exerted on it by each of the poles of the bar magnet. The force of repulsion on the unit north pole at A due to the north pole N is, from Equation (29-1a)

$$F = \frac{800 \times 1}{400} \text{ dynes} = 2 \text{ dynes}$$

directed away from N along NA. The force of attraction on the unit north pole at A due to the south pole S is also 2 dynes but is directed toward S along AS. Adding these two forces vectorially, we get 2.4 dynes acting parallel to the axis of the magnet as the force on a unit north pole at A. The intensity of the magnetic field at A is therefore

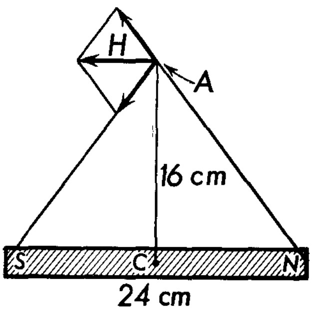


Fig. 29-5

$$H = \frac{2.4 \text{ dynes}}{\text{cgs pole}} = 2.4 \text{ oersteds.}$$

If the north pole is taken away and if a south pole is placed at A, it will experience an equal force in the opposite direction. A magnet so small that it can be considered as though both poles are practically at A will line up parallel to the direction of the field at A, with the north pole pointing in the direction of the field. A magnetic field may therefore be mapped by small compass needles, which will orient themselves parallel to the direction of the field.

When iron filings are introduced into a magnetic field, the filings become magnetized. The magnetized filings orient themselves in the direction of the field. Figure 29-6 shows a typical arrangement of iron filings in the neighborhood of a bar magnet as seen in one plane.

As we shall see in a subsequent chapter, a magnetic field may also be generated by a wire carrying current, and, in fact, this is the most common method for the generation of a magnetic field. This has been recognized in the choice of the units of magnetic field intensity H in the mks system where H is generally stated in units of *amperes per meter*. Until we have developed the concepts appropriate to the generation of the magnetic field by current in a wire, we shall represent the magnetic field in units of *newtons per weber* in the mks system of units, a form appropriate to the definition of H in Equation (29-4). The relationship between the mks and cgs units of the magnetic field intensity H may be stated as

$$1 \frac{\text{nt}}{\text{weber}} = 4\pi \times 10^{-3} \text{ oersted.} \tag{29-5}$$

This conversion factor may be verified by recalling that

$$1 \text{ nt} = 10^5 \text{ dynes,}$$

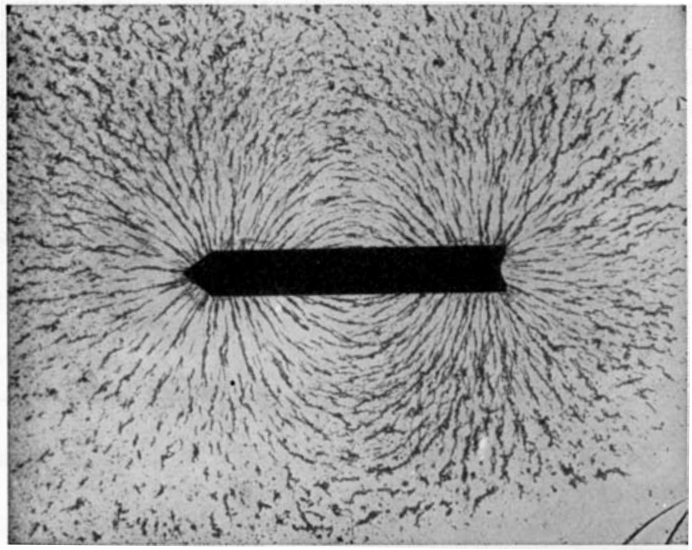


Fig. 29-6 Pattern formed by iron filings in the neighborhood of a bar magnet.

and
$$1 \text{ weber} = \frac{10^8}{4\pi} \text{ cgs poles.}$$

Thus
$$1 \frac{\text{nt}}{\text{weber}} = 1 \frac{\text{nt}}{\text{weber}} \times \frac{10^5 \text{ dynes}}{1 \text{ nt}} \times \frac{1 \text{ weber}}{10^8/4\pi \text{ cgs pole}}$$
$$= 4\pi \times 10^{-3} \frac{\text{dyne}}{\text{cgs pole}}$$
$$= 4\pi \times 10^{-3} \text{ oersted.}$$

29-4 Magnetic Moment

If a bar magnet of length s and pole strength p is placed in a uniform magnetic field of intensity H, its north pole will experience a force Hp in the direction of the field, while its south pole will experience an equal force in the opposite direction, as shown in Figure 29-7(a). If the magnet is not parallel to the field, the magnet will experience a torque tending to rotate it so that its alignment is parallel to the field. The magnet will experience two forces equal in magnitude but opposite in direction, so that the net force on the magnet is zero.

Following the terminology developed in electrostatics where we called an assemblage of two equal and opposite charges an electric dipole, we call a magnet with its two equal and opposite poles a *magnetic dipole.* The *magnetic dipole moment* $\mathbf{m}$ is a vector quantity directed from the south pole

of the magnet to its north pole and is of magnitude given by the product of the pole strength p by the separation of the poles s. In the form of an equation the magnitude of the dipole moment is given by

$$m = ps. \tag{29-6}$$

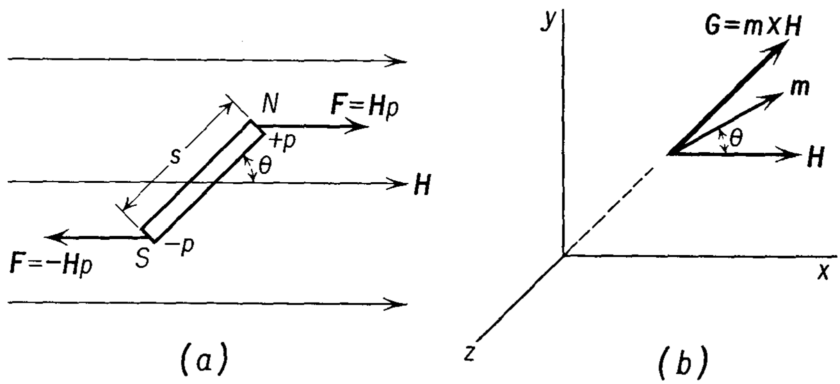


Fig. 29-7 (a) Forces on a bar magnet in a uniform magnetic field. The plane of the paper is the x-y plane. (b) When the magnetic moment vector **m** and the magnetic field **H** lie in the x-y plane as shown, the torque **G** on the magnet is parallel to the z axis.

If **s** is the displacement vector directed from the south pole of the magnet to its north pole, the magnetic dipole moment vector **m** is given by the equation

$$\boxed{\mathbf{m} = p\mathbf{s},} \tag{29-6a}$$

which may be compared to the corresponding definitions in electricity given in Equation (25-11). The cgs electromagnetic units of magnetic dipole moment are *cgs pole centimeters*, while the mks units of dipole moment may be expressed as *weber meters*.

Let us calculate the magnitude of the torque G about the center of the magnet due to the effect of the magnetic field **H** upon the magnetic dipole **m**, as shown in Figure 29-7(b). If the dipole-moment vector makes an angle θ with **H**, the torque is

$$G = Hps \sin \theta,$$

or, in terms of the magnetic moment m as defined in Equations (29-6), the torque is given by

$$G = Hm \sin \theta. \tag{29-7}$$

As shown in the figure, the torque tends to rotate the magnet in the clockwise direction. Following the right-hand rule, the torque vector is in the

$-z$ direction. Using the notation of the vector product developed in Section 7-4, we may write

$$\boxed{\mathbf{G} = \mathbf{m} \times \mathbf{H}.} \tag{29-7a}$$

We may apply this result to the measurement of the magnetic field. If a small bar magnet of moment of inertia I is suspended in a magnetic field H, the torque on the magnet will be given by Equation (29-7), and if the angle made by the magnet with the field direction is small, we may replace $\sin\theta$ by θ. Observing that the direction of the torque vector is opposite to the direction of the angular displacement, we obtain

$$G = -Hm\theta \tag{29-8}$$

By comparing Equation (29-8) with Equation (12-2), we see that the magnet is subjected to an elastic restoring torque which is proportional to the angular displacement θ and opposite in direction. The magnet will therefore vibrate in simple harmonic motion. The period of the motion T will be given by Equation (12-17) as

$$T = 2\pi\sqrt{\frac{I}{mH}}\cdot \tag{29-9}$$

Thus the magnetic field intensity H may be determined by measuring the period of vibration, T of a magnet of known magnetic moment M and moment of inertia I.

Illustrative Example. A bar magnet 8 cm long has a mass of 20 gm and a pole strength of 60 cgs unit poles. Determine (a) its magnetic moment and (b) its period when placed in a field whose intensity is 100 oersteds.

(a) The magnetic moment of the bar magnet is, from Equation (29-6),

$$m = 60 \text{ poles} \times 8 \text{ cm} = 480 \text{ pole cm}.$$

To determine the period it is necessary to find the moment of inertia. The moment of inertia of a uniform bar about an axis through its center of gravity is given by the expression

$$I = \tfrac{1}{12}ML^2,$$

where M is the mass of the bar and L is its length. Therefore

$$I = \tfrac{1}{12} \times 20 \text{ gm} \times 64 \text{ cm}^2 = 106.7 \text{ gm cm}^2.$$

(b) Substituting these values in Equation (29-9), we find

$$T = 2\pi\sqrt{\frac{106.7}{480 \times 100}} \text{ sec} = 0.296 \text{ sec}.$$

29-5 Atomic Magnetic Moments

When a small magnet is placed in a uniform magnetic field, it experiences a torque. When the magnet is placed in a *nonuniform* field, the field acting at the south pole of the magnet may be given by $\mathbf{H}_s$, while that acting at the north pole of the magnet may be given by $\mathbf{H}_n = \mathbf{H}_s + \Delta\mathbf{H}$, as shown in Figure 29-8, where $\Delta\mathbf{H}$ is the vector difference between the fields at these two poles. The net force on the magnet is due to the change in the field $p\,\Delta\mathbf{H}$, for if p is the pole strength of the magnet, the force on the magnet is

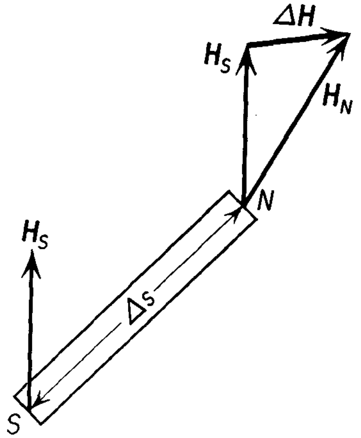


Fig. 29-8

$$\mathbf{F} = p\mathbf{H}_n - p\mathbf{H}_s = p\,\Delta\mathbf{H}.$$

Let us suppose that the magnet is physically small, of length Δs. We multiply and divide the right-hand side of the equation above by the quantity Δs to find

$$\mathbf{F} = p\,\Delta s\,\frac{\Delta\mathbf{H}}{\Delta s}\,.$$

The product $p\,\Delta s$ is the magnetic moment m of the magnet, while the quantity $\Delta\mathbf{H}/\Delta s$ is the rate of change of the magnetic field directed from the south to the north pole. Thus we have, in the limit of extremely small magnets,

$$\mathbf{F} = m\,\frac{\Delta\mathbf{H}}{\Delta s} = m\,\frac{d\mathbf{H}}{ds}\,, \tag{29-10}$$

where the direction in which we take the derivative of the magnetic field intensity is the direction of the magnetic moment.

A method of measuring atomic magnetic moments devised by Stern and Gerlach (1921) was to generate a beam of atoms, to send them through a nonuniform or inhomogeneous magnetic field of known inhomogeneity, and to observe the deflection of the beam. In the Stern-Gerlach experiment a beam of silver atoms was obtained by heating silver to a high temperature in an oven in vacuum. A narrow beam of silver atoms coming from an oven O, Figure 29-9, after passing through defining slits S_1 and S_2, was allowed to pass through an inhomogeneous magnetic field and strike a plate P. The inhomogeneous field was produced by an electromagnet with specially designed pole pieces. One pole piece was in the form of a knife-edge, while the other pole piece had a channel cut in it parallel to the

knife-edge. From the pattern on the photographic plate, shown in Figure 29-10, and a knowledge of the speed of the particles, the magnetic moment

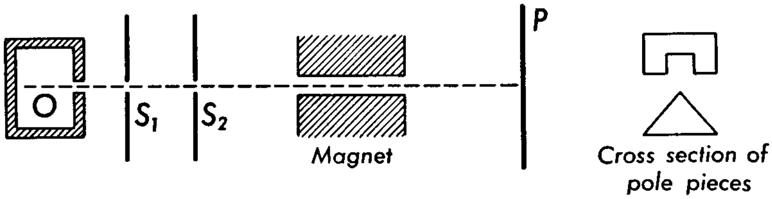


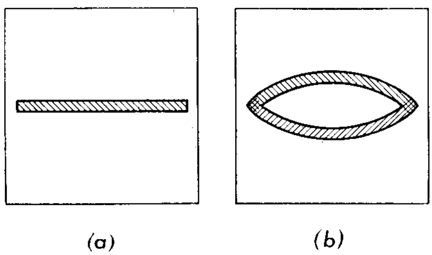
Fig. 29-9 Arrangement of apparatus in Stern-Gerlach experiment.

of silver could be calculated. The magnetic dipole moment of silver, m_{Ag}, is now known to be

$$m_{Ag} = 0.93 \times 10^{-20} \text{ emu.}$$

It may be remarked parenthetically that the magnetic moment of the silver atom, as measured in the Stern-Gerlach experiment, is actually the magnetic moment of the valence electron of silver. A glance at Table 5 of Appendix A shows that silver, in the normal state, has a single electron,

Fig. 29-10 Type of pattern made by a beam of silver atoms (a) without magnetic field on; (b) with magnetic field on.



an *s* electron outside of a closed, stable configuration of the other 46 electrons. The sum of the magnetic moments of these 46 electrons is zero. The magnetic moment of an electron is ascribed to its intrinsic spin. One interesting result of the above experiment was that the silver atom was deflected in only one of two possible directions, as indicated by the splitting of the line into two lines. The interpretation of this is that a spinning electron, when in a magnetic field, cannot have any arbitrary direction in space. The spin of an electron may be considered as a vector quantity, since there is an angular momentum associated with it directed along the axis of spin. The magnetic moment is also a vector quantity directed along this axis. When an electron enters a magnetic field, its axis can take only one of two possible orientations, either parallel to the direction of the

magnetic field or antiparallel, that is, oppositely directed, depending upon the direction of its spin. This limitation of the possible orientations of the axis of spin of an electron is called *space quantization.*

Measurements of the magnetic moment, by the methods of Stern and Gerlach and by other methods developed subsequently, have been made of atoms, molecules, and subatomic particles such as electrons, protons, neutrons, and many nuclei. These measurements are an important part of our knowledge of the properties of matter.

29-6 Terrestrial Magnetism

We have already shown that a bar magnet, when suspended so that it can swing freely about a vertical axis, will ultimately come to rest in an approximately north-south position. The use of a compass is based upon this

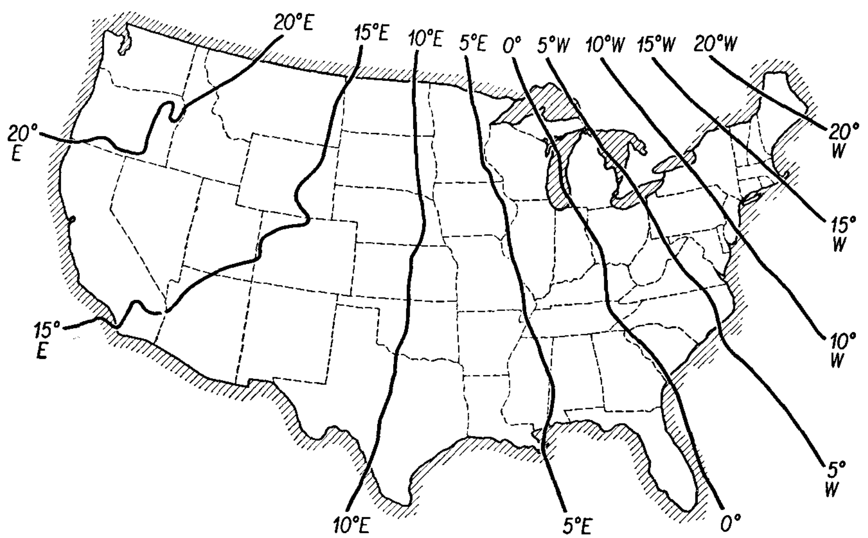


Fig. 29-11 Chart showing isogonic lines for the United States, 1955. From the 1955 edition of chart 3077, of the U.S. Coast and Geodetic Survey, showing distribution of magnetic declination throughout the United States. (Courtesy of Coast and Geodetic Survey.)

observation. It is important, however, to know how much the magnetic north-south direction differs from the geographic or true north-south direction. Various surveys have been made to determine this difference for various points of the earth's surface. This difference is termed the *declination* of the magnetic from the geographic north-south direction. It differs from point to point on the earth's surface and also is found to vary in a

complex manner with time. Charts are plotted for any region of the earth's surface with lines showing equal declinations, or *isogonic* lines, marked on them. One such chart is shown in Figure 29-11. In New York the declination is about 11°W, in Detroit it happens to be 3°W, while in San Francisco it is about 18°E.

The declination gives the direction of the horizontal component only of the earth's magnetic field at any one point. The fact that there is a vertical component to the earth's magnetic field can be shown by the following simple experiment. A long unmagnetized steel needle is mounted so that it can pivot about a horizontal axis through its center of gravity. It is balanced in the horizontal position and pointed in the direction in which a compass needle would point. The needle is now magnetized by means of a strong magnet and then allowed to swing freely. It will be found that the magnetized needle no longer balances in the horizontal position but dips with its north pole downward. The angle of dip varies from place to place over the earth. At New York the angle of dip is 72° with the horizontal, near the equator it is almost 0°, and at the magnetic pole in the Northern Hemisphere it is 90°. In this manner the earth's magnetic poles are located.

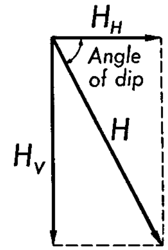


Fig. 29-12 Angle of dip.

When the dip needle is properly used as described above, its direction is that of the earth's magnetic field at the given place. The horizontal component H_H of the intensity of the earth's magnetic field can easily be determined by means of an oscillating horizontal magnet, as described above. By means of a simple vector diagram, illustrated in Figure 29-12, the total magnetic field intensity H and its vertical component H_V can be computed.

The earth behaves as though it contained a very large magnet with its south pole in the Northern Hemisphere located at and below Boothia Peninsula in northern Canada, and the other pole in the Southern Hemisphere near Ross Sea in Antarctica. The origin of the earth's magnetic field is at present unknown. The hypothesis that seems most reasonable is that the earth's magnetic field is produced by electric currents within the earth, probably within its core. If the above hypothesis is accepted, one must explain the source of these electric currents.

29-7 A Molecular Theory of Magnets

An insight into the origin of the magnetic properties of magnetic substances can be obtained from the following simple experiment. First magnetize a long steel needle or steel strip; then break the magnet in half. By dipping

these parts in iron filings, it will be found, as sketched in Figure 29-13, that each half is itself a magnet. By testing the polarity of each part, it will be found that opposite poles appear on the faces where the break was made. If one of these parts is again broken in half, we will again have two magnets. This process can be repeated until the magnets become too small to handle. Each part broken from the original magnet will always be found to be magnetized. This suggests that a magnet is made up of very tiny magnets, perhaps of molecular size.

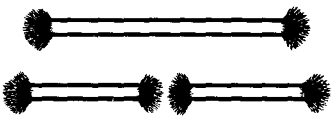


Fig. 29-13 A bar magnet broken in half forms two bar magnets.

On this theory of magnetism, magnetic substances are assumed to contain little magnets of approximately molecular size. When the bar of steel is unmagnetized, these molecular magnets are oriented at random, as sketched in Figure 29-14(a), so that the bar shows no polarity and has no net magnetic moment. But when the bar of steel is brought into a magnetic field, its molecular magnets rotate until they line up with the field, as shown in Figure 29-14(b); the bar thus acquires magnetization, magnetic poles, and a magnetic moment. When the steel bar is removed from the magnetic field, most of these little magnets retain their orientation so that the steel bar remains permanently magnetized, although its magnetic moment may not be as big as it was when in the magnetic field. Other substances, such as soft iron, lose most of their magnetism when taken out of the magnetic field.

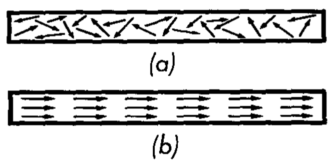


Fig. 29-14 (a) Molecular magnets in an unmagnetized steel bar have random orientations. (b) When the steel bar is magnetized, the molecular magnets line up parallel to the external field, producing poles at the ends of the bar.

Other evidence in support of this molecular theory of magnetism may be cited. If a magnetized steel bar is dropped, hammered, or otherwise jarred, it loses some of its magnetism. If it is heated to a temperature beyond about 800°C, it loses its magnetic properties, probably because of the fact that the increased vibratory motion of the molecules destroys the orientation of the molecular magnets. This molecular theory is not, of course, the complete explanation. One would like to know the origin of these elementary magnets and why some substances are magnetic while others apparently are not magnetic. We shall consider this topic again after we have shown the relationship between electricity and magnetism.

29-8 A Theory of Magnetism

From Coulomb's law it is possible to develop a theory of magnetism based upon the magnetic dipole as the basic element of magnetism, and upon the auxiliary concept of the magnetic pole. Such a theory of magnetism is quite analogous to the theory of electrostatics developed in earlier chapters. In this theory the *magnetic field intensity* **H** is analogous to the electric field intensity **E**. The contribution to the magnetic field intensity at a point due to a magnetic pole of strength p can be written in an equation analogous to Equation (23-2) as

$$\mathbf{H} = k_1 \frac{p}{r^2} \mathbf{1}_r, \tag{29-11}$$

where, in the cgs system of units,

cgs units $$k_1 = 1$$

in vacuum, and **H** is expressed in oersteds when p is expressed in cgs unit poles and r is expressed in centimeters. In the mks system of units,

mks units $$k_1 = \frac{1}{4\pi\mu_0} \tag{29-12}$$

in vacuum, and **H** is expressed in newtons per weber when p is expressed in webers, r is expressed in meters, and

$$\mu_0 = 4\pi \times 10^{-7} \frac{\text{weber}^2}{\text{nt m}^2}.$$

In many problems associated with the magnetic fields due to permanent magnets, it is helpful to define a magnetic potential similar to the electric potential defined in electrostatics. One essential difference between electrostatics and magnetostatics lies in the fact that there are no *conductors* of magnetism. Another difference which we have already emphasized is that the isolated magnetic pole is a useful mental construct rather than an experimentally observed entity.

To carry the analogy further, we utilize the concept of *magnetic lines of force,* developed as in Section 23-5 for the electric field. There are 4π lines of force emanating from a cgs unit north pole, and $1/\mu_0$ lines of force emanating from an mks unit north pole of pole strength 1 weber, as in Equations (23-5). Since the source of the magnetic field is the magnetic dipole, every volume element must contain as much positive magnetism as negative magnetism, so that the net number of magnetic lines of force emerging from any Gaussian surface is zero.

In dealing with the magnetic field in the material medium, we define the *magnetization* **M** as the *magnetic moment per unit volume*, just as, in a dielectric, we spoke of the polarization **P** as the electric moment per unit volume. The magnetic medium differs from the dielectric medium in that a permanent state of magnetization is commonly found in magnetic materials, while the polarization of most dielectric materials is only found in the presence of an electric field.

It is possible to produce a state of permanent electric polarization in some waxes by cooling them slowly in an electric field. These permanently polarized waxes are called *electrets*, and generate an electric field similar to the magnetic field generated by magnets. When electrets are left alone for a time, electric charges, in the form of atmospheric ions, collect on their surfaces in such a way as to neutralize the electric field generated by the electret. The electret may be revived by scraping away its surface, removing the collected ions. The fact that permanent magnets are not diminished in strength by exposure to the atmosphere may be cited as evidence that free magnetic poles do not exist in the atmosphere.

In addition to the permanent magnetization, we describe the induced magnetization in a magnetic material by the magnetic susceptibility χ_m (chi sub m) so that the induced magnetization could be written as

$$\mathbf{M} = \chi_m \mathbf{H}, \tag{29-13}$$

analogous to Equation (25-12).

Just as it was convenient to describe the electric field in the material medium by means of the electric displacement **D**, so it is convenient to describe the magnetic field in terms of the magnetic induction **B**, defined by equations analogous to Equation (25-24) as

$$\mathbf{B} = \mu_0 \mathbf{H} + \mathbf{M}. \tag{29-14a}$$

In the mks system of units, the magnetization **M**, the magnetic moment per unit volume, is expressed in webers per square meter, and the product $\mu_0\mathbf{H}$ is similarly expressed in webers per square meter. Thus the units of magnetic induction **B** are expressed in webers per square meter. We may describe the relationship between **B** and **H** in a linear medium through the equation

$$\mathbf{B} = \mu \mathbf{H}, \tag{29-15a}$$

where

$$\mu = \kappa_m \mu_0. \tag{29-16}$$

In these equations μ is called the *permeability* of the medium, and κ_m is called the relative permeability of the medium. The value of κ_m for vacuum is unity.

In the cgs system of units Equation (29-14a) may be re-expressed as

cgs units
$$\mathbf{B} = \mathbf{H} + 4\pi\mathbf{M}. \tag{29-14b}$$

The units of magnetic induction **B** are expressed in *gausses*. The terminology for this system of units is not so well worked out as in the case of mks units, so that each quantity in Equation (29-14b) is expressed in different units. Thus **H** is expressed in oersteds and **M** is expressed in cgs poles per square centimeter. The linear relationship between **B** and **H**, for a linear medium, is then expressed as

cgs units $$\mathbf{B} = \kappa_m \mathbf{H}. \tag{29-15b}$$

This neat analogy between the electric field and the magnetic field is very useful, but it neglects the most important of all magnetic effects, the generation of a magnetic field by an electric current. It was discovered by Hans Christian Oersted in 1820 that an electric current in a wire caused a compass needle to be deflected. Subsequent investigations have shown that the magnetic field exerts a force on a moving charge, and that a changing magnetic field induces a current in a wire. The theory outlined above must be supplemented to include these effects.

In contrast to the case of the electric field in which the electric displacement **D** plays a minor role, important in simplifying calculations but not at the foundation of electrostatics, the magnetic induction **B** is of great significance, for, as we shall see, **B** describes the effects of the field on a moving charge, while **H** describes the effects of the field upon a pole. From a fundamental physical viewpoint the pole is a property of bulk matter and is not a fundamentally significant magnetic parameter. Since isolated poles do not exist in nature, we are forced to relegate the magnetic field intensity **H** to the role of an auxiliary quantity and to think of the magnetic induction **B** as the fundamental magnetic-field vector. The sources of the magnetic field are today understood to be electric currents and the magnetic dipole moments of the fundamental physical particles, while the probe by means of which we are enabled to examine the magnetic field is either a moving charged particle or an elementary magnetic dipole.

Problems

29-1. A bar magnet having a pole strength of 400 cgs poles and a length of 5 cm is placed so that its south pole is at the origin and its north pole is located at the point whose coordinates are (5 cm, 0). Find the magnitude and direction of the force exerted by this magnet on the north pole of a long, slender, needlelike magnet whose pole strength is 20 cgs poles, when that pole is placed at a point whose coordinates are (a) (0, 10 cm), (b) (−10 cm, 0), and (c) (5 cm, 3 cm).

29-2. A bar magnet having a pole strength of 10^{-5} weber is placed along the y axis so that its south pole is at the origin and its north pole is located at $y = 0.1$ m. Find the magnitude and direction of the force exerted by this magnet on the north pole of a needlelike magnet of pole strength 10^{-6} weber when that north pole is located at a point whose coordinates are (a) (0, −0.2 m), (b) (0.24 m, 0), and (c) (0.03 m, 0.05 m).

29-3. Two identical bar magnets are placed side by side with their axes parallel and separated by a distance of 10 cm. Find the magnitude and direction of the force between them when like poles face each other. The pole strength of each magnet is 300 cgs poles, and the length of each magnet is 20 cm.

29-4. Repeat the calculation of Problem 29-3 in the case that unlike poles face each other and the pole strength of the magnets is 10^{-5} weber. Perform your calculations in mks units.

29-5. What is the intensity of the magnetic field at a point 20 cm from the north pole of a bar magnet, measured along the axis of the magnet, if the magnet is 30 cm long and has a pole strength of 400 cgs poles?

29-6. What is the magnetic field intensity generated by the bar magnet at each of the points indicated in Problem 29-1?

29-7. What is the magnetic field intensity generated by the bar magnet at each of the points indicated in Problem 29-2?

29-8. A bar magnet 4 cm long, having a pole strength of 12 cgs poles, is placed in a uniform magnetic field with the axis of the magnet perpendicular to the direction of the magnetic field. The intensity of the magnetic field is 45 oersteds. Determine the torque acting on this magnet.

29-9. A small compass needle oscillates with a period of 2.5 sec when placed at a point where the horizontal component of the earth's magnetic field is 0.2 oersted. Determine the magnetic moment of the compass needle if its moment of inertia about the axis of rotation is 24 gm cm^2.

29-10. A bar magnet 6 cm long is suspended by a wire passing through its center. The magnet is placed in a uniform field with the axis of the magnet perpendicular to the direction of the field. The torque, as measured by the twist of the wire, is 1,500 dynes cm. Determine (a) the magnetic moment of the magnet and (b) the pole strength. The intensity of the field is 50 oersteds.

29-11. A small bar magnet has a period of oscillation of 2 sec when pivoted in a uniform magnetic field of 36 nt/weber. When it is used to measure the intensity of a second magnetic field, the period is found to be 0.40 sec. Determine the intensity of the second magnetic field.

29-12. A uniformly magnetized rod has a pole strength of 10 webers. The magnet has a cross-sectional area of 1 cm^2 and a length of 10 cm. (a) What is the magnetic moment of the magnet? (b) What is the magnetization of the magnet?

29-13. By analogy with the electric field in a material medium (Sections 25-6 to 25-8), write a formula for the magnetic field intensity H due to a magnetic pole of pole strength p when that pole is immersed in a medium of relative permeability κ_m (a) in cgs units and (b) in mks units.

29-14. From the result of Problem 29-13, state Coulomb's law of force between two poles immersed in a medium of relative permeability κ_m (a) in mks units and (b) in cgs units.

30

Magnetic Fields of Currents

30-1 Magnetic Field around an Electric Current

The first evidence for the existence of a magnetic field around an electric current was observed in 1820 by Hans Christian Oersted (1777–1851). He found that a wire carrying current caused a freely pivoted compass needle

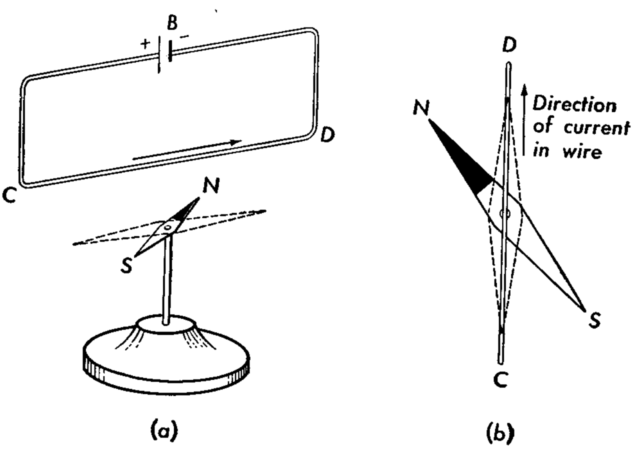


Fig. 30-1 Oersted's experiment. Compass needle is deflected toward the west when the wire CD carrying current is placed above it and the direction of the current is toward the north, from C to D.

in its vicinity to be deflected. If the current in a long straight wire is directed from C to D, as shown in Figure 30-1, a compass needle below it, whose initial orientation is shown in dotted lines, will have its north pole deflected to the left and its south pole deflected to the right. If the current in the wire is reversed and directed from D to C, then the north pole will be deflected to the right, as seen from above. In terms of the forces acting on the poles, these forces are clearly perpendicular to the direction of the current and to the line from the nearest portion of the wire to the pole itself.

The magnetic field in the neighborhood of a wire carrying current can be investigated either by exploring the region with a small compass or by using iron filings. When a wire carrying current is passed perpendicularly through a plane board and iron filings are sprinkled on the board, the

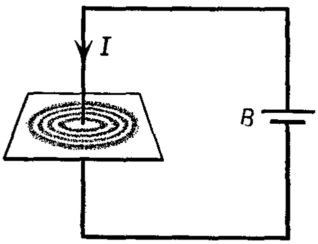


Fig. 30-2 Pattern formed by iron filings showing the circular magnetic field around a wire carrying current.

filings form a circular pattern, as shown in Figure 30-2. Thus the magnetic field generated by the wire carrying current is circular in a plane at right angles to the current. The circles are concentric, with their common center at the position of the wire. The direction of the magnetic field can be

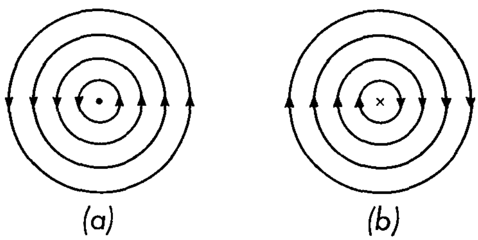


Fig. 30-3 Direction of the magnetic field around a wire (a) when the current is out of the paper; (b) when the current is into the paper. The dot represents a head-on view of an arrow while the cross represents a rear view.

determined with the aid of a small compass. If we look along the wire so that the current is coming toward us, the magnetic field is counterclockwise. If we draw a dot to represent a head-on view of an arrow and a cross to represent a rear view of an arrow, we may show the direction of the magnetic field, associated with current in a wire perpendicular to the plane of the paper when the current is coming toward the reader in Figure 30-3(a), and when the current is away from the reader in Figure 30-3(b). A small compass placed anywhere in the field will orient itself tangent to one of these circles with its north pole in the direction of the arrow.

When the wire carrying current is bent in the form of a circular loop, as shown in Figure 30-4, the direction of the magnetic field around each small portion of the wire may be determined from the above observations.

If the current in the loop is counterclockwise when viewed from the positive x axis, the direction of the magnetic field at the center of the loop is perpendicular to the plane of the loop and in the positive x direction.

A simple way of determining the direction of the magnetic field relative to the direction of the current is given by the *right-hand rule.* If the current

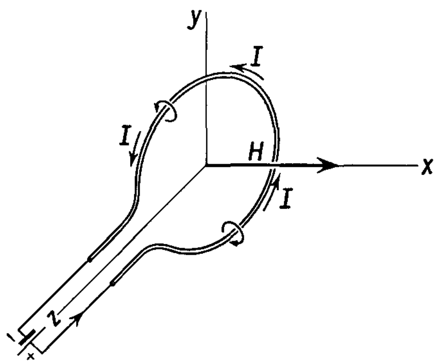


Fig. 30-4 Magnetic field produced by a current in a circular loop of wire. The magnetic field at the center is at right angles to the plane of the loop.

is in a straight portion of the wire, then, if we imagine the thumb of the right hand to be placed along the wire and pointing in the direction of the current, the curled fingers of the right hand will point in the direction of the magnetic field. If the current flows in a circular path, then, if we imagine the fingers of the right hand curving in the direction of the current, the thumb will point in the direction of the magnetic field inside the coil, as in Figure 30-4.

When, as in Figure 30-5, the wire is wound in the form of a helix which is very long in comparison with its diameter, the magnetic field outside the coil is very similar to that of a bar magnet, while the field inside the coil is uniform and parallel to the axis of the coil. At large distances the field produced by a circular loop of wire carrying current is indistinguishable from the field generated by a small magnet. According to a theory first presented by Ampère, it is possible to imagine that all magnetic effects are due to circulating currents, and to attribute the magnetic fields generated by permanent magnets to circulating currents within molecules. These currents are associated with orbital motions of electrons about some attractive center within the molecule.

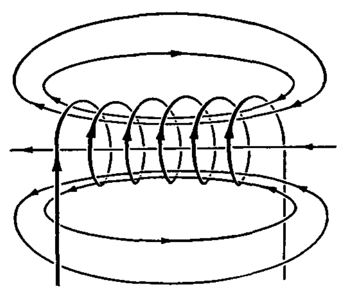


Fig. 30-5 Magnetic field due to a current in a long cylindrical coil or helix.

30-2 Magnetic Field Generated by a Wire Carrying Current

As soon as Oersted's discovery was announced, scientists in Europe began repeating and extending these experiments. Interest in Oersted's results was aroused because this was the first time a force had been observed which was not directed along the line joining the two bodies but was perpendicular to this line. Since Newton had developed the concept of universal gravitation, both electrostatic and magnetostatic forces had been studied. All these forces obeyed an inverse square law and were directed along the line joining the bodies responsible for the forces. Biot and Savart (1820) determined the magnetic field produced by the current in a long straight wire. The experiments of Biot and Savart showed that the magnetic field intensity H at a distance a from a long straight wire carrying current I was directly proportional to the magnitude of the current and inversely proportional to the distance from the wire; that is,

$$H = k_2 \frac{2I}{a}, \tag{30-1}$$

where k_2 is a constant of proportionality. The direction of the magnetic field intensity was always normal to the wire and to the perpendicular from

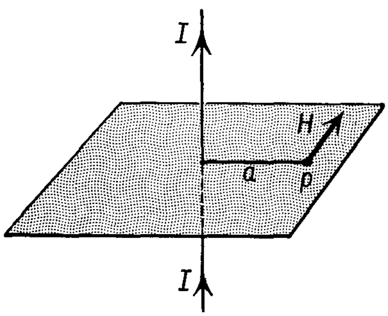


Fig. 30-6 The field at a point P due to a long straight wire is perpendicular to the plane formed by the wire and the normal a dropped from P to the wire. In this figure H is directed into the plane of the paper, as given by the right-hand rule.

the point where the field was being evaluated to the wire. The direction of the magnetic field was given by the right-hand rule, as shown in Figure 30-6.

The value of the constant k_2 depends upon the system of units used to measure H, I, and a. In the mks system of units, the current I is expressed in amperes, the distance a is expressed in meters, and the magnetic field intensity H is expressed in mks units, stated in Section 29-3 as newtons per weber. In this system of units the constant k_2 is $\frac{1}{4\pi}$, and Equation (30-1) becomes

$$H = \frac{I}{2\pi a}. \tag{30-1a}$$

In Section 29-3 we mentioned that the units of H in the mks system were commonly expressed as amperes per meter. The justification for expressing H in these terms is provided by Equation (30-1a). The units of H may be stated equivalently as newtons per weber or as amperes per meter, and the choice of units will depend upon which form is more convenient for the problem in hand.

In one form of cgs units called the *Gaussian* system of units, toward which we have been building our cgs unit system, the units of electrical quantities such as charge, current, electric field intensity, electric displacement, potential, and resistance are based upon the statcoulomb as the unit of charge, while units of magnetic quantities, the oersted and the gauss, are based upon the cgs unit pole. In this system the constant k_2 is $1/c$, where c is the velocity of light in centimeters/per second, given approximately as $c = 3 \times 10^{10}$ cm/sec. To simplify the presentation of the material of this chapter, all equations will be presented in the mks system of units.

Traditionally, Equation (30-1) has been used to define a new unit of current called the *electromagnetic unit of current*, by setting k_2 equal to 1. This unit is called the emu of current, or the *abampere*. Electromagnetic units of current will rarely be used in this book.

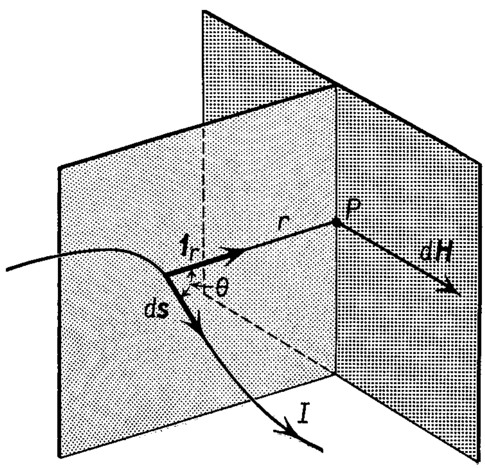


Fig. 30-7 The magnetic field intensity $d\mathbf{H}$ at any point P a distance r from an element $d\mathbf{s}$ of wire carrying current I is perpendicular to the plane of $d\mathbf{s}$ and $\mathbf{1}_r$, where $\mathbf{1}_r$ is the unit vector directed from $d\mathbf{s}$ to P. The element $d\mathbf{H}$ is shown coming out of the plane of the paper when $d\mathbf{s}$ and $\mathbf{1}_r$ lie in the plane of the paper.

30-3 Magnetic Field Generated by Current Element

The results of the experiments of Biot and Savart on the magnetic field around a wire carrying current form the basis of a fundamental physical law for the determination of magnetic fields produced by currents. If we consider any small element of wire of length ds in which there is a current I (see Figure 30-7), its contribution $d\mathbf{H}$ to the magnetic field intensity at a

point P located a distance r from it, is given by

$$\boxed{d\mathbf{H} = \frac{I\, d\mathbf{s} \times \mathbf{1}_r}{4\pi r^2}.} \tag{30-2}$$

The unit vector $\mathbf{1}_r$ is directed from the current element to the field point P. The vector $d\mathbf{s}$ is tangent to the direction of the current. The direction of $d\mathbf{H}$ is perpendicular to both $d\mathbf{s}$ and $\mathbf{1}_r$. Equation (30-2) is sometimes called the law of Biot and Savart, and also *Ampère's law.*

The total magnetic field intensity $\mathbf{H}$ at a point P due to a given current distribution is found by summing vectorially (that is, integrating) the contributions $d\mathbf{H}$ from all the elements $d\mathbf{s}$ of the wire. For most geometrical arrangements of the conductor, this summation is beyond the reach of the calculus and must be carried out by numerical methods of integration or by use of an electronic computer. We shall consider only two cases, the flat circular coil and the long straight wire.

30-4 Magnetic Field of a Flat Circular Coil

Let us use Ampère's law to calculate the magnetic field intensity at the center of a single circular loop of wire carrying a current I. As shown in

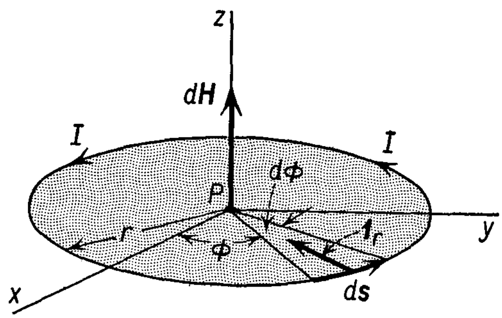


Fig. 30-8

Figure 30-8, each element of the loop of wire makes a right angle with the line drawn from that element to the field point P at the center of the circle. The direction of the magnetic field $d\mathbf{H}$ due to each element of the loop $d\mathbf{s}$ is in the positive z direction when the current is in the counterclockwise direction, as shown in the figure. Since the contribution to the magnetic field from every element of current is in the same direction, the problem of obtaining the vector sum of the contributions to the total magnetic field intensity is reduced to a scalar sum, and we may set up the problem in the form of an integral.

From Equation (30-2) we have

$$d\mathbf{H} = \frac{I\,d\mathbf{s} \times \mathbf{1}_r}{4\pi r^2}.$$

The factor $d\mathbf{s} \times \mathbf{1}_r$ is equal in magnitude to the product of the magnitude ds by the magnitude of $\mathbf{1}_r$ (which is unity), by the sine of the angle between them. Since the angle between the two vectors is 90°, the magnitude of this vector product is simply the magnitude ds. The direction of this vector product is in the $+z$ direction, so that we may represent it as

$$d\mathbf{s} \times \mathbf{1}_r = ds\,\mathbf{1}_z,$$

where $\mathbf{1}_z$ is a unit vector in the positive z direction. Thus we have

$$d\mathbf{H} = \frac{I\,ds}{4\pi r^2}\,\mathbf{1}_z,$$

where the magnitude of $d\mathbf{H}$ is the factor multiplying $\mathbf{1}_z$.

In order to carry out the integration, we must represent the variables of the problem in terms of some simple parameter. Let us measure the angular position of an element of current by the angle ϕ, from the x axis, positive in the counterclockwise direction. The angle subtended by the element of length ds is $d\phi$ such that

$$ds = r\,d\phi.$$

Thus we have

$$dH = \frac{Ir\,d\phi}{4\pi r^2},$$

and, integrating around the circumference of the circle by letting ϕ go from 0 to 2π, we find

$$H = \int_0^{2\pi} \frac{I\,d\phi}{4\pi r} = \frac{I}{2r}.$$

The complete representation of the magnetic field intensity at the center of a single circular loop of wire of radius r carrying a current I in the counterclockwise direction is given by

$$\mathbf{H} = \frac{I}{2r}\,\mathbf{1}_z. \tag{30-3}$$

In Equation (30-3) the appropriate choice of units to represent H is amperes per meter when I is in amperes and r is expressed in meters.

In the event that we have a coil of N turns of wire wound as a flat circular coil rather than a single circular loop, each turn of the coil contributes a magnetic field at the center of the coil, as given by Equation (30-3). Since these field contributions are all in the same direction, the total magnetic field is N times that given by Equation (30-3). A circular coil of wire of N turns carrying current I in the counterclockwise direction, when viewed from the positive z axis, as in Figure 30-8, has at its center a magnetic field $\mathbf{H}$ given by

$$\mathbf{H} = \frac{NI}{2r}\mathbf{1}_z. \tag{30-4}$$

One of the early forms of current-measuring instruments was the *tangent galvanometer*, which consisted of a single circular loop of wire with a short compass needle placed at its center, as shown in Figure 30-9. When a current I was sent through the loop, it set up a magnetic field of intensity

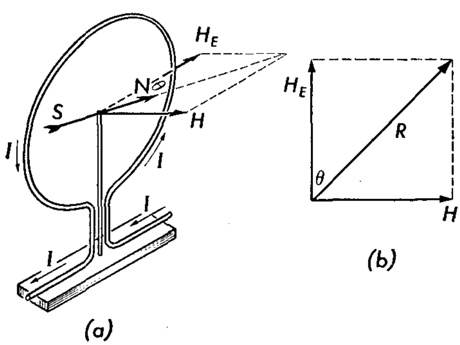


Fig. 30-9 (a) Tangent galvanometer. The plane of the circular coil is parallel to the earth's magnetic field. (b) The two magnetic fields which act on the N-pole of the small compass needle. H_E is the horizontal component of the earth's magnetic field; H is the intensity of the magnetic field produced by the current. R is the resultant of these two field intensities. $H = H_E \tan\theta$.

H at its center, producing a force Hp on each pole of the magnet. In the figure the force on the north pole is shown directed to the right, while that on the south pole is directed to the left. These two forces produced a torque causing the compass needle to rotate out of the plane of the coil. The horizontal component of the earth's magnetic field produced an opposite torque on the compass needle. The compass needle came to rest at some angle θ with the plane of the loop, at which angle the two opposing torques were equal in magnitude. A simple calculation shows that the current in the loop is proportional to the tangent of the angle θ, hence the name tangent galvanometer. This was the instrument refined by Lord Kelvin as the detector for use in the first successful operation of the Atlantic cable.

Illustrative Example. A circular coil of 30 closely wound turns of 15 cm average radius carries a current of 2 amp. Determine the magnitude and direction of the magnetic field at the center of this coil.

Assuming that the turns are so closely wound that the current in each one contributes the same amount to the intensity of the magnetic field at the center of the coil, the magnetic field intensity is of magnitude

$$H = \frac{NI}{2r}.$$

Substituting numerical values, we have

$$H = \frac{30 \text{ turns} \times 2 \text{ amp}}{2 \times 0.15 \text{ m}} = 200 \frac{\text{amp}}{\text{m}}.$$

Note that the *turn* is a dimensionless quantity. The direction of the magnetic field is perpendicular to the plane of the coil, as given by the right-hand rule.

30-5 Magnetic Field of a Long Straight Wire

The experimental result of Biot and Savart given in Equation (30-1) may be calculated from Equation (30-2) by integrating the contributions to the field from each element of the wire. The field point P is located a distance a from the wire, as shown in Figure 30-10. Let us measure the distance s along the wire from the foot of the perpendicular from P to the wire.

In terms of the angle θ between the element ds and the unit vector $\mathbf{1}_r$ from the element ds to the field point P, the distance r may be represented as

$$r = \frac{a}{\sin\theta} = a \csc\theta.$$

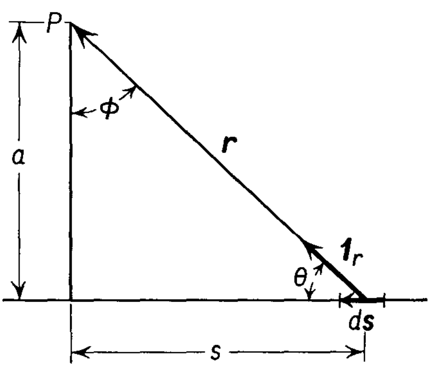


Fig. 30-10

To express the length of the element ds in terms of the distance a and the angle θ, we first note that the length s is given by

$$s = a \cot\theta,$$

so that, taking differentials to obtain the magnitude of ds,

$$ds = -a \csc^2\theta \, d\theta.$$

We note that contribution to the magnetic field from each element of wire is in the same direction at P; that is, it is directed out of the plane of the

paper, toward the reader. If we call this direction the direction of z, then

$$d\mathbf{H} = +\frac{I\,ds\sin\theta}{4\pi r^2}\mathbf{1}_z.$$

Substituting appropriate expressions into the above equation, we find

$$dH = -\frac{Ia\csc^2\theta\,d\theta\sin\theta}{4\pi a^2\csc^2\theta}$$

$$= -\frac{I\sin\theta\,d\theta}{4\pi a}.$$

Since all contributions to the magnetic field at P are in the same direction, we may find the magnetic field intensity at P by integrating between the limits of $\theta = \pi$ to $\theta = 0$, corresponding to limits on s of $s = -\infty$ to $s = +\infty$. Thus we have

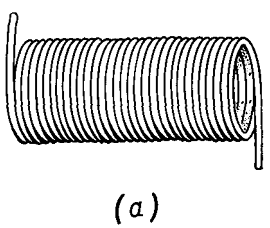
$$H = \int_{\pi}^{0} -\frac{I}{4\pi a}\sin\theta\,d\theta,$$

$$H = \frac{I}{2\pi a},$$

as previously stated in Equation (30-1a).

30-6 Field of a Solenoid and a Toroid

A solenoid is made by winding wire on a cylindrical form; the length s of the cylinder is generally much larger than its radius r. The adjacent turns of the wire are generally close together, as shown in Figure 30-11. The field inside the solenoid may be calculated from Ampère's law. The field is uniform over the entire region of the solenoid except near the ends, as



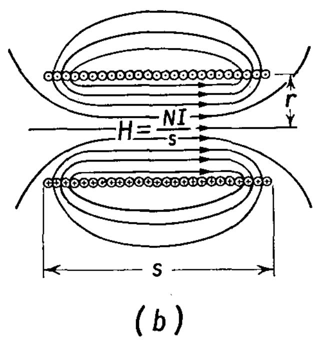


Fig. 30-11 (a) Windings of a solenoid. (b) Magnetic field intensity inside a solenoid. Current is coming out of the wires on top and going into the wires on bottom.

shown in the figure. The field intensity within a solenoid will be shown in Section 33-6 to be

$$H = \frac{NI}{s}, \tag{30-5}$$

where N is the number of turns of the solenoid, s is its length, and I is the current in it.

A toroid is made by winding wire on a ring or doughnut-shaped form called a *torus*. The field of the toroid is entirely confined to the space within the toroid. A toroid may be generated by bending a long straight solenoid

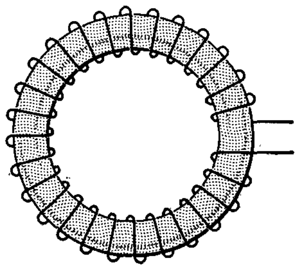


Fig. 30-12 Toroid.

in the form of a ring. The magnetic field intensity within the toroid is given by Equation (30-5), where s now represents the mean circumference of the ring, shown in Figure 30-12.

The solenoid and the toroid are frequently used as standard means of achieving known, uniform magnetic fields.

30-7 Equivalence of a Moving Charge and a Current

Our original definition of current was based on the flow of charges through a surface in a given time interval; it remains to be shown that a set of moving charges will produce the same magnetic effect as a current in a wire. This was first shown experimentally by H. A. Rowland in 1876. He used an ebonite disk having metallic sectors distributed near the rim of the disk. The metallic sectors were charged electrically, and the disk was set into rapid rotation. A magnetic needle suspended near the disk was deflected by the magnetic field set up by the moving charges. The direction of the deflection was the same as that which would have been produced by the current in a circular loop of wire coinciding with the rim of the disk. When the direction of rotation was reversed, the deflection of the magnetic needle was also reversed. More recently (1929) R. C. Tolman set a charged cylinder oscillating about its axis and observed that this produced an alternating magnetic field, the same as that produced in the neighborhood

of an alternating current. These experiments show that, as far as the magnetic effect is concerned, a moving charge and a current in a wire are equivalent.

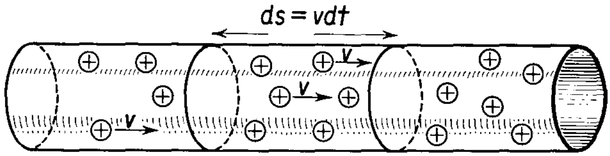


Fig. 30-13

Figure 30-13 shows a tube in which a current I consists of the motion of positive charges with uniform velocity v. From the definition of current

$$I = \frac{dq}{dt},$$

where dq is the quantity of charge which passes through a complete cross section of the tube in a time dt. The charge will traverse a length ds of the tube, where

$$ds = v\,dt.$$

Eliminating dt from these two equations yields

$$I\,ds = v\,dq.$$

In other words, a charge dq moving with velocity v may be considered to be equivalent to a current element of length ds carrying a current I.

We may find the magnetic field H generated by a moving charge q by substituting the above results into Equation (30-2). We find

$$\boxed{\mathbf{H} = \frac{q\mathbf{v}\times\mathbf{1}_r}{4\pi r^2}.} \tag{30-6}$$

30-8 Coulomb's Law and Ampère's Law

Let us consider the magnetic field generated by a charged particle mounted on an airplane. If we calculate the magnetic field generated by the moving charge from Equation (30-6), we note that, to an observer on the ground, the charge is moving with the velocity v of the airplane, but, to an observer on the airplane, the charge is moving with zero velocity. Thus an observer on the surface of the earth is able to detect a magnetic field due to the motion of the charged particle on the airplane, but an observer on the airplane cannot detect a magnetic field due to the charge

on the airplane, for that charge has no motion with respect to him. By definition, the magnetic field intensity H in a particular reference frame is the force on a unit north pole at rest in that frame. If our formulation is to be consistent, we must expect that the force exerted by a charge upon a magnetic pole is the same when these are in relative motion, regardless

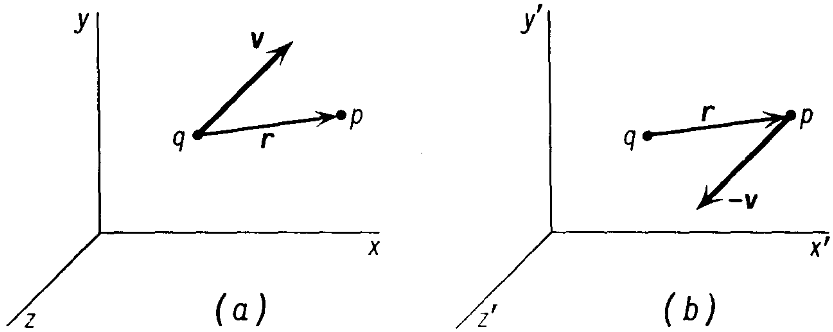


Fig. 30-14 (a) System of axes in which the magnetic pole is at rest. (b) System of axes in which the charge q is at rest.

of whether the pole is considered to be at rest or the charge is considered to be at rest.

In the coordinate frame in which the pole of strength p is at rest, as shown in Figure 30-14(a), the charge q is moving with velocity $\mathbf{v}$, and the magnetic field $\mathbf{H}$ at the position of the pole, due to the moving charge, is given by Equation (30-6) as

$$\mathbf{H} = q\frac{\mathbf{v}\times\mathbf{1}_r}{4\pi r^2}\cdot$$

The force on the magnetic pole at this point is therefore

$$\mathbf{F} = \mathbf{H}p = qp\frac{\mathbf{v}\times\mathbf{1}_r}{4\pi r^2}\cdot$$

If we consider the frame of reference to be fixed with respect to the charge q, as shown in Figure 30-14(b), we must consider the charge to be at rest and the pole to be moving with a velocity $\mathbf{v}' = -\mathbf{v}$. Substituting this into the above equation, and rearranging the position of q in the equation we find

$$\mathbf{F} = p(-\mathbf{v}')\times\frac{q\,\mathbf{1}_r}{4\pi r^2}$$

for the force exerted on the moving pole by the fixed charge. From the illustrative example of Section 25-8, we note that the terms grouped to the

right of the cross in the above equation represent precisely the electric displacement **D** generated by the charge q at the position of the pole p, for

$$\mathbf{D} = \frac{q\,\mathbf{1}_r}{4\pi r^2}.$$

Thus the force on a pole p, moving with velocity **v** (dropping the prime) in a coordinate frame in which the electric displacement is **D**, is given by

$$\mathbf{F} = -p(\mathbf{v}\times\mathbf{D}).$$

Thus the force experienced by a moving magnetic pole consists of two parts. One part is due to the magnetic field **H**; this force on the pole is independent of its state of rest or motion. The second part depends upon the motion of the poles; this force is due to the velocity of the pole and to the electric displacement **D**. In the form of an equation, the force on a magnetic pole, moving with velocity **v** in a coordinate frame in which the magnetic field intensity is **H** and the electric displacement is **D**, is given by

$$\mathbf{F} = p(\mathbf{H} - \mathbf{v}\times\mathbf{D}). \tag{30-7}$$

Another interpretation of Equation (30-7) is that the force exerted on a magnetic pole by a charged particle may be calculated in one of two ways: (a) for the case in which the pole is considered to be at rest and the charge in motion, we can use Ampère's law; (b) for the case in which the charge is considered to be at rest and the pole is in motion, we can use Coulomb's law, utilizing the concept of electric displacement. In Rowland's experiment a moving charge was seen to deflect a compass needle. From Equation (30-7) we see that we must expect a compass needle to be deflected when the needle is moving in a region of space in which there is an electric

TABLE 30-1 PRINCIPAL EQUATIONS IN MKS AND GAUSSIAN UNITS

Equation	MKS	Gaussian	
(30-2)	$d\mathbf{H} = \dfrac{I\,d\mathbf{s}\times\mathbf{1}_r}{4\pi r^2}$	$d\mathbf{H} = \dfrac{I\,d\mathbf{s}\times\mathbf{1}_r}{r^2 c}$	Current element
(30-1a)	$H = \dfrac{I}{2\pi a}$	$H = \dfrac{2I}{ac}$	Straight wire
(30-5)	$H = \dfrac{NI}{s}$	$H = \dfrac{4\pi NI}{sc}$	Solenoid or toroid
(30-6)	$\mathbf{H} = \dfrac{q\mathbf{v}\times\mathbf{1}_r}{4\pi r^2}$	$\mathbf{H} = \dfrac{q\mathbf{v}\times\mathbf{1}_r}{r^2 c}$	Moving charge
(30-7)	$\mathbf{F} = p(\mathbf{H} - \mathbf{v}\times\mathbf{D})$	$\mathbf{F} = p\left(\mathbf{H} - \dfrac{\mathbf{v}}{c}\times\mathbf{D}\right)$	Pole

field, for example, if a compass needle moves between the plates of a charged capacitor.

TABLE 30-2 CONVERSION FACTORS RELATING MKS AND GAUSSIAN UNITS

Quantity	Symbol	MKS Unit	Gaussian Unit	
Pole	p	1 weber	$= \frac{10^8}{4\pi}$ unit pole	(emu)
Magnetic field intensity	H	$1\ \frac{\text{nt}}{\text{weber}} = 1\ \frac{\text{amp}}{\text{m}}$	$= 4\pi \times 10^{-3}$ oersted	(emu)
Charge	q	1 coul	$= 3 \times 10^9$ stcoul	(esu)
Current	I	1 amp	$= 3 \times 10^9$ statamperes	(esu)
Displacement	D	1 coul/m^2	$= 3 \times 10^5$ stcoul/cm^2	(esu)

Problems

30-1. Determine (a) in mks units and (b) in Gaussian units the intensity of the magnetic field at the center of a circular coil of 6 cm radius when it carries a current of 15 amp.

30-2. Determine (a) in mks units and (b) in Gaussian units the magnetic field intensity at the center of a closely wound circular coil of 75 turns, whose average radius is 9 cm, when the coil carries a current of 4 amp.

30-3. A circular coil of 20 cm radius consists of a single turn of wire and has a small compass needle suspended at its center so that it can swing freely about a vertical axis. The plane of the coil is set parallel to the earth's magnetic field. If the horizontal component of the earth's magnetic field is 0.2 oersted, determine the angle through which the compass needle is deflected when a current of 3 amp is sent through the coil.

30-4. Determine (a) in mks units and (b) in Gaussian units the magnetic field intensity at a distance of 0.25 m from a long straight wire in which there is a current of 32 amp.

30-5. A small compass needle, whose magnetic moment is 40 cgs poles cm and whose moment of inertia is 6.0 gm cm^2, is placed at a point 15 cm from a long straight wire. When a current is sent through the wire, the needle oscillates with a period of 7.2 sec. Determine (a) the magnetic field intensity at the position of the compass and (b) the current in the wire.

30-6. A solenoid 0.75 m long and 0.08 m in diameter is wound with 400 turns of wire. Determine the magnetic field intensity inside this solenoid when it carries a current of 6.5 amp.

30-7. A long straight wire carries a current of 25 amp. The north pole of a long bar magnet is placed 5 cm from this wire. If the pole strength of this magnet is 40 cgs poles, determine the direction and magnitude of the force on this north pole.

30-8. A long straight wire carries a current of 20 amp. The wire is parallel

to the z axis, and the current flows in the positive z direction. In addition to the magnetic field generated by the current in the wire, there is an external magnetic field in the positive z direction of intensity 5 amp/m. Find the resultant magnetic field (a) at a point in the x-y plane whose coordinates are (5 m, 0) and (b) at a point in the x-y plane whose coordinates are (0, 10 m).

30-9. Derive an equation, in Gaussian units, for the magnetic field intensity at a point in the neighborhood of a long straight wire carrying current.

30-10. Derive an equation, in Gaussian units, for the magnetic field intensity at the center of a circular loop of wire.

30-11. Find the magnetic field intensity within a toroid of mean radius 10 cm wound with 100 turns of wire when the current in the toroid is 4 amp. The circular cross section of the toroid has a diameter of 1 cm.

30-12. Two long straight wires are 18 cm apart and carry currents of 36 amp each. Determine the magnetic field intensity at a point midway between them (a) when these currents are in the same direction and (b) when the currents are in opposite directions. Express your answers in both Gaussian and mks units.

30-13. In the Oersted experiment a long straight wire placed 4 cm above a compass needle carries a current of 28 amp directed to the north. Determine the torque on the compass needle if its magnetic moment is 75 cgs pole cm.

30-14. Find the force on a magnetic pole of 10^{-4} weber moving in the $+x$ direction with a speed of 10^3 m/sec between the plates of a parallel-plate capacitor in vacuum. The plates of the capacitor are 1 cm apart and are charged to a potential difference of 500 volts. The electric field between the plates of the capacitor is in the $+y$ direction.

31

Forces on Moving Charges and Currents

31-1 Force on a Charge Moving in a Magnetic Field

Ampère was the first to show that wires carrying current experienced forces when placed in magnetic fields. Later it was shown that charged particles moving in magnetic fields also experience forces. Common applications of these phenomena are the electric motor, the galvanometer, and the cathode-ray tube. Since a moving electric charge is equivalent to a current, we shall first consider the force acting on a charge q moving with velocity $\mathbf{v}$ in a magnetic field of induction $\mathbf{B}$. Experiment shows that the force $\mathbf{F}$ acting on the charge q is at right angles to the directions of both $\mathbf{v}$ and $\mathbf{B}$. This provides us with a basic distinction between the two magnetic field vectors $\mathbf{B}$ and $\mathbf{H}$.

We have already shown that the force on a magnetic pole at rest is determined by the magnetic intensity $\mathbf{H}$. We have seen that moving charges and currents produce $\mathbf{H}$. The relation between the magnetic induction $\mathbf{B}$ and the magnetic intensity $\mathbf{H}$ has been given in Equation (29-15a) as

$$\mathbf{B} = \mu\mathbf{H},$$

where μ is the permeability of the medium.

The force $\mathbf{F}$ on a charge q moving with velocity $\mathbf{v}$ in a field of induction $\mathbf{B}$ is given by

$$\mathbf{F} = k_3 q\mathbf{v}\times\mathbf{B}. \tag{31-1}$$

As shown in Figure 31-1 the force is perpendicular to the plane formed by the vectors $\mathbf{v}$ and $\mathbf{B}$ and is directed in accordance with the right-hand rule for the vector product. The constant k_3 is a constant of proportionality which is assigned the value unity in the mks system of units. Thus if the force $\mathbf{F}$ is expressed in newtons, the charge q in coulombs, the velocity $\mathbf{v}$ in meters per second, and the magnetic induction $\mathbf{B}$ in webers per square meter, Equation (31-1) may be rewritten as

$$\boxed{\mathbf{F} = q\mathbf{v}\times\mathbf{B}.} \tag{31-2}$$

We may consider Equation (31-2) as a definition of **B**. From the above equation we see that if **v** is parallel to **B**, the force on the moving charge is zero. This suggests a method of determining the direction of **B** by use of cathode-ray tubes. If a charge of 1 coul moving perpendicular to the field with a velocity of 1 m/sec experiences a force of 1 nt, then the field has an induction of 1 weber/m^2.

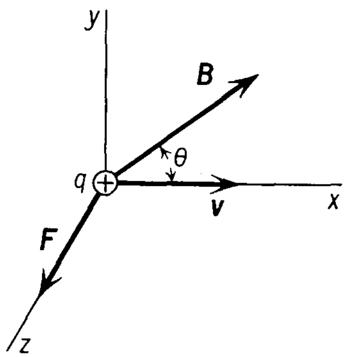


Fig. 31-1 The force **F** exerted on a charge q moving with velocity **v** in a magnetic field of induction **B**. **B** and **v** are in the x-y plane; **F** is along the z axis.

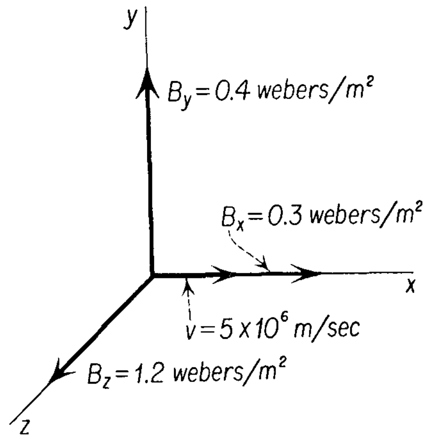


Fig. 31-2

Illustrative Example. A proton whose charge is 1.60×10^{-19} coul is moving in the x direction with a velocity of 5×10^6 m/sec in a magnetic field whose components are $B_x = 0.3$ weber/m^2, $B_y = 0.4$ weber/m^2, $B_z = 1.2$ weber/m^2. See Figure 31-2. Find the x, y, and z components of the force on the proton.

The x component of the field is parallel to the velocity; thus the force on the moving charge due to B_x is zero.

From Equation (31-2) the direction of the force on the particle due to the y component of the field is in the z direction, and since the angle between **v** and $\mathbf{B}_y$ is 90° we have

$$F_z = qvB_y = 1.60 \times 10^{-19}\text{ coul} \times 5 \times 10^6\text{ m/sec} \times 0.4\text{ weber/m}^2,$$

so that

$$F_z = 3.2 \times 10^{-13}\text{ nt.}$$

Applying the right-hand rule for the vector product, we see that the force on the particle due to the z component of the field is in the $-y$ direction. The angle between **v** and $\mathbf{B}_z$ is 90°. The y component of the force on the particle is therefore

$$F_y = -1.60 \times 10^{-19}\text{ coul} \times 5 \times 10^6\text{ m/sec} \times 1.2\text{ weber/m}^2,$$

$$F_y = -9.6 \times 10^{-13}\text{ nt.}$$

The resultant force F is of magnitude

$$F = (F_y^2 + F_z^2)^{1/2} = 9.74 \times 10^{-13}\,\text{nt}.$$

In the Gaussian system of units the constant k_3 has the value $1/c$, and we must express the units of B in gausses and of H in oersteds. As before, we must express q in statcoulombs, I in statamperes, v in centimeters per second, and F in dynes. The relationship between B and H is

$$\mathbf{B} = \kappa_m \mathbf{H}. \tag{31-3}$$

While κ_m is a pure number, we may think of it as expressed in units of gausses per oersted for convenience in calculation.

The relationship between gausses and webers per square meter may be found by calculating, from Equation (31-1), the magnetic induction in gausses required to exert a force of 1 nt on a charge of 1 coul moving perpendicularly to the magnetic field with a velocity of 1 m/sec. By definition, this field is 1 weber/m^2. The known numerical values are

$$\begin{aligned} F &= 1\,\text{nt} = 10^5\,\text{dynes},\\ q &= 1\,\text{coul} = 3 \times 10^9\,\text{stcoul},\\ v &= 1\,\text{m/sec} = 10^2\,\text{cm/sec},\\ c &= 3 \times 10^{10}\,\text{cm/sec},\\ \theta &= 90^\circ. \end{aligned}$$

From Equation (31-1) expressed in cgs units, we have

$$F = q\frac{v}{c}B\sin\theta,$$

$$10^5\,\text{dynes} = \frac{3 \times 10^9\,\text{stcoul} \times 10^2\,\text{cm/sec}}{3 \times 10^{10}\,\text{cm/sec}} \times B,$$

$$B = 10^4\,\text{gausses}.$$

Thus a magnetic field whose induction is 10^4 gausses will produce the same effect on a moving charge as a field whose induction is 1 weber/m^2. In the form of an equation we may write

$$1\,\text{weber/m}^2 = 10^4\,\text{gausses}. \tag{31-4}$$

31-2 Path of a Charged Particle in a Magnetic Field

When a charged particle is projected into a uniform magnetic field, the component of the field parallel to the velocity exerts no force on the particle. It is convenient to resolve the velocity of the particle into components parallel and perpendicular to the field, and to consider separately the motion of the particle perpendicular to the field.

Let us consider a charged particle that is moving with velocity $\mathbf{v}$ perpendicular to a magnetic field of induction $\mathbf{B}$. The particle will experience a force Bqv constant in magnitude and directed perpendicularly to the velocity of the particle. Hence the particle will move in a circular path with uniform speed v. The force exerted by the magnetic field on the charged particle is the centripetal force which produces an acceleration $a = v^2/r$, where r is the radius of the circular path. Applying Newton's second law to this case, we get

$$F = \frac{mv^2}{r} = Bqv,$$

from which

$$r = \frac{mv}{Bq}. \tag{31-5}$$

If the motion of the charged particle is not transverse to the field, the symbol v in Equation (31-5) must be interpreted as the component of the velocity which is perpendicular to the magnetic field. The motion of the

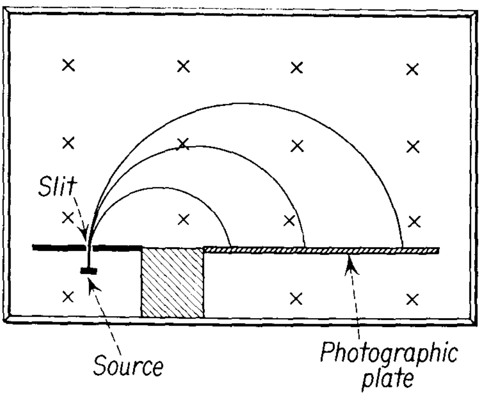


Fig. 31-3 Beta-ray spectrometer.

particle will then consist of a motion of translation parallel to the field and, at the same time, a motion in a circular orbit in a plane perpendicular to the field. The effect of these two simultaneous movements is to produce a corkscrewlike motion, or a helical motion whose axis is parallel to the magnetic field.

Note that the force on the particle moving in the magnetic field is always perpendicular to its velocity. The magnetic field therefore *does no work* on the charged particle, but only deflects it. This property of a uniform magnetic field may be utilized to measure the velocities of charged particles. In a *beta-ray spectrometer*, used to measure the velocities of electrons emitted in radioactive disintegrations, electrons or beta rays are emitted into a uniform magnetic field of known magnetic induction B in a

direction perpendicular to the field. The particles are deflected into circular orbits and strike a photographic film after completing a semicircle, as shown in Figure 31-3. The diameter of the orbit can be measured from the photographic film; the momentum mv of the electrons can then be determined, and the energy of the emitted particles may also be calculated.

The deflection of charged particles in magnetic fields is today widely utilized in scientific and technical apparatus such as beta-ray spectrometers,

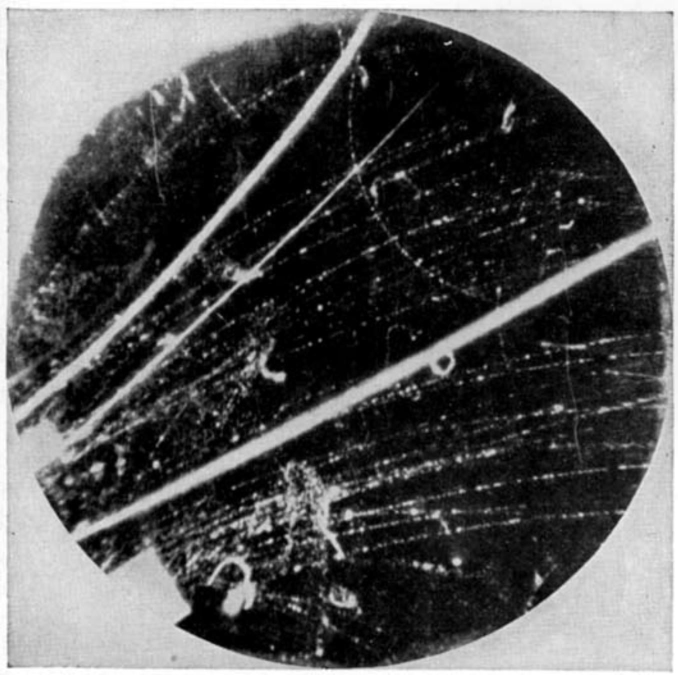


Fig. 31-4 Cloud chamber photograph showing curved paths of particles moving at right angles to a magnetic field. The magnetic field is directed into the paper and the particles originate from a source at the left. The three heavy tracks are those of protons; the numerous light tracks curved in the opposite direction are those of electrons. (Photograph by H. R. Crane.)

mass spectrometers, cathode-ray tubes (as used in television), the cyclotron (accelerates charged particles for nuclear research), the magnetron (a vacuum tube for the production of microwaves used in radar), and many others. The deflection of charged particles in a magnetic field may be seen in the accompanying photograph of tracks in a Wilson cloud chamber, shown in Figure 31-4. In this photograph the magnetic field is perpendicular to the paper and is directed into the plane of the paper.

31-3 Force on a Charged Particle in an Electromagnetic Field

We have already seen that a charged particle in an electric field of intensity **E** experiences a force $q\mathbf{E}$. This force does not depend upon the velocity of the particle. When a charged particle moves with a velocity **v**, it will experience a force given by the equation

$$\mathbf{F} = q(\mathbf{E} + \mathbf{v}\times\mathbf{B}). \tag{31-6}$$

Equation (31-6) is simply a combination of results already stated in Equations (23-1a) and (31-2). We may think of Equation (31-6) as the defining equation for the electric field intensity **E** and the magnetic induction **B**. To determine these fields in a region of space, we may imagine that a probe charge q is directed through the region. The force on the particle is determined by observing the deflection of the particle, and the electric and magnetic fields are determined by calculation from Equation (31-6). In a similar way we may imagine that Equation (30-7) is the defining equation for the magnetic intensity **H** and the electric displacement **D**, through the force on a moving pole.

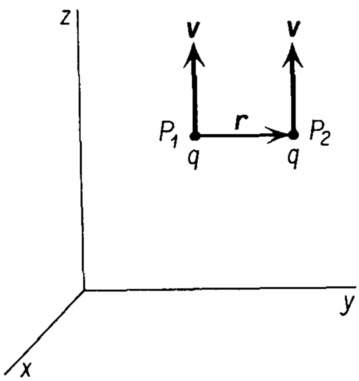


Fig. 31-5

In our initial discussion of Coulomb's law of force between magnetic poles, in Section 29-2, we stated that the constant μ_0 was to be assigned to a value of $4\pi \times 10^{-7}$ for reasons which had to be deferred to a later chapter. At this point we wish to justify the choice of a numerical value for μ_0. To do this we will calculate the ratio of the magnetic force to the electric force between two moving charged particles.

Let us consider the force between two equal charges q moving with the same velocity **v** in a direction parallel to the z axis, as shown in Figure 31-5. The instantaneous positions of the charges are P_1 and P_2, as shown in the figure. The particles are fixed in position relative to each other. The electric force on the charge at P_2 due to the presence of an equal charge at P_1 is $\mathbf{F}_e$, whose magnitude is

$$F_e = \frac{q^2}{4\pi\epsilon_0 r^2}\cdot$$

The magnetic force on the charge at P_2 may be obtained by first computing the magnetic field intensity H generated by the motion of the charge at P_1

with respect to the coordinate frame from Equation (30-6),

$$\mathbf{H} = \frac{q\mathbf{v}\times\mathbf{1}_r}{4\pi r^2}\,.$$

The relationship between **B** and **H** in vacuum is given by $\mathbf{B} = \mu_0\mathbf{H}$, so that

$$\mathbf{B} = \frac{\mu_0 q\mathbf{v}\times\mathbf{1}_r}{4\pi r^2}\,.$$

Since the velocity $\mathbf{v}$ is in the z direction and the unit vector $\mathbf{1}_r$ is in the y direction, from P_1 to P_2, the magnetic induction **B** is in the $-x$ direction. The magnitude of the magnetic force F_m on the charged particle at P_2 may be obtained from

$$\mathbf{F}_m = q\mathbf{v}\times\mathbf{B},$$

$$F_m = qvB\sin\theta.$$

At P_2 the magnetic induction **B** is in the $-x$ direction, and the velocity of the particle is in the z direction. Since the two vectors $\mathbf{v}$ and **B** are perpendicular to each other, the factor $\sin\theta$ in the vector product is equal to 1, and the magnitude of the magnetic force is equal to

$$F_m = \frac{\mu_0}{4\pi}\,\frac{q^2v^2}{r^2}\,.$$

The ratio of the magnetic force to the electric force must be a pure number, without dimensions. We obtain

$$\text{mks units} \qquad \frac{F_m}{F_e} = \mu_0\epsilon_0 v^2. \tag{31-7a}$$

Thus the quantity $\mu_0\epsilon_0$ must have the dimensions of $1/v^2$; that is, its dimensions must be $(\text{sec/m})^2$.

When the same calculation is carried out in Gaussian units, we find that

$$\text{Gaussian units} \qquad \frac{F_m}{F_e} = \frac{v^2}{c^2}\,. \tag{31-7b}$$

The ratio of the magnetic force to the electric force must be the same in whatever system of units the calculation is carried out. In Gaussian units both the velocity of the particle and the velocity of light are expressed in centimeters per second, while in mks units the velocity of the particle is expressed in meters per second. In order to equate the results of the two calculations, let us now express all velocities in meters per second. This will not alter the ratio in Equation (31-7b). We obtain

$$\mu_0\epsilon_0 v^2 = \frac{v^2}{c^2}\,,$$

or

mks units
$$c = \frac{1}{\sqrt{\mu_0 \epsilon_0}}, \tag{31-8}$$

where c is the velocity of light in meters per second.

Our procedure in developing the mks system of electrical units was first to define the size of the coulomb from the statcoulomb. This determined the choice of the constant ϵ_0. Once a choice of ϵ_0 had been made, there could be no freedom of choice of the constant μ_0, for the magnitude of that constant was determined by the connection between electric and magnetic forces, as developed in Equation (31-8).

31-4 Force on a Wire Carrying Current in a Magnetic Field

We have already seen in Section 30-7 that a charge dq moving with a velocity $\mathbf{v}$ could be considered equivalent to a current I in an element of wire $d\mathbf{s}$ through the relationship

$$\mathbf{v}\,dq = I\,d\mathbf{s}. \tag{31-9}$$

From Equation (31-2) the force on a moving charge element dq is given by

$$d\mathbf{F} = dq\,\mathbf{v}\times\mathbf{B},$$

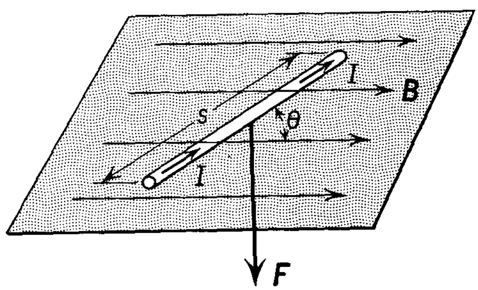


Fig. 31-6

and, substituting from Equation (31-9), we find the force $d\mathbf{F}$ on a current element $I\,d\mathbf{s}$ to be

$$d\mathbf{F} = I\,d\mathbf{s}\times\mathbf{B}. \tag{31-10}$$

This force is perpendicular to the plane formed by the current element and the direction of the field. For the special case of a long straight conductor of length s in a uniform field of magnetic induction B, the magnitude of the force F on the wire is given by

$$F = BIs\sin\theta, \tag{31-11}$$

where θ is the angle between the direction of the current in the wire and the direction of B, as shown in Figure 31-6. The force is perpendicular to the

plane formed by s and B, and its direction may be found by the right-hand rule for the vector product of Equation (31-10). The force on a wire carrying current in a magnetic field is the basis for the operation of electric motors and galvanometers.

Illustrative Example. A wire 5 m long lies along the x axis. Find the components of the force on the wire when it carries a current of 2 amp in the positive x direction if it is placed in a uniform magnetic field for which

$$B_x = 0.3 \text{ weber/m}^2,$$

$$B_y = 0.4 \text{ weber/m}^2,$$

$$B_z = 1.2 \text{ webers/m}^2.$$

Since the wire lies along the x axis, the force on the wire due to B_x is zero. The angle made by the wire with the y axis and with the z axis is 90°. The force exerted on the wire due to the y component of the field is in the $+z$ direction. Calling this force F_z, we have

$$\begin{aligned} F_z &= B_y I s \\ &= 0.4 \text{ weber/m}^2 \times 2 \text{ amp} \times 5 \text{ m} \\ &= 4.0 \text{ nt.} \end{aligned}$$

The force exerted on the wire due to the z component of the field is in the $-y$ direction. Calling this force F_y, we have

$$\begin{aligned} F_y &= -B_z I s \\ &= -1.2 \text{ weber/m}^2 \times 2 \text{ amp} \times 5 \text{ m} \\ &= -12 \text{ nt.} \end{aligned}$$

31-5 Torque on a Coil Carrying Current

Let us assume that a rectangular coil of sides a and b is carrying current I in a region of space in which there is a uniform field of magnetic induction B. If the plane of the coil is perpendicular to the direction of B, as shown in Figure 31-7 where B is directed into the paper, symmetrical elements of the coil at g and h at the top and bottom of the coil experience equal and opposite forces. Since the forces on symmetrical elements of the wire are directed along the same straight line, the resultant force on the coil and the resultant torque on the coil due to these elements are zero. Similarly, the elements at c and d contribute a net force of zero and a net torque of zero.

When the orientation of the coil is changed so that the normal to the plane of the coil makes an angle ϕ with the direction of B, as shown in Figure 31-8, the symmetrical elements g and h at the top and bottom of the coil still experience collinear forces in the $-z$ and $+z$ direction, respectively, which are equal and opposite, and therefore contribute zero net force and

torque. The forces on elements at c and d are still equal and opposite, directed parallel to the x axis in the figure, but they are no longer in the same straight line; they are now separated by a distance $a \sin \phi$ [Figure 31-8(a)].

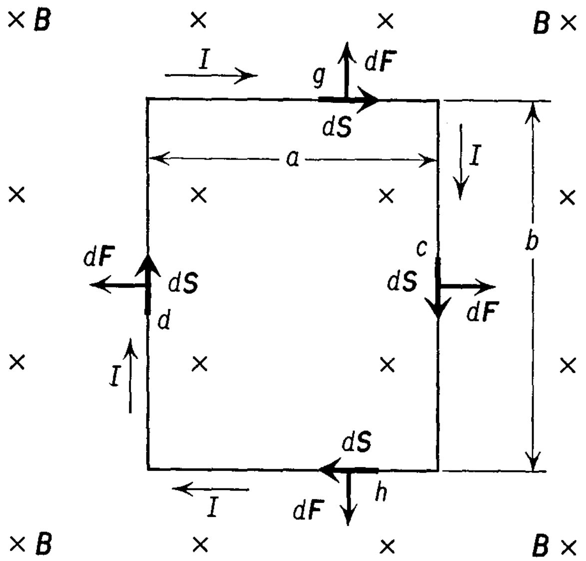


Fig. 31-7

The force on each side b of the coil is given by Equation (31-11) as

$$F = Bib$$

for the angle θ between the vertical wires, and the direction of B is 90°. The resultant force on the coil is zero, for the forces on each side of the coil are in opposite directions, but the torque G on the coil is now

$$G = BIba \sin \phi.$$

Since the area A of the coil is given by

$$A = ba,$$

we may express the torque on the coil as

$$G = BIA \sin \phi. \tag{31-12}$$

If the coil is closely wound, having N turns, a torque equal to that given by Equation (31-12) is contributed by each turn so that we have

$$G = NBIA \sin \phi. \tag{31-13}$$

Although Equation (31-13) was derived for rectangular loops or coils

of wire, it may be seen that coils of any shape, whose conductors lie in a single plane, may be approximated by a collection of rectangular coils laid side by side, as shown in Figure 31-9. Thus Equation (31-13) gives the torque on a coil of any shape when placed in a uniform magnetic field.

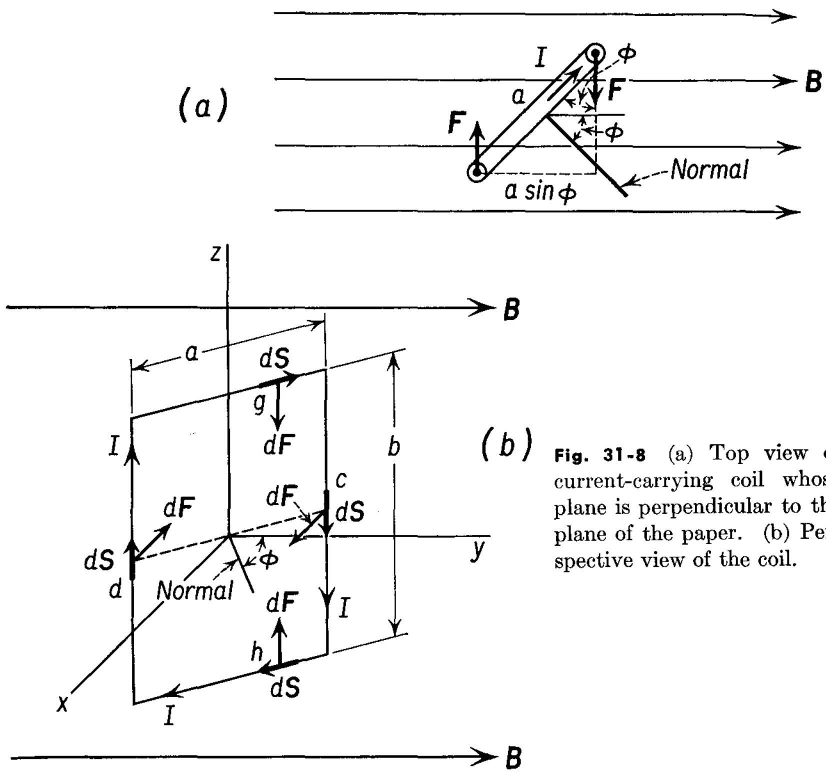


Fig. 31-8 (a) Top view of current-carrying coil whose plane is perpendicular to the plane of the paper. (b) Perspective view of the coil.

The operation of a moving-coil galvanometer, or *D'Arsonval galvanometer*, is based upon the torque on a coil in a magnetic field. The essentials of such a galvanometer are a strong permanent magnet shaped so that the two poles face each other across a gap in which the magnetic induction is fairly uniform, a rectangular coil of many turns of fine wire, and a fine spring attached to the coil. This spring may be in the form of a spirallike fine watch spring, in which case the coil is mounted on jeweled bearings, or the spring may be in the form of a long wire, one end of which is rigidly attached at the top of the frame housing the coil and the other end is attached to the coil and is part of the conductor leading current into the coil. As seen from Equation (31-13), the torque on the coil is proportional to the current through the coil. At a particular value of the current, the coil is deflected until the restoring torque of the spring is equal to the torque

generated by the magnetic field. The deflection may be indicated by a pointer moving over a scale or by means of a beam of light reflected from

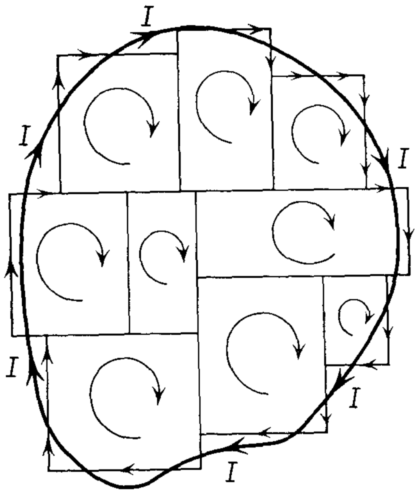


Fig. 31-9 A plane coil of arbitrary shape may be approximated by a set of rectangular coils. Note that the currents on all interior wires cancel out.

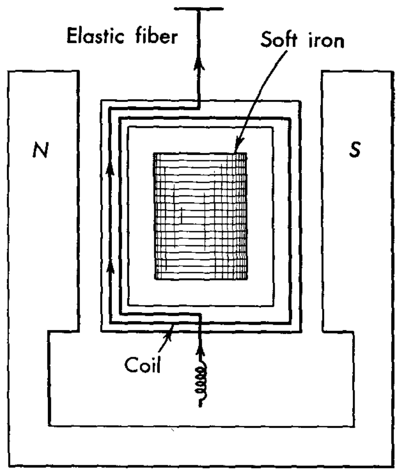


Fig. 31-10 Essentials of a D'Arsonval type of galvanometer. The soft iron core is introduced to shape the magnetic field so that it is radial, thus obtaining a galvanometer in which the deflection is proportional to the current through the coil.

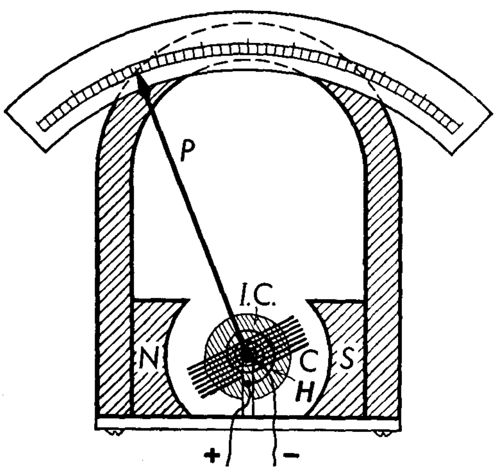


Fig. 31-11 Essentials of a moving coil galvanometer. C is the moving coil, N and S are the poles of a permanent magnet, $I.C.$ is the soft iron core, P is the pointer, and H is the hairspring.

a mirror which turns with the coil. Schematic diagrams of the essentials of a D'Arsonval galvanometer, and a typical panel-type moving-coil galvanometer are shown in Figures 31-10 and 31-11. In practice the pole

pieces are curved so that the magnetic field is nearly radial. This produces a galvanometer in which the deflection is nearly proportional to the angular displacement.

31-6 Force between Two Parallel Conductors

When two parallel conductors carry current, there is a force between them which is repulsive when the currents are in opposite directions and attractive when the currents are in the same direction. The force is due to the interaction between the current in one wire and the magnetic field at that wire due to current in the other wire. We may calculate the force between the two conductors by utilizing the results of the preceding chapter. If the current in the wire shown as 1 in Figure 31-12 is I_1, the magnetic field intensity at the position of wire 2, due to the current I_1 in wire 1, is

$$H = \frac{I_1}{2\pi a}, \qquad (30\text{-}1a)$$

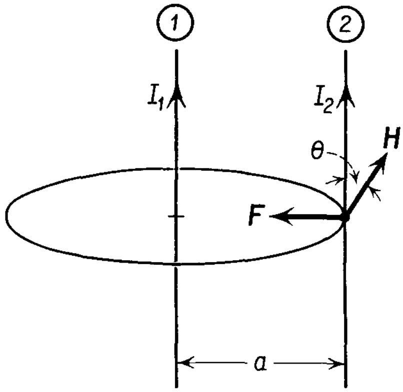


Fig. 31-12

where a is the distance between the wires. The magnetic induction B is related to the magnetic field intensity H in vacuum or in air through the equation

$$B = \mu_0 H;$$

thus the magnetic induction due to the current in wire 1 at the position of wire 2 is

$$B = \frac{\mu_0 I_1}{2\pi a}.$$

The direction of **B** is parallel to the direction of **H** and is shown in the figure as directed into the plane of the paper, perpendicular to the wire 2. We may find the force on a length s of wire 2 by application of Equation (31-11). We note that the angle θ between the direction of the current in wire 2 and the direction of **B** at wire 2 is 90°. Thus we have

$$F = BsI_2$$

$$= \frac{\mu_0 I_1 I_2 s}{2\pi a}.$$

The force per unit length of wire is given by

$$\frac{F}{s} = \frac{\mu_0 I_1 I_2}{2\pi a}. \tag{31-14}$$

From application of the rule for the vector product, as given in Equation (31-10) for the direction of the force, the force on wire 2 may be seen to be toward wire 1. An equal but opposite force is exerted on wire 2, as may be seen from an application of the same procedure to wire 2. When the currents are in opposite directions, the force is of the same magnitude but is repulsive rather than attractive.

From Equation (31-14) we note that the units of μ_0 may be expressed as nt/amp^2 as well as weber/nt m^2 or weber/amp m. The units in which μ_0 is expressed is a matter of convenience. Thus we may write

$$\mu_0 = 4\pi \times 10^{-7}\ \text{nt/amp}^2. \tag{31-15}$$

It is possible through Equation (31-14) to establish a definition of the ampere in terms of the mechanical quantities of force and distance and the constant μ_0. Thus an ampere may be defined as that current flowing in two infinitely long parallel wires separated by a distance of 1 m in air which causes a force of 2×10^{-7} nt per meter of length between them. In a device called the current balance, the force between two coils of wire connected in series and separated by a distance of a few centimeters may be used to make an absolute determination of current.

When there is an alternating current in a coil, the current is first in one direction, then in the other, following a sine or cosine function of time. The currents in adjacent turns are in the same direction, so that there is an attractive force between adjacent turns which varies from some maximum value to zero and back again. This periodic force may cause the wires of the coil to vibrate and produce an audible hum; such a noise is sometimes heard in a transformer of a radio set.

Illustrative Example. Find the force between two parallel conductors in a power distributing station. The conductors are 10 m long, carry a current of 150 amp each, and are separated by a distance of 1 cm.

From Equation (31-14)

$$F = \frac{\mu_0 I_1 I_2 s}{2a}.$$

Substituting numerical values, we have

$$F = \frac{4\pi \times 10^{-7}\ \text{nt/amp}^2 \times 150\ \text{amp} \times 150\ \text{amp} \times 10\ \text{m}}{2 \times 0.01\ \text{m}},$$

$$F = 4.5\ \text{nt}.$$

TABLE 31-1 PRINCIPAL EQUATIONS IN MKS AND GAUSSIAN UNITS

Equation	MKS	Gaussian	
(31-2)	$\mathbf{F} = q\mathbf{v} \times \mathbf{B}$	$\mathbf{F} = q\dfrac{\mathbf{v} \times \mathbf{B}}{c}$	Moving charge
(31-3)	$\mathbf{B} = \kappa_m \mu_0 \mathbf{H}$	$\mathbf{B} = \kappa_m \mathbf{H}$	
(31-5)	$r = \dfrac{mv}{Bq}$	$r = \dfrac{mvc}{Bq}$	Circular path
(31-6)	$\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B})$	$\mathbf{F} = q\left(\mathbf{E} + \dfrac{\mathbf{v} \times \mathbf{B}}{c}\right)$	Charge
(31-10)	$d\mathbf{F} = I\, d\mathbf{s} \times \mathbf{B}$	$d\mathbf{F} = \dfrac{I}{c}\, d\mathbf{s} \times \mathbf{B}$	Element of wire carrying current
(31-13)	$G = NBIA \sin \phi$	$G = \dfrac{NBIA \sin \phi}{c}$	Torque on coil
(31-14)	$\dfrac{F}{s} = \dfrac{\mu_0 I_1 I_2}{2\pi a}$	$\dfrac{F}{s} = \dfrac{2I_1 I_2}{ac^2}$	Parallel wires

TABLE 31-2 CONVERSION FACTORS RELATING MKS AND GAUSSIAN UNITS

Quantity	Symbol	MKS Unit	Gaussian Unit	
Magnetic intensity	H	1 amp/m = 1 nt/weber	$= 4\pi \times 10^{-3}$ oersted	(emu)
Magnetic induction	B	1 weber/m^2	$= 10^4$ gausses	(emu)
Electric intensity	E	1 nt/coul = 1 volt/m	$= \dfrac{10^{-4}}{3} \dfrac{\text{statvolt}}{\text{cm}}$	(esu)
			$= \dfrac{10^{-4}}{3} \dfrac{\text{dyne}}{\text{stcoul}}$	(esu)
Electric displacement	D	1 coul/m^2	$= 3 \times 10^5$ stcoul/cm^2	(esu)
Charge	q	1 coul	$= 3 \times 10^9$ stcoul	(esu)
Current	I	1 amp	$= 3 \times 10^9$ statamperes	(esu)

Problems

31-1. A proton of charge $+4.8 \times 10^{-10}$ stcoul moves with a speed of 10^9 cm/sec at right angles to the direction of a magnetic field of 300 gausses. Find the force on the proton due to its motion with respect to the magnetic field.

31-2. The mass of a proton is 1.67×10^{-24} gm. What will be the radius of the circle in which the proton of Problem 31-1 moves?

31-3. In a beta-ray spectrometer an electron moving perpendicularly to a magnetic field of magnetic induction 150 gausses moves in a circle of 15 cm diameter. Find the momentum of the electron.

31-4. A proton moves with a velocity of 10^4 m/sec in the $+x$ direction in a magnetic field whose components are $B_x = 0.03$ weber/m^2, $B_y = 0.04$ weber/m^2, $B_z = 14$ webers/m^2. Find the components of the force on the proton.

31-5. A particle of charge 10 μcoul has a velocity whose components are $v_x =$ 3 m/sec, $v_y = 4$ m/sec. The particle moves in a magnetic field whose components are $B_x = 0.5$ weber/m^2, $B_y = 1.2$ webers/m^2, and $B_z = 1.5$ webers/m^2. Find the components of the force on the particle.

31-6. The magnetic induction between the poles of a large electromagnet is 1,000 gausses in a direction from east to west. The field extends vertically for a distance of 5 cm. A wire is suspended vertically in the field, and a current of 15 amp is sent through it. Determine the magnitude and direction of the force on the wire.

31-7. A wire parallel to the y axis carries a current of 10 amp. The wire is 0.5 m long. The magnetic induction in the vicinity of the wire has an x component of 0.3 weber/m^2, a y component of -1.2 webers/m^2, and a z component of 0.5 weber/m^2. Find the components of the force on the wire.

31-8. A current of 8 amp is directed downward in a vertical wire 80 cm long which is suspended in a place where the intensity of the earth's magnetic field has a horizontal component of 0.25 oersted and a vertical component of 0.75 oersted. Determine the magnitude and the direction of the force on the wire.

31-9. A long uniform solenoid is wound with 1,000 turns/m of length and carries a current of 10 amp. An electron, whose velocity component parallel to the axis of the solenoid is 10^5 m/sec and whose velocity perpendicular to the axis of the solenoid is 10^6 m/sec, is ejected from a radioactive source at the center of the solenoid. The path described by the electron is a helix. (a) Find the radius of the cross section of this helix and (b) find the pitch (number of turns per meter) of the helix. The charge of an electron is -4.8×10^{-10} stcoul, and the mass of an electron is 9.10×10^{-28} gm.

31-10. A rectangular coil of 150 turns of fine wire is suspended in a magnetic field with its plane parallel to the direction of the field. The dimensions of the coil are 2 cm by 4 cm. When a current of 0.5 amp is sent through the coil, a torque of 18,000 dynes cm acts on it. Determine the magnetic induction of this field.

31-11. A straight wire 12 cm long is placed in a magnetic field of 350 gausses directed at right angles to the length of the wire. A delicate spring balance attached to the wire measures a force of 160 dynes when a current is sent through the wire. Determine the current in the wire.

31-12. A proton whose mass is 1.65×10^{-27} kg travels in a circular path of 0.35 m radius in a magnetic field of 1.2 webers/m^2. Determine (a) the velocity of the proton and (b) its period of revolution.

31-13. A rectangular galvanometer coil has 500 turns wound on a frame 2 cm by 4 cm. It hangs between the poles of a magnet where the flux density is 0.080 weber/m^2. Determine the torque which acts on this coil when the current in it is 3×10^{-8} amp.

31-14. Derive Equation (31-7b).

31-15. Two electrons are moving down the axis of a television tube with a velocity of 10^7 m/sec. The electrons are separated by a distance of 0.01 cm. (a) What is the electric force between them? (b) What is the magnetic force between them? (c) What is the resultant force between them?

31-16. A circular coil of wire of radius 10 cm and containing 50 turns is placed in a magnetic field in which B is 1.5 webers/m^2. The current in the coil is 3 amp. (a) What is the torque on the coil when its plane is parallel to the field? (b) When the normal to the plane of the coil makes an angle of 30° with the field?

31-17. One end of a flexible copper wire is fastened to a post while the other goes over a fixed pulley and has a weight W hanging from it. The portion of the wire from A to the pulley is horizontal and is in a uniform magnetic field of flux density B in the vertical direction. A current I is sent through the wire; determine the radius of the circular arc into which the wire is bent.

32

Electromagnetic Induction

32-1 Motion of a Wire in a Magnetic Field

When a wire moves through a uniform magnetic field of induction B, in a direction at right angles to the field and to the wire itself, the electric charges within the conductor experience forces due to their motion through this magnetic field. The positive charges are held in place in the conductor by the action of interatomic forces, but the free electrons, usually one or two per atom, are caused to drift to one side of the conductor, thus setting up an electric field E within the conductor which opposes the further drift of electrons. The magnitude of this electric field E may be calculated by equating the force it exerts on a charge q, to the force on this charge due to its motion through the magnetic field of induction B; thus

$$Eq = Bqv,$$

from which

$$E = Bv.$$

If, as a result of the motion of the wire through the magnetic field, a charge q is moved a distance s along the wire against the internal electric field E, a quantity of work $\mathscr{W}$ is done by the agency moving the wire, given by the expression

$$\mathscr{W} = Eqs = Bvqs.$$

Thus an electromotive force is generated within the wire as a result of its motion through the magnetic field. The electromotive force across the ends of the wire is the work per unit charge done by the agency moving the wire. The emf $\mathcal{E}$ is thus

$$\mathcal{E} = \frac{\mathscr{W}}{q} = \frac{Bqvs}{q},$$

so that

$$\boxed{\mathcal{E} = Bsv.} \tag{32-1}$$

As shown in Figure 32-1, the direction of the emf is the direction in which positive charges are made to move by the action of the magnetic field, and therefore is opposite to the direction of the induced electric field E within the wire.

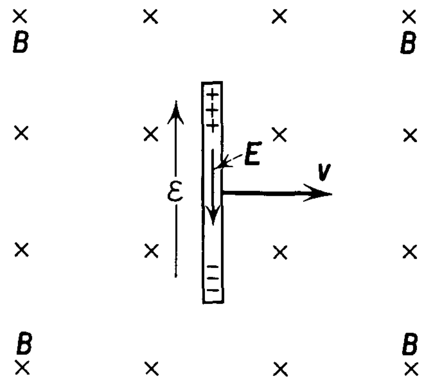


Fig. 32-1

To gain further insight into the effect of moving a wire through a magnetic field, let us suppose that the wire of length s slides over a fixed conductor a consisting of two parallel tracks which are electrically connected at one end, as shown in Figure 32-2. As the wire moves to the right with velocity v, the induced emf in the wire produces a current I in the closed circuit, in the direction of the emf $\mathcal{E}$ in the moving wire. We have seen in Section 31-4 that a wire carrying current in a direction perpendicular to the magnetic field experiences a force given by

$$F = BIs.$$

In the figure this force is directed to the left. In order to satisfy the principle of conservation of energy, the agency moving the wire to the right

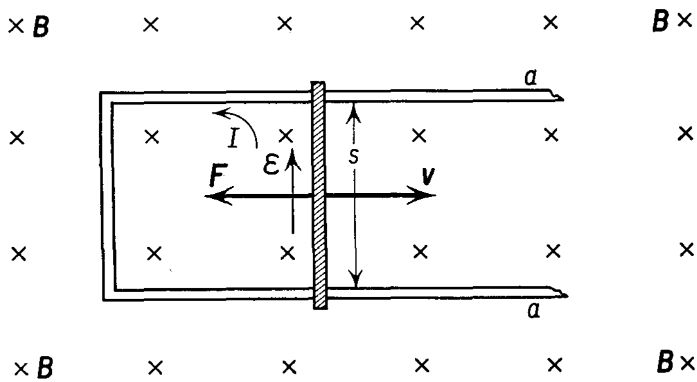


Fig. 32-2

must exert a force equal and opposite to the force F above, and expend mechanical power $\mathcal{P}$ such that

$$\mathcal{P} = Fv = BIsv.$$

At the same time the electrical power generated is

$$\mathcal{P} = \mathcal{E}I.$$

Equating the mechanical power expended and the electrical power generated, we find

$$\mathcal{E} = Bsv,$$

which is Equation (32-1).

In calculating the emf developed by a wire moving through a magnetic field, we have used two different points of view. The first calculation was made from essentially a microscopic point of view in which our attention was directed to the forces on isolated charges within the wire. The second calculation was made from a macroscopic viewpoint, in which our attention was directed to the force on the wire and to the emf. The same result was obtained in each case.

Many practical devices, such as electric generators and motors, are designed so that conductors move across magnetic fields. In using Equation (32-1) to discuss the operation of these devices, it must be remembered that B, s, and v were all considered to be perpendicular to one another. If they are not mutually perpendicular in a particular case, then only the components of the three quantities which are mutually perpendicular are to be considered. As the wire moves through the magnetic field, it is often described as "cutting" the lines of magnetic induction. Equation (32-1) then shows that the emf induced in a wire depends upon the number of lines of magnetic induction cut per unit time.

32-2 Magnetic Flux and Flux Density

It is convenient to represent the magnetic induction B by lines of magnetic induction, sometimes called lines of magnetic *flux*. The direction of the magnetic induction is tangent to the flux lines, and the magnitude of the magnetic induction is given in the usual way by the number of lines per unit area passing through a surface perpendicular to the flux lines. The total number of lines passing perpendicularly through an element of area is then called the *magnetic flux* Φ (capital phi) through that area. If the area of an element perpendicular to B is ΔA, then the flux $\Delta\Phi$ through that element is given by

$$\Delta\Phi = B\,\Delta A. \tag{32-2}$$

In the mks system of units, the magnetic induction is stated in units of webers per square meter, and the area is stated in units of square meters. The flux is expressed in units of webers. For this reason the magnetic induction B is often referred to as the *flux density*.

In the Gaussian system of units, the magnetic induction B is expressed in gausses, the area in square centimeters, and the flux in *maxwells*. We have already seen that

$$1 \text{ weber/m}^2 = 10^4 \text{ gausses},$$

so that

$$1 \text{ weber} = 10^4 \text{ gausses} \times 10^4 \text{ cm}^2,$$

or

$$1 \text{ weber} = 10^8 \text{ maxwells}.$$

Following the procedure we have used throughout the development of electricity and magnetism, unless otherwise indicated all equations are expressed in the mks system of units, and the principal ones will be restated in the Gaussian system of units in a table at the end of this chapter.

The magnetic flux Φ is a scalar quantity, but the magnetic induction $\mathbf{B}$ is a vector quantity; the area $\Delta\mathbf{A}$ may be considered as a vector quantity. In dealing with closed surfaces, as in Gauss's theorem in electrostatics, we considered the direction of an area as that of an outward drawn normal. Although in the present instance the area is not part of a closed surface, it may be thought of as a *film* or a *cap* bounded by a closed conducting boundary. In choosing the direction of the area vector, we must associate a positive direction of circulation around the boundary of the area in accordance with a *right-hand rule*. Thus if in Figure 32-2 the direction of the area vector is chosen as pointing toward the reader, the positive direction of the current in the wire may be found by directing the thumb of the right hand in the direction of the area vector. The curled fingers of the right hand indicate the positive direction of the current or the emf as in the counterclockwise direction. If the area vector is pointing into the paper, the positive direction of the current is clockwise. Following this convention, we may rewrite Equation (32-2) in vector form as the scalar product of $\mathbf{B}$ and $\Delta\mathbf{A}$ as

$$\Delta\Phi = \mathbf{B}\cdot\Delta\mathbf{A}. \tag{32-2a}$$

32-3 Faraday's Law of Electromagnetic Induction

The phenomenon of electromagnetic induction was discovered in 1831 by Michael Faraday (1791–1867) in England and independently by Joseph Henry (1797–1878) in the United States. One example of electromagnetic induction is the emf generated in a wire moving through a magnetic field, as discussed in Section 32-1. In Faraday's original experiment the apparatus consisted essentially of two neighboring circuits, shown in Figure 32-3; one circuit, which we shall call the primary circuit, contained a battery B, a coil P, and a key K, for opening and closing the circuit; the second circuit, or secondary circuit, consisted of a coil S and a galvanometer G. Faraday observed that when the key was closed, the galvanometer

in the S circuit gave a momentary deflection and then returned to its zero position and remained there as long as the key was closed. When the key was opened, there was another momentary deflection of the galvanometer, opposite in direction to the previous deflection, and then the galvanometer needle returned to its zero position.

Analyzing this simple experiment, we find that when the key in the primary circuit was closed, a current started flowing through the primary coil P. This current produced a magnetic field in the neighborhood of P and also around the coil S; that is, a change was produced in the magnetic field around the coil S in the secondary circuit. The fact that the galvanometer showed only a momentary deflection can be interpreted by saying that a current was *induced* in the secondary circuit momentarily and that this induced current was due to the *change* in the magnetic field around the secondary circuit. As long as the current in the primary circuit remained constant, the magnetic field around both P and S remained constant, but the galvanometer read zero during this time. But when the magnetic field was again changed, say by opening the key, a current was again induced in the secondary circuit, this time in a direction opposite to that produced when the key was closed.

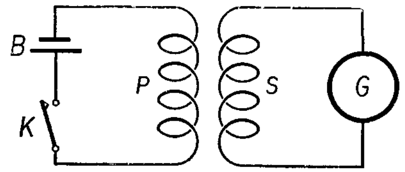


Fig. 32-3 Faraday's experiment on electromagnetic induction.

The results of the above experiment on electromagnetic induction can be explained qualitatively by stating that the change in the magnetic field around the secondary coil *induced an electromotive force* in the coil, and, since the coil is part of a closed circuit, this induced emf produced a current in the secondary circuit.

There are many ways in which the magnetic field around the coil S may be varied. Suppose, for example, that the key of the primary circuit is kept closed so that a steady current flows through the coil P. As long as the magnetic field around the secondary coil S remains constant, there will be no emf induced in it, and the galvanometer will read zero. But if we move S away from P so as to decrease the magnetic field around S, the galvanometer will show a deflection; similarly, if we move S toward P, the galvanometer will show a deflection but now in the opposite direction.

If we put a variable resistor in the primary circuit so that the current in it may be varied, and keep the distance between S and P constant, there will be an induced emf and hence an induced current in the secondary circuit whenever the current in the primary circuit is changed. When the

current in the primary coil is increased, the induced current in the secondary coil will be in one direction; when the current in the primary coil is decreased, the induced current in the secondary coil will be in the opposite direction.

Fig. 32-4 Michael Faraday (1791–1867). Chemist and physicist. Discovered the laws of electrolysis and electromagnetic induction. Introduced the concept of lines of force to help understand the phenomena associated with electric and magnetic fields. (Courtesy of *Scripta Mathematica.*)

The results of many experiments on electromagnetic induction can be stated in the form of a law known as *Faraday's law of electromagnetic induction* as follows:

The electromotive force induced in each turn of wire in any circuit is equal to the time rate of decrease of the magnetic flux through it, or, in the form of an equation:

$$\boxed{\mathcal{E} = -\frac{d\Phi}{dt}.} \tag{32-3}$$

Faraday's law may be considered as one of the fundamental empirical laws of electromagnetism, or it may be derived by applying Ampère's law and the principle of conservation of energy to typical cases. To show this, let us reconsider the case of a wire moving perpendicularly with constant velocity through a magnetic field of flux density **B**, as shown in Figure 32-2. If the wire moves a distance Δx to the right, the change in area is

$$\Delta A = s\,\Delta x.$$

In accordance with the sign convention, the positive direction of the normal to ΔA is upward toward the reader. The flux density **B** is directed down-

ward, so that $\mathbf{B} \cdot \Delta\mathbf{A}$, the change in flux $\Delta\Phi$ through this area, is

$$\Delta\Phi = -Bs\,\Delta x.$$

If this change in flux takes place in a short time interval Δt, then

$$\frac{\Delta\Phi}{\Delta t} = -Bs\frac{\Delta x}{\Delta t},$$

and, in the limit of short time intervals, we find

$$\frac{d\Phi}{dt} = -Bs\frac{dx}{dt} = -Bsv.$$

Substituting for the quantity on the right-hand side of the above equation from Equation (32-1), we find

$$\mathcal{E} = -\frac{d\Phi}{dt}\cdot \tag{32-3}$$

Although we have derived the result for a moving wire, Equation (32-3) has been found to be true for the emf induced in any closed circuit when the flux through that circuit changes with time.

Whenever the magnetic flux passing through a single turn of wire is changing, the instantaneous emf developed in the loop is given by Equation (32-3), regardless of the reason for the change in the flux. If the coil consists of N turns of wire, an emf will be induced in each turn by the changing magnetic flux. The turns of a coil may be considered as connected in series, so that the emf in the coil will be the sum of the emf's induced in the individual turns. If the rate of change of magnetic flux is the same through each turn, then the total emf induced in the coil will be given by

$$\boxed{\mathcal{E} = -N\frac{d\Phi}{dt}\cdot} \tag{32-4}$$

Illustrative Example. A coil containing 750 turns of wire is wound on a rectangular frame 20 cm by 30 cm. The magnetic induction normal to the area of the coil is 0.3 weber/m². The magnetic field is reduced to zero at a uniform rate in 0.25 sec. Determine the magnitude of the emf induced in the coil.

The total flux through the coil is

$$\begin{aligned}\Phi &= BA\\ &= 0.3\,\frac{\text{weber}}{\text{m}^2} \times 0.06\ \text{m}^2\\ &= 0.018\ \text{weber}\end{aligned}$$

The rate of change of flux is given by

$$\frac{d\Phi}{dt} = \frac{0.018 \text{ weber}}{0.25 \text{ sec}}$$

$$= 0.072 \text{ weber/sec.}$$

Since $N = 750$ turns, we have, from Equation (32-4), neglecting the sign of the emf,

$$\mathcal{E} = 750 \times 0.072 \text{ volts,}$$

$$\mathcal{E} = 54 \text{ volts.}$$

32-4 Lenz's Law

The method for determining the direction of the current in a coil produced by an induced emf was first clearly stated by H. F. E. Lenz (1804-1864) in 1834; it is based upon the application of the principle of conservation of energy to the process of electromagnetic induction. Lenz's law states that *the induced current is in such a direction as to oppose, by its magnetic action, whatever change produces the current.*

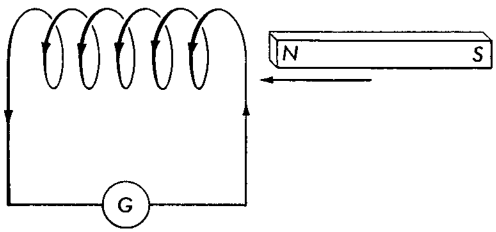


Fig. 32-5 An emf is induced in the coil by the motion of a bar magnet. Direction of the induced current is shown by the arrows on the wire.

Another way of stating Lenz's law is that the direction of the induced current is such as to oppose the change in the magnetic flux in the circuit. If for any reason whatever there is an increase in the magnetic flux through the circuit, the induced current will be in such a direction as to set up a magnetic field to oppose the increase in the magnetic flux through it. Similarly, if there is a decrease in the magnetic flux through the circuit, the induced current will be in such a direction that it will set up a magnetic field which will oppose the decrease in the flux through it. In the case of the moving wire of Figure 32-2, the motion of the wire to the right tended to increase the magnetic flux enclosed within the closed circuit. The increase in magnetic flux was directed into the plane of the paper. According to Lenz's law the induced current in this circuit had to be in the counterclockwise direction, as shown in the figure. The magnetic field generated by the induced current was directed out of the plane of the paper, so as to oppose the change in the flux through the circuit.

Let us suppose that a bar magnet is brought near a coil whose terminals are connected to a galvanometer, as shown in Figure 32-5. When the north pole of the magnet approaches the coil, the galvanometer registers a current in one direction. When the magnet is removed from the coil, the galvanometer registers a current in the opposite direction. We may find the direction of the induced currents by the application of Lenz's law. When the north pole of the bar magnet is brought near one end of the coil, the induced current in the coil will be in such a direction as to set up a magnetic field which will oppose the motion of the north pole toward it; that is, the magnetic field caused by the induced current will repel the north pole of the bar magnet. If the north pole of the magnet is moved away from one end of the coil, the induced current will set up a magnetic field so as to attract the north pole of the bar magnet and oppose its motion away from the coil. Thus work must be done in moving the bar magnet with respect to the coil because of the force which is generated when the magnet is moved. This work is transformed into electric energy, as evidenced by the existence of an induced current in the coil, and the conversion of this energy into heat.

32-5 Electric Dynamo or Generator

The moving wire of Figure 32-2 is a simple form of electric generator. In this illustrative dynamo, mechanical work is done on the wire, and this is converted into electrical energy. More practical forms of generators are usually constructed so that the motion of the conductor is rotational rather than translational. In the simplest case a coil of N turns is rotated about an axis in the plane of the coil, in a magnetic field which is perpendicular to the axis of rotation, as shown in Figure 32-6. The two ends of the coil are connected to two insulated conducting rings called *slip rings*, mounted on the axis and rotating with the coil. Two blocks of carbon, called *brushes*, press against these rings as they rotate and provide electrical contact with the external circuit.

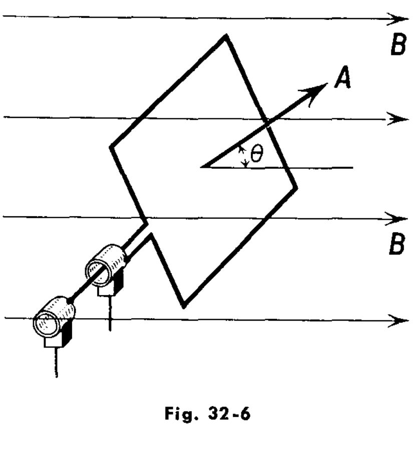


Fig. 32-6

Let us suppose that the coil of area A is rotating in a field of uniform flux density B. If the coil is oriented so that the normal to the plane of the coil makes an angle θ with the lines of flux, the component of the magnetic

field which is normal to the plane of the coil is $B \cos \theta$, so that the flux passing perpendicularly through the coil is

$$\Phi = BA \cos \theta.$$

The emf $\mathcal{E}$ generated in a coil of N turns, as a result of the changing flux through the coil as it rotates, may be obtained from Equation (32-4) as

$$\begin{aligned}\mathcal{E} &= -N \frac{d\Phi}{dt} \\ &= -NBA \frac{d}{dt}(\cos \theta) \\ &= NBA \sin \theta \frac{d\theta}{dt}.\end{aligned}$$

Let us assume that the coil rotates with uniform angular velocity ω, and that at time $t = 0$, the angle $\theta = 0°$. We have

$$\theta = \omega t,$$

and

$$\frac{d\theta}{dt} = \omega,$$

so that

$$\mathcal{E} = NBA\omega \sin \omega t. \tag{32-5}$$

The emf is zero at time $t = 0$ and varies as a sine function of the time. The emf changes direction after each half revolution of the coil. The emf is said to be an *alternating* emf. The maximum value of the emf occurs when $\theta = \omega t = 90°$; that is, when the plane of the coil is parallel to the

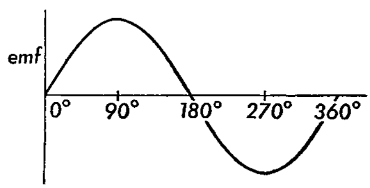


Fig. 32-7 Graph of the alternating emf induced in the coil of the generator during one revolution.

magnetic field. The emf is zero when the plane of the coil is perpendicular to the magnetic field ($\theta = 0°$). The variation in the emf induced in the coil is one revolution, as shown in Figure 32-7. The rate of change of the magnetic flux through the coil is greatest when it is passing through the position where it is parallel to the field, and is zero when the coil is perpendicular to the field.

In some small generators, such as those operated by hand and used to supply current to ring bells in rural telephones, the magnetic field may be supplied by a permanent magnet; these generators are called *magnetos.*

In most generators the magnetic field is produced by current in *field coils*. This current may be supplied by a battery or it may be supplied by the generator itself. Instead of a single rotating coil of wire, there are usually several coils, each consisting of many turns of wire, wound on an iron core, rotating in the magnetic field. The whole assembly is called an *armature*. Most large generators have a complicated field structure, with two or more pairs of poles. The induced emf in the coils of an armature is always an alternating emf, and the current in these coils is always an alternating current.

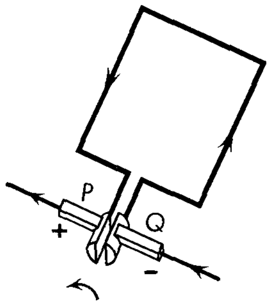


Fig. 32-8 Split-ring commutator. P and Q are brushes in contact with the segments of the commutator.

For some purposes it is desired to have a current which does not reverse its direction, as in electroplating baths. For such cases it is necessary to change the alternating current developed in the armature to current which is always in the same direction in the outside circuit. In a simple form of d-c generator, this is accomplished by a *split-ring commutator*, shown in Figure 32-8. The two ends of the armature are connected to the two insulated halves of the split ring. As the coil rotates, a given brush is always connected to that part of the coil moving in a particular direction through the field, so that one of the brushes is always the positive terminal of the generator and the other brush is always the negative terminal. The current from an armature with a single turn is not constant but is pulsating, as shown in Figure 32-9. The wave form is essentially a sine wave with the negative half cycles reversed.

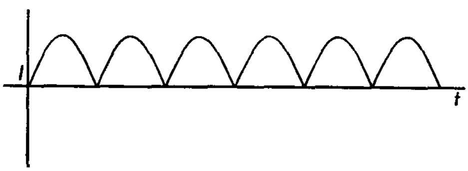


Fig. 32-9 Direct current from a single coil.

In modern d-c generators the armature consists of many coils connected in series, and the commutator contains many segments. Figure 32-10 shows the current from a generator containing two coils. The small variations in the current are referred to as commutator ripple.

A simple generator coil rotated in an unknown magnetic field may be used to measure the magnetic induction, using Equation (32-5), if the dimensions of the coil and the speed of rotation are known, and the emf generated by the rotation of the coil is measured.

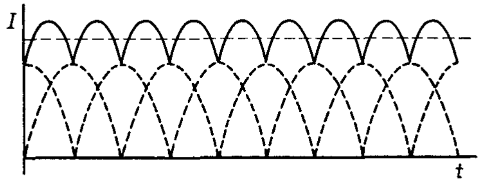


Fig. 32-10 Direct current from a d-c generator having a commutator with many segments. The dotted lines show the current from the two pairs of segments, while the solid line shows the current from such a generator. Note that the current is not constant but fluctuates about some average value shown as a dashed horizontal line.

32-6 Self-Inductance

The law of electromagnetic induction states that an emf is induced in any circuit in which the magnetic flux is changing. The manner in which the change in the magnetic flux is produced does not matter; the changes may be induced by external currents or magnets, or by changes in the circuit itself. The coil of a motor is caused to rotate by current passing through the armature, and as a result of the rotation of the armature, the flux through the coil is changed and a *back emf* is developed. When current is sent through a coil, a magnetic field is established through it, and any changes in the current generate changes in the magnetic flux through the coil. These changes in flux induce an emf in the coil, which, according to Lenz's law, must be in such a direction as to oppose the change in current. The emf induced in the coil is proportional to the rate of change of current in it, or, in the form of an equation,

$$\boxed{\mathcal{E} = -L\frac{di}{dt}\cdot} \tag{32-6}$$

The constant of proportionality L represents a property of the coil which depends upon its dimensions and its geometrical shape; L is called the *self-inductance* of the coil.

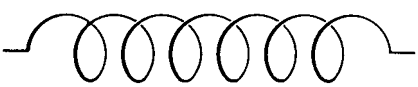


Fig. 32-11 Schematic representation of an inductor.

The minus sign is used to express the fact that the emf induced in a coil by a *change* in current is opposite to the direction of the *change.* In the mks system the unit of inductance is the *henry,* after Joseph Henry. Thus a coil has an inductance of 1 henry if an emf of 1 volt is induced in the coil when the current through it is changing at the rate of 1 amp/sec. A device having inductance is called an *inductor* and is represented schematically in Figure 32-11.

Another interpretation of the self-inductance L of a circuit may be obtained by comparing Equations (32-4) and (32-6), yielding

$$L\frac{di}{dt} = N\frac{d\Phi}{dt},$$

from which

$$Li = N\Phi,$$

so that

$$L = \frac{N\Phi}{i}. \tag{32-7}$$

The quantity $N\Phi$ is called the *flux linkage* of the circuit, hence the self-inductance L is the flux linkage per unit current of a circuit.

Let us determine the self-inductance of a uniformly wound toroid of N turns, mean length s, and cross-sectional area A. The magnetic field intensity within the toroid is uniform and given by

$$H = \frac{Ni}{s}. \tag{30-5}$$

When a toroid is in vacuum, the magnetic induction within the toroid is given by the equation

$$B = \frac{\mu_0 Ni}{s},$$

and the flux of magnetic induction within the toroid is

$$\Phi = \frac{\mu_0 NiA}{s}.$$

Since

$$L = \frac{N\Phi}{i}, \tag{32-7}$$

we get

$$L = \frac{\mu_0 N^2 A}{s}. \tag{32-8}$$

Thus the self-inductance of a toroid in air is a property of the geometry of the toroid, just as the capacitance of a capacitor is a property of its geometry. Any conducting element in an electrical circuit has the property of inductance. The conductors which connect the various parts of an electric circuit also generate a magnetic field when current passes through them. The inductance associated with the leads is often called *stray* inductance or distributed inductance.

From Equation (32-8) we see that μ_0 may be expressed in terms of the unit of inductance as

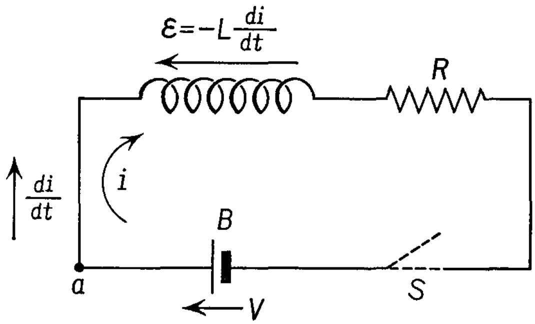
$$\mu_0 = 4\pi \times 10^{-7}\,\frac{\text{henry}}{\text{meter}}, \tag{32-9}$$

and, indeed, these are the units in which μ_0 is most commonly expressed.

32-7 Inductive Transients

Let us consider the simple circuit of Figure 32-12 in which an inductor L and a resistor R are connected in series to a battery B. Initially the switch S is open. We may analyze the behavior of the circuit by applying Kirchhoff's laws (Section 27-6) to the circuit at any instant after the switch is closed. Let us suppose that at a particular time t the current in the circuit

Fig. 32-12



is in the clockwise direction, as shown in the figure, and that the current is increasing so that the direction of the rate of change of current is in the same direction as the current itself. Starting at the point a and applying Kirchhoff's laws by moving a probe charge around the circuit in the direction of the current, we find

$$-L\frac{di}{dt} - iR + V = 0, \tag{32-10}$$

where V is the emf of the battery, and $-L\ (di/dt)$ is the back emf in the inductor.

Equation (32-10) is a differential equation whose solution is given by

$$i = \frac{V}{R}(1 - e^{-\frac{R}{L}t}). \tag{32-11}$$

We may verify this solution by differentiation of Equation (32-11) and substitution into Equation (32-10).

$$\frac{di}{dt} = \frac{V}{R}\frac{R}{L}e^{-\frac{R}{L}t}. \tag{32-12}$$

Substituting for i and for di/dt from Equations (32-11) and (32-12) into Equation (32-10), we find

$$-L\frac{V}{L}e^{(-R/L)t} - V + Ve^{(-R/L)t} + V = 0,$$

establishing the correctness of the solution.

The current in the circuit is zero at the instant the switch is closed, and gradually increases to a maximum value which is determined only by the magnitude of the resistance and the emf of the cell, as shown in Figure 32-13. Initially, the rate of change of current is very large, and the emf induced in the inductor limits the flow of current. At the instant the switch is closed, the emf of the inductor is equal and opposite to the emf of the

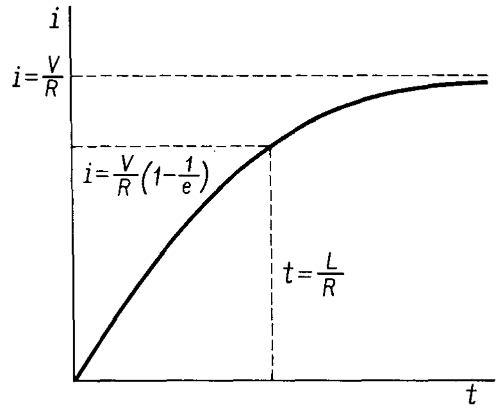


Fig. 32-13 Growth of current in a circuit containing inductance and resistance.

battery, so that the current is zero. When the current reaches a steady value and is no longer changing, there is no induced emf, and the current is determined by Ohm's law. At a time $t = L/R$, called the *time constant* of the circuit, the current has reached to within $1/e$ of its maximum value.

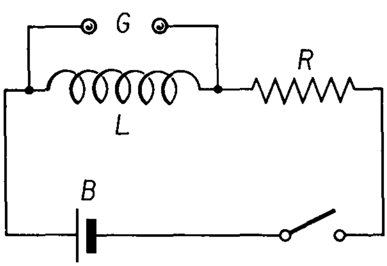


Fig. 32-14 A spark-gap G placed across the terminals of an inductor.

The curve describing the current as a function of time is called a *transient*, for it describes the current shortly after the switch is closed rather than the steady-state current that is established after a long interval of time.

In the process of establishing a current in the circuit, a magnetic field is established in the inductor. If the switch is suddenly opened after the current has reached a steady value, an emf will be induced in the inductor

whose value will depend upon the rate of change of current. The more quickly the current is interrupted, the greater will be the induced emf. If a spark gap is placed across the terminals of the inductor, as shown in Figure 32-14, a spark may pass between the terminals of the spark gap when the switch is opened, because of the large induced emf in the inductor. The energy of the magnetic field is then dissipated into the heat, sound, and radiant energy generated at the spark gap. The emf generated upon opening a circuit containing inductance is often called a switching transient, and is responsible for the large arcs which are often observed when electrical switches are opened.

32-8 Energy Stored in an Inductor

Let us calculate the energy of an inductor when there is a steady-state current I in it. During the transient interval when the current is changing from zero to the maximum value I, the emf of the inductor is given by

$$\mathcal{E} = -L\frac{di}{dt}. \tag{32-6}$$

An amount of power $\mathcal{P}$ is expended by an external source of electrical energy to establish this current. The applied potential difference is opposite to the direction of the induced emf so that

$$\mathcal{P} = -\mathcal{E}i = Li\frac{di}{dt}.$$

The work $d\mathcal{W}$ done by an outside agency in driving current through the inductor against the induced emf in a time dt is

$$d\mathcal{W} = \mathcal{P}\,dt = Li\frac{di}{dt}\,dt,$$

or

$$d\mathcal{W} = Li\,di.$$

Integrating between the limits of $i = 0$ and $i = I$, the final current through the inductor, we have,

$$\mathcal{W} = \int_0^I Li\,di,$$

yielding

$$\mathcal{W} = \tfrac{1}{2}LI^2 \tag{32-13}$$

for the energy of an inductor. When L is in henrys and I is in amperes, $\mathcal{W}$ is in joules.

When the current in a circuit builds up from zero to a value I, energy is supplied to the magnetic field, its value being $\frac{1}{2}LI^2$. As long as the

current remains constant, no additional energy is supplied to the magnetic field. In the circuit of Figure 32-12, all of the energy supplied during the steady state is transformed into heat. When the current decreases from I to zero, the magnetic field also decreases to zero; the energy that was stored in the magnetic field is returned to the circuit.

We can use the equation for the energy of an inductor to determine the energy per unit volume $\mathscr{W}_V$ in a magnetic field by considering an inductor in the form of a toroid. Its magnetic field is confined entirely to the volume within the toroid. The inductance of a toroid is given by

$$L = \frac{\mu_0 N^2 A}{s}, \tag{32-8}$$

hence the energy stored in the magnetic field of the toroid is, from Equation (32-13),

$$\mathscr{W} = \frac{1}{2}\frac{\mu_0 N^2 A}{s} I^2.$$

The magnetic field intensity within the toroid is given by

$$H = \frac{NI}{s}. \tag{30-5}$$

Substituting for N from Equation (30-5) into the above equation for the energy of the toroid, we have

$$\mathscr{W} = \tfrac{1}{2}\mu_0 H^2 A s.$$

The volume within the toroid is given by the product of its cross-sectional area A by its mean circumference s. Thus the energy per unit volume $\mathscr{W}_V$ of the magnetic field is given by

$$\mathscr{W}_V = \tfrac{1}{2}\mu_0 H^2 = \tfrac{1}{2}BH. \tag{32-14}$$

Notice that our procedure for finding the energy stored in the magnetic field has been very similar to the calculation by which we found the energy per unit volume stored in the electric field. In the case of the magnetic field, we utilized the energy in the field of a toroid, while in the electric field we utilized the energy in the field of a capacitor. Recalling that result from Equation (25-5), we may write the energy per unit volume $\mathscr{W}_V$ in the electromagnetic field as

$$\mathscr{W}_V = \tfrac{1}{2}(\epsilon_0 E^2 + \mu_0 H^2). \tag{32-15a}$$

This formula may be rewritten as

$$\mathscr{W}_V = \tfrac{1}{2}(DE + BH). \tag{32-15b}$$

In the mks system of units, the energy density $\mathscr{W}_V$ in the electromagnetic field is in joules per cubic meter when the electric field is expressed in volts per meter and the magnetic field intensity is expressed in amperes per meter.

TABLE 32-1 PRINCIPAL EQUATIONS IN MKS AND GAUSSIAN UNITS

Equation	MKS	Gaussian	
(32-2)	$\Delta\Phi = B\,\Delta A$	Same form as mks	Flux change
(32-3)	$\mathcal{E} = -\dfrac{d\Phi}{dt}$	$\mathcal{E} = -\dfrac{1}{c}\dfrac{d\Phi}{dt}$	Electromagnetic induction
(32-4)	$\mathcal{E} = -N\dfrac{d\Phi}{dt}$	$\mathcal{E} = -\dfrac{N}{c}\dfrac{d\Phi}{dt}$	N-turn coil
(32-6)	$\mathcal{E} = -L\dfrac{di}{dt}$	Same form as mks	Inductance
(32-8)	$L = \dfrac{\mu_0 N^2 A}{s}$	$L = \dfrac{4\pi N^2 A}{sc^2}$	Toroid or long solenoid
(32-13)	$\mathscr{W} = \frac{1}{2}LI^2$	Same form as mks	Energy
(32-14)	$\mathscr{W}_V = \frac{1}{2}\mu_0 H^2 = \frac{1}{2}BH$	$\mathscr{W}_V = \dfrac{H^2}{8\pi} = \dfrac{BH}{8\pi}$	Energy density in vacuum
(32-15b)	$\mathscr{W}_V = \frac{1}{2}(DE + BH)$	$\mathscr{W}_V = \dfrac{1}{8\pi}(DE + BH)$	Energy density

TABLE 32-2 CONVERSION FACTORS RELATING MKS AND GAUSSIAN UNITS

Quantity	Symbol	MKS Unit	Gaussian Unit	
Flux	Φ	1 weber	$= 10^8$ maxwells	(emu)
Inductance	L	1 henry	$= \dfrac{1}{9 \times 10^{11}}$ stathenry	(esu)
Pole	p	1 weber	$= \dfrac{10^8}{4\pi}$ unit pole	(emu)
Magnetic intensity	H	$1\,\dfrac{\text{nt}}{\text{weber}} = 1\,\dfrac{\text{amp}}{\text{m}}$	$= 4\pi \times 10^{-3}$ oersted	(emu)
Magnetic induction	B	1 weber/m^2	$= 10^4$ gausses	(emu)
Electric intensity	E	$1\,\dfrac{\text{nt}}{\text{coul}} = 1\,\dfrac{\text{volt}}{\text{m}}$	$= \dfrac{10^{-4}}{3}\dfrac{\text{statvolt}}{\text{cm}} = \dfrac{10^{-4}}{3}\dfrac{\text{dyne}}{\text{stcoul}}$	(esu)
Electric displacement	D	1 coul/m^2	$= 3 \times 10^5$ statcoul/cm^2	(esu)

Problems

32-1. The magnetic flux in a coil having 40 turns changes steadily from zero to 20,000 maxwells in 2 sec. Find the induced emf in the coil.

32-2. The magnetic flux through a coil of 125 turns is changed at a constant rate from zero to 0.40 weber in 2.5 sec. Determine the emf induced in this coil.

32-3. A circular coil of 50 turns and radius 15 cm lies in the x-y plane. The magnetic induction is changed in 0.001 sec at a constant rate from zero to a value whose x component is 0.3 weber/m^2, whose y component is 0.4 weber/m^2, and whose z component is 0.5 weber/m^2. Find the emf induced in the coil.

32-4. A wire 50 cm long is at rest along the x axis. A large magnet generating a uniform field directed along the $+y$ direction of 0.2 weber/m^2 is moved in the $+z$ direction with a speed of 30 m/sec. Find the magnitude and the direction of the emf induced in the wire.

32-5. A wire 75 cm long is moved in the y direction at a speed of 25 m/sec, so that the wire is always parallel to the x axis. The magnetic field has components $B_x = 0.2$ weber/m^2, $B_y = -0.3$ weber/m^2, and $B_z = 0.4$ weber/m^2. Find the emf induced in the wire.

32-6. A short solenoid, connected to a galvanometer, stands on one end upon a table. The north pole of a long bar magnet is brought down from above into the solenoid. Apply Lenz's law to find if the direction of the current induced in the solenoid is clockwise or counterclockwise, as viewed from above.

32-7. A rectangular coil of wire having 10 turns with dimensions of 20 cm by 30 cm is rotating at constant speed of 600 rpm in a magnetic field in which the magnetic induction is 600 gausses. The axis of rotation is perpendicular to the field. Find the maximum value of the emf produced.

32-8. Part of a closed circuit consists of a straight wire 1.5 m long moving at a speed of 2 m/sec perpendicular to a magnetic field of 10,000 gausses. (a) What is the emf induced in the circuit? (b) What is the force on the wire when the induced current is 5 amp?

32-9. A coil of 300 concentrated turns and an area of 800 cm^2 is lying flat on a horizontal table. When the coil is turned over through 180° in 0.10 sec, the average induced emf is 0.024 volt. What is the vertical component of the magnetic intensity of the earth's magnetic field?

32-10. A rectangular coil 12 cm by 25 cm and containing 15 turns is rotating at a constant speed of 1,800 rpm in a magnetic field in which the magnetic induction is 0.15 weber/m^2. The axis of rotation is perpendicular to the field. (a) Determine the maximum emf induced in this coil. (b) If the zero of time is taken at the point where the coil is parallel to the magnetic induction, find the emf in the coil when it has rotated by 53°.

32-11. When the current in a coil is changed at a constant rate of 5 amp/sec, the emf induced in the coil is 0.25 volt. Determine the self-inductance of the coil.

32-12. Derive a formula for the self-inductance of a long solenoid. Assume that the field is uniform everywhere within the solenoid.

32-13. What is the self-inductance of a solenoid 50 cm long and 5 cm in diameter, wound with 400 turns of wire?

32-14. What is the self-inductance of a toroid of mean circumference 25 cm, wound with 500 turns of wire, if the cross-sectional area of each turn is 2 cm^2?

32-15. An inductor has an inductance of 0.01 henry and an internal resistance of 5 ohms. The inductor is connected to the terminals of a battery having an emf of 12 volts. What will be the current in the circuit (a) in 0.001 sec? (b) In 0.01 sec? (c) In 0.1 sec? (d) Determine the time constant of this circuit.

32-16. An inductor of inductance 0.1 henry is connected in series with a 50-ohm resistor. This series combination is connected across the terminals of a 100-volt battery. What will be the energy stored in the magnetic field of the inductor when the current reaches a steady value?

32-17. Solve Equation (32-10) by separating the variables so that it becomes

$$\frac{di}{i - \frac{V}{R}} = -\frac{R}{L}\,dt,$$

and integrating to obtain Equation (32-11). Evaluate the constant of integration, letting $i = 0$ when $t = 0$.

32-18. Suppose that a switching arrangement is used in the circuit of Figure 32-12 so that the battery is removed from the circuit and a connecting wire is substituted in its place. (a) Show that Equation (32-10) becomes

$$L\frac{di}{dt} + iR = 0.$$

(b) Solve this equation for the current as a function of the time. Evaluate the constant of integration letting $i = I$ when $t = 0$. (c) Plot a graph of this equation and compare it with the graph of Figure 32-13.

32-19. Referring to the toroid of Problem 32-14, calculate (a) the energy in the magnetic field when the current is 10 amp and (b) the energy per unit volume of this field.

33

Magnetic Properties of Matter

33-1 Introduction

Matter is composed of atoms consisting of positively charged nuclei and negative electrons. These electrons occur in shells, and the periodic nature of chemical properties of atoms as the atomic weight increases is a reflection of the fact that the chemical behavior of an atom depends largely upon the number of electrons in the outermost shell. In some parts of the periodic table (see Table 5 of Appendix A), electrons occupy places in an outer shell before an inner shell is completely filled; it is then observed that a number of different elements have very similar chemical properties. The same number of electrons lies in the same outermost shell of these different atoms, but the inner shells contain different numbers of electrons. The chemical similarity of the *rare-earth* elements, atomic numbers 57 to 71, may be explained on this basis. A similar state of affairs exists in the group of elements of atomic number 26 (iron), 27 (cobalt), and 28 (nickel), all of which contain 2 electrons in their outermost shell but have 6, 7, and 8 electrons respectively, in their next inner shell, where 10 electrons are required to fill that shell. As a consequence the electrons in the unfilled shell of one atom may exert an important influence on the electrons of the unfilled shell of a properly spaced adjacent atom in a crystal. The *ferromagnetism* of iron, nickel, and cobalt is explicable in terms of the electronic configuration of these partially filled shells.

The magnetic properties of matter arise from two sources. An electron in orbital motion about the nucleus constitutes a small circulating current, which generates a magnetic field. The electron moving in its orbit has an *angular momentum* about the nucleus. In addition to the magnetism due to its orbital motion, an electron has an intrinsic magnetic moment and an intrinsic angular momentum, owing to its *spin*.

In the absence of a magnetic field, the orbital magnetic moments and the intrinsic magnetic moments of different electrons are randomly oriented within matter. There may be relatively large local magnetic fields in small

regions, but when these magnetic fields are averaged over a volume of even a cubic millimeter, the average field is zero, so that, macroscopically, no magnetism is displayed. Most materials are only very slightly magnetic in the presence of external magnetic fields. They are said to be either diamagnetic or paramagnetic. *A diamagnetic material is one for which* κ_m *is less than* 1. The magnetic effects induced in the material are opposed to the external field. We would expect materials to be diamagnetic if the orbital electronic effects predominated, for, in accordance with Lenz's law, the magnetic effects induced in a circuit must be in such a direction as to oppose the change in magnetic field in the substance. *A paramagnetic material is one for which* κ_m *is greater than* 1. When placed in a nonuniform field, a diamagnetic substance will experience a force directed from the stronger to the weaker part of the field. A paramagnetic substance will experience a force in the opposite direction. In general, in all materials except those called *ferromagnetic* (that is, those which behave like iron), the magnetic effects are quite small, and these materials may be treated as though their relative permeability κ_m is 1, to an accuracy of about 0.1 per cent. The magnetic behavior of most substances is not substantially different from vacuum. A *ferromagnetic* substance is attracted into a magnetic field with a *large* force. The relative permeability κ_m of a ferromagnetic substance may be as large as 10^4 or 10^5. Ferromagnetic substances are therefore special cases of the general class of paramagnetic substances. The subject of ferromagnetism is of great importance in electrical engineering, and is as complex as it is important. The properties of ferromagnetic substances form the basis of the practical design of motors, generators, transformers, magnetic amplifiers, tape recorders, loud-speakers, permanent magnets, and a host of other devices. We shall attempt to develop only some of the basic ideas of ferromagnetism, and to solve problems of the simplest type in which symmetry considerations enable us to see the important principles most clearly without the confusion of detail which cannot be neglected in practical engineering design.

33-2 Permeability

The basis of our study of the electrical properties of matter was the observed change in the capacitance of a parallel-plate capacitor when the space between the plates was filled with a dielectric. The dielectric constant was defined in this manner in Section 25-5 and was interpreted in subsequent sections in terms of the induced electric polarization. Lacking magnetic conductors, we must find some other way to define the magnetic properties of matter. One way is to compare the inductance of a long solenoid or toroid in vacuum with the inductance of that same solenoid or toroid when the space within the coil is filled with a medium. If L_0 is the inductance

of a toroid in vacuum, and L is its inductance when filled with a particular material, then

$$\frac{L}{L_0} = \kappa_m, \tag{33-1}$$

where κ_m is the relative permeability of the medium. The inductance of a toroid in vacuum is given by the equation

$$L_0 = \frac{\mu_0 N^2 A}{s}. \tag{32-8}$$

Remembering that

$$\mu = \kappa_m \mu_0, \tag{29-16}$$

we may write

$$L = \frac{\mu N^2 A}{s} \tag{33-2}$$

for the inductance of a toroid filled with a magnetic medium.

Upon measurement of the inductance of a toroid filled with a ferromagnetic core, it is found that the inductance of the toroid is not constant but depends upon the magnitude of the current. That is, the permeability of the medium in the core is not constant but depends upon the magnetic field intensity within the toroid. Hence it is advantageous to study the variation of the permeability with H, the magnetic field intensity within the toroid.

In order to understand properly the effect of the induced magnetization of the medium on the magnetic field within the toroid, let us first consider the behavior of a long rod of ferromagnetic material placed inside a long solenoid, as shown in Figure 33-1. When current is passed through

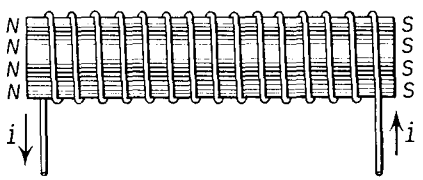


Fig. 33-1 A solenoid containing a ferromagnetic core.

the solenoid, there is a tendency for the intrinsic electronic magnetic moments of the electrons in the unfilled inner shells to align themselves with the direction of the field, like a collection of compass needles. Just as in the case of electric polarization, we may imagine the magnetized rod to be replaced by a layer of north poles at one end of the rod and a layer of south poles at the other end of the rod. The magnetic field intensity within the rod may then be calculated as being due to two causes. First there is the applied field intensity due to the current in the solenoid, and, second, there is the magnetic field intensity due to the induced poles in the material,

at the ends of the rod. Let us suppose that the field is such as to induce a magnetic moment M per unit volume in the rod. The total magnetic moment in the rod is the product of the magnetization M by the volume of the rod. If the total strength of the induced pole at each end of the rod is represented by p, the magnitude of the induced pole strength may be found by representing the total magnetic moment in terms of the pole strength and in terms of the magnetization, and equating these two quantities. Thus we have

$$MAs = ps,$$

or

$$p = MA,$$

where A is the cross-sectional area of the rod and s is its length. We see that the induced pole strength does not depend upon the length of the rod but only on the magnetization and the area of the rod. If the rod is made very long, the induced poles contribute very little to the magnetic field intensity at the center of the solenoid, so that H at the center of the solenoid is the same, whether the solenoid is filled with magnetic material or is in vacuum. The same result is accomplished if the two ends of the solenoid are joined to form a toroid. In this case the rod has no free ends, so that there are no induced poles, and the value of H within the toroid is the same, regardless of whether the toroid is filled with matter or is in vacuum. For this reason experimental measurements are often made on a ringlike specimen, called a *Rowland ring*, after H. A. Rowland who first (1873) utilized toroids wound on iron rings to measure permeability.

While the magnetic field intensity H within the toroid is not changed when it is filled with a ferromagnetic material, the inductance, which depends upon the magnetic induction B, does change when the toroid is filled with matter. To study the changes in permeability with H we shall examine, through use of the Rowland ring, the way in which the magnetic induction within a substance changes with the imposed magnetic field intensity H.

33-3 Magnetic Measurements with a Rowland Ring

To measure the magnetic properties of a material by means of the Rowland-ring method, a specimen is machined in the form of a ring and is wound with a toroidal coil, called the primary coil, as shown in Figure 33-2. The primary coil is connected in series with a battery, an ammeter, and a rheostat; the latter is used to change the current in the primary coil. A small secondary winding is wound around the toroid and connected to a *ballistic galvanometer*. Such a galvanometer is constructed with a coil which has a relatively large moment of inertia, and a long period of vibration.

When a short burst of current passes through the galvanometer, there is very little rotation of the galvanometer coil until after the current has ceased. The angular impulse delivered to the galvanometer depends upon the torque, which is proportional to the current, and upon the time interval during which the torque is applied. The ballistic galvanometer thus re-

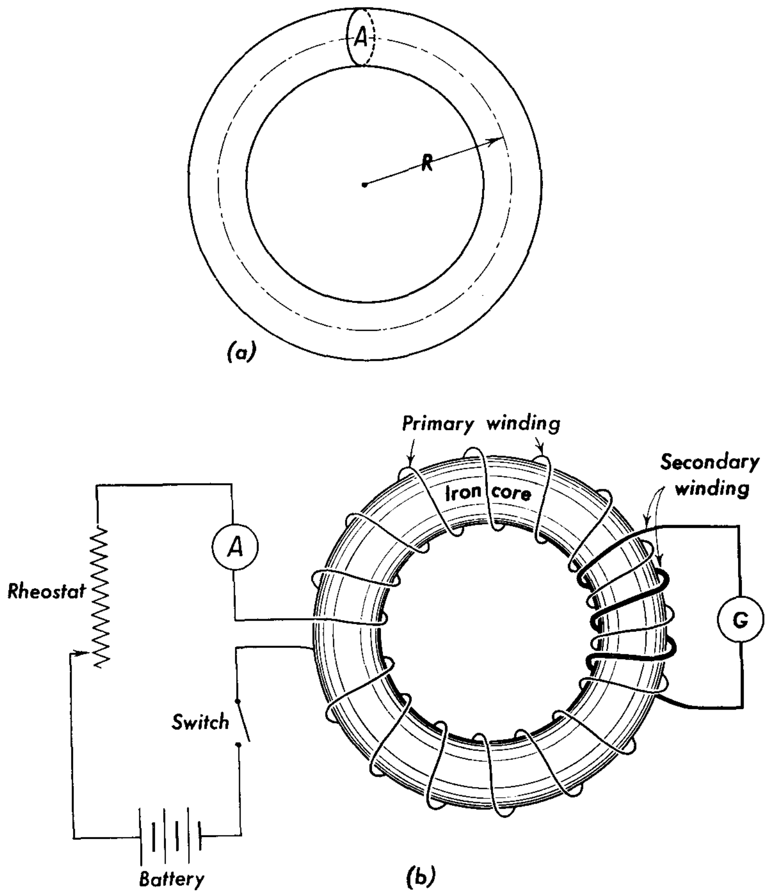


Fig. 33-2 Rowland ring.

ceives an angular impulse which depends upon the product of the current times the time, or the charge passing through it; its deflection is proportional to the total charge passing through in the time interval Δt, rather than the instantaneous value of the current.

Suppose we start with an unmagnetized piece of iron and vary the current through the toroid in a series of small steps. At each step the

magnetic field intensity H may be calculated from Equation (30-5), for, as we have seen, the magnetic intensity H within the toroid is unchanged by the presence of the iron core. Each time H is changed, there is a corresponding change in B and, consequently, a change in the flux through the secondary coil. From Faraday's law of induction, an emf is thus induced in the secondary coil.

Let us suppose that the change in B due to a change in H in a time interval Δt is represented by ΔB. If the cross-sectional area of the toroid is A, the change in flux $\Delta\Phi$ through the secondary coil is given by

$$\Delta\Phi = A\,\Delta B.$$

The induced emf in the secondary coil of N turns is

$$\mathcal{E} = -N\frac{\Delta\Phi}{\Delta t} = -NA\frac{\Delta B}{\Delta t}\cdot$$

If the total resistance of the secondary coil and galvanometer is R, the magnitude of the current induced in the secondary circuit is given by Ohm's law as

$$i = \frac{\mathcal{E}}{R} = \frac{NA}{R}\frac{\Delta B}{\Delta t}\cdot$$

Multiplying through by the time interval Δt, we find that the charge Δq flowing through the galvanometer when the current in the primary winding of the toroid is changed is given by

$$\Delta q = -\frac{NA}{R}\Delta B. \tag{33-3}$$

Thus each time the current through the primary coil is changed, a measurement of the charge flowing through the ballistic galvanometer enables us to determine the change in B, ΔB, from the constants of the measuring circuit.

If we start with $H = 0$ and with unmagnetized iron in which the magnetization M is zero, then since

$$B = \mu_0 H + M, \tag{29-14a}$$

the magnetic induction within the coil is zero initially. By measuring the successive changes in B as H is varied, we get a series of values of B, as shown in Figure 33-3. The magnetic induction B increases slowly at first, with increasing H, then increases more rapidly until the flat portion of the curve is reached at A. Along the flat portion of the curve the iron is said to be *saturated.* Very little further magnetization in the iron is observed when H is increased, and the increase in B is due to the increase in H itself. The curve appears flat because of the difference in the scales used for B and

H. The relative permeability κ_m may be obtained from such a curve and the relationship

$$B = \kappa_m \mu_0 H,$$

when the scales of the axes are represented in mks units. The shape of the curve and the permeability itself depend upon the types of iron used. In some laboratory samples of iron, a relative permeability as high as 10^6 has been achieved.

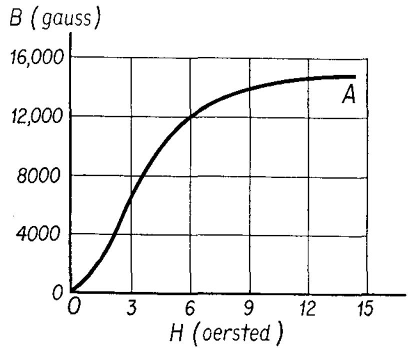
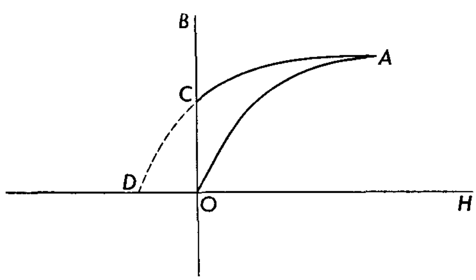


Fig. 33-3 Magnetization curve of soft iron.

33-4 Hysteresis

Let us take a piece of unmagnetized iron in the form of a ring of circular cross section, wind a toroid around it, then wind a secondary coil around it, and connect it to a battery and meters, as shown in Figure 33-2. If we

Fig. 33-4 Hysteresis.



send current through the coil and magnetize the iron until the magnetic induction has reached its saturation value, as shown in Figure 33-4, the variation of B with H follows the magnetization curve OA. If we now decrease the current in small steps, it is found that the demagnetization curve does not follow the original curve OA but, instead, follows the curve AC. The value of the magnetic induction remaining when the current in the toroid is reduced to zero is given by OC. The iron is now permanently magnetized, and the value of the induction OC is called the *retentivity*. In

permanent magnets a high retentivity is required, while in other applications a low value of the retentivity is required.

To reduce the magnetic induction to zero, that is, to *demagnetize* the iron, it is necessary to reverse the current in the magnetizing coil in order to reverse the direction of H. When H has reached the negative value

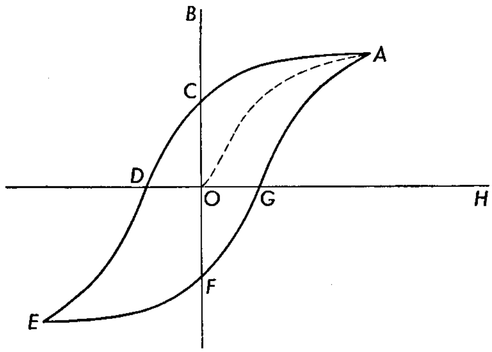


Fig. 33-5 Hysteresis loop in cycle of magnetization of iron.

given by OD, the induction within the iron is reduced to zero. The value of H needed to reduce B to zero is called the *coercive force*. The fact that B lags behind its previous value while H returns to its former value is called *hysteresis*.

All ferromagnetic materials exhibit the phenomenon of hysteresis. One of the effects of hysteresis is that the value of B for any given value of H is not always the same but rather depends upon the magnetic history of the material. If an alternating current is sent through the toroid wound on a ferromagnetic core, the B–H curve will be similar to Figure 33-5 for each complete cycle of current. From Equation (32-14) the energy per unit volume in the magnetic field depends upon the product BH. While the iron is being magnetized, energy is being stored in the magnetic field; when the iron is demagnetized, some of that energy is recovered as electrical energy. In traversing one cycle of the hysteresis loop, the energy per unit volume of iron dissipated as heat is the area inside the hysteresis loop. In the design of a-c machinery, it is important that a type of iron be used in which the area enclosed within the hysteresis loop is small, so that a minimum of energy is lost through this mechanism.

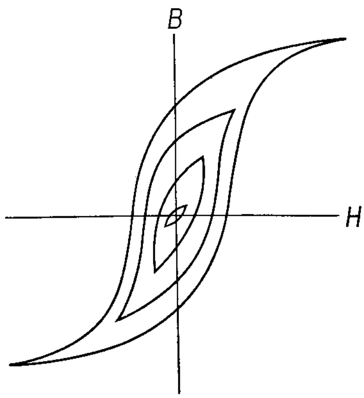


Fig. 33-6 Hysteresis curves obtained by using alternating current of successively smaller values in the primary circuit.

A simple way of demagnetizing a substance is to place it inside a coil and pass alternating current through the coil. The amplitude of the alternating current is decreased slowly, so that the hysteresis loop gets smaller and smaller in successive cycles, as shown in Figure 33-6, until finally, when the current in the coil is zero, the values of B and H are zero.

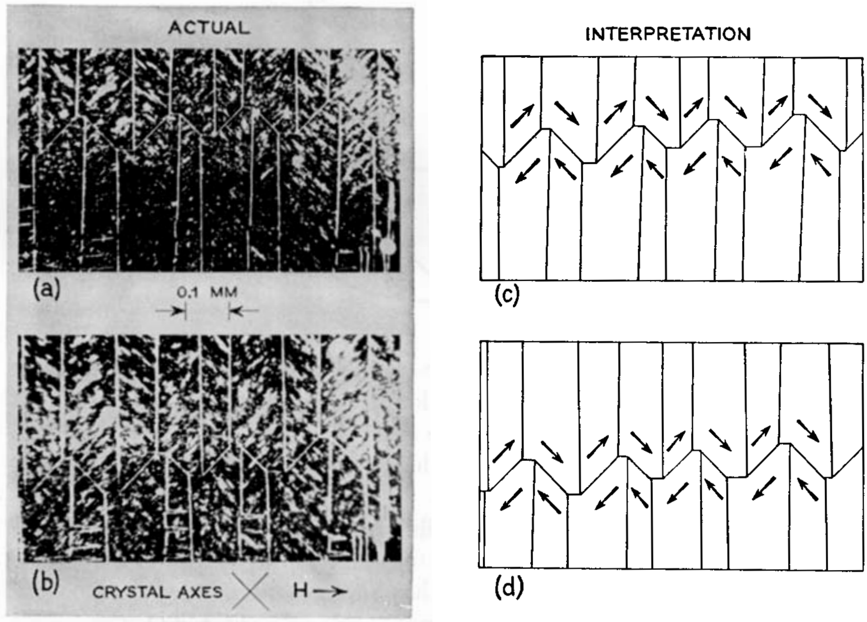


Fig. 33-7 (a) Photograph of magnetic domains of a magnet; (c) diagram of (a) showing the directions of magnetization at the boundaries of the domains. (b) Photograph showing the movement of the domain walls when an external magnetic field H is applied; (d) diagram of (b) showing that the domains that are magnetized favorably with respect to H grow at the expense of the other domains. (Photograph by H. J. Williams and R. M. Bozorth; reproduced by permission of *Electrical Engineering*, **68**, 1949, 471.)

The hysteresis loop is explained by means of the *domain theory of magnetization*, first stated by P. Weiss in 1907. According to this theory a ferromagnetic material is composed of many small regions, or domains, each magnetized to saturation, and about 5×10^{-3} cm in width. When a suspension of colloidal magnetite is applied to a highly polished piece of iron while it is being magnetized, the colloidal particles collect near the domain boundaries and may be observed with a microscope, as shown in Figure 33-7.

In the unmagnetized state the directions in which the domains are magnetized are distributed at random or in some other way such that the resultant magnetization of the specimen is zero. Changes in the total

magnetization of the specimen are produced by changes in the direction of magnetization of the domains or by motion of the boundaries of the walls of the domains, as shown in Figure 33-8. At weak fields the magnetization proceeds by boundary displacement. When the material is very

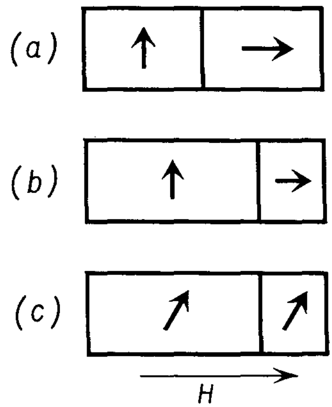


Fig. 33-8 (a) Original magnetization of two adjacent domains. (b) Change in magnetization as a result of the growth of one domain at the expense of the other. (c) Change in magnetization resulting from a change in the direction of magnetization within the domains.

pure and homogeneous, the boundaries of the domains are easily changed, and so the coercive force is small and the permeability is high. When impurities are present, or the material is in a state of internal stress, the material is inhomogeneous, so that the boundaries of the domains are less easily displaced. The initial permeability is lowered, and the coercive force is increased. By sintering a magnetic material of very fine powders, or by precipitating an impurity or another metallurgical phase, the displacement of domain boundaries is hindered, and good permanent magnets are made which require a large demagnetizing field to alter their magnetization. At high magnetic fields the magnetization is accomplished by rotation of the direction of magnetization within the domains. When the direction of magnetization in all the domains is parallel to the applied magnetic field, the material is saturated.

The motion of the boundary walls of the magnetic domains does not occur smoothly, particularly along the steep portion of the magnetization curve. This irregular motion produces sudden changes in the magnetization of the specimen as the magnetizing force H is changed. If the galvanometer in the secondary circuit of Figure 33-2 is replaced by an amplifier and loud-speaker, a succession of clicks will be heard when the magnetization is changing. This effect, known as the *Barkhausen effect,* was discovered by H. Barkhausen in 1919.

From Figure 33-6 we see that the magnetization which remains in a specimen when the applied magnetic field intensity is reduced to zero depends upon the magnitude of the applied field. Through the residual magnetization the specimen remembers the amplitude of the applied field.

The magnetic memory is the basis of magnetic tape recorders and of many present-day electronic computers.

When a specimen is magnetized, it behaves as though there were magnetic poles at the boundaries of the specimen. If there is a crack in the specimen, poles of opposite polarity appear on the adjacent faces of the crack. This is the basis of magnetic inspection of ferrous machine parts, widely used in aircraft maintenance and production. A part is magnetized and is then flushed with light oil which carries a suspension of magnetic powder. The powder tends to cling to the crack, enabling an inspector to identify a defective part. This is also the basis of the techniques used in making photographs of domains such as Figure 33-7.

33-5 Other Magnetic Effects

The intrinsic magnetic moment of the electron, associated with electron spin, rather than the orbital motion of the electron, is responsible for ferromagnetism. Electrons in completed shells are arranged so that he total magnetic moment of the shell is zero. In iron, cobalt, and nickel the electrons in the unfilled inner shell are responsible for ferromagnetism.

In ferromagnetic materials the inner electrons of neighboring atoms are bound by forces called *exchange forces* which depend upon the orientation of the spins of these inner electrons. In ferromagnetic materials the electrons of adjacent atoms are held parallel, whereas in other materials these forces generally tend to align the spins so that they are antiparallel, that is, in opposite directions, so that adjacent atoms tend to neutralize each other's magnetic effects. When iron is heated to such a temperature that the thermal energy of the electrons exceeds the energy associated with the exchange force, the electrons are no longer able to maintain their parallel orientation, so that the iron is no longer ferromagnetic. The temperature at which this magnetic transition occurs is called the *Curie temperature*, which, in iron, is 760°C.

One would expect that ferromagnetism would also depend upon the separation and arrangement of iron atoms in the crystal lattice. If the atoms were sufficiently far apart, or were not laid out according to a proper pattern, it would be impossible for adjacent atoms to influence each other; indeed, this is the case. In single crystals the magnetization curve is different in different crystal directions. Furthermore, while iron is ferromagnetic, a type of stainless steel containing 18 per cent chromium and 8 per cent nickel is nonmagnetic. Similarly for the two common oxides of iron, one, called magnetite, Fe_3O_4, is magnetic, while the other, called hematite, Fe_2O_3, is nonmagnetic. It is also possible to make alloys which display ferromagnetic properties out of elements, such as copper, manganese, and aluminum, which themselves are not ferromagnetic. These are called *Heusler's alloys*. The way in which the atoms are arranged in the

alloy or in a crystal is of fundamental importance in determining its ferromagnetic properties.

Two extremely interesting effects have been observed, using macroscopic specimens, which confirm the theory that the spin of the electron and its inherent magnetic moment are responsible for ferromagnetism. Let us suppose that a rod of ferromagnetic material is suspended by a fiber, inside a solenoid, as shown in Figure 33-9. When current is passed through the solenoid, the orientation of the magnetic moment of a large number of electrons is changed. This implies that the angular-momentum vector has been changed for each of these electrons. According to the principle of conservation of angular momentum, a system which has experienced no external torques must retain a constant value of its angular momentum, and so the rod itself must rotate in such a direction that the total angular momentum of the system, made up of the crystal lattice and the electrons, remains equal to zero. This is called the *Einstein-de Haas effect,* in which a specimen is observed to rotate when it is being magnetized.

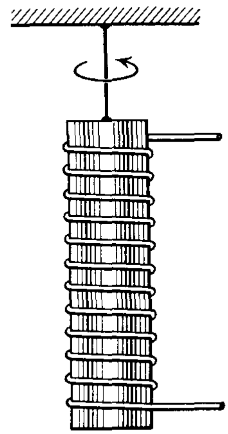


Fig. 33-9 Demonstration of the Einstein-de Haas effect. Magnetic specimen will rotate when it is being magnetized.

The inverse of this effect is called the *Barnett effect.* In the Barnett effect a specimen of ferromagnetic material is mechanically rotated, and the specimen may be observed to become magnetized.

Both of the above phenomena are classed as *gyromagnetic effects.*

33-6 Magnetic Circuits

Lines of magnetic induction always form closed loops. If we think of a tube of magnetic induction as a tube bounded by lines of induction, the lines of induction may never cross each other, and the total number of lines of induction contained within such a tube must be constant. *Thus the magnetic flux within a tube of induction is constant.* This is the first principle used in the calculation of a *magnetic circuit,* which may be considered as a *closed path of magnetic material.*

Let us consider the work done in carrying a magnetic pole around a wire carrying current. The magnetic field intensity at a distance a from a long straight wire is given by

$$H = \frac{I}{2\pi a}\cdot \qquad \textbf{(30-1a)}$$

Suppose that a pole of strength p is carried around the circle of radius a, concentric with the wire carrying current into the plane of the paper in Figure 33-10, opposite to the direction of the magnetic field. The pole

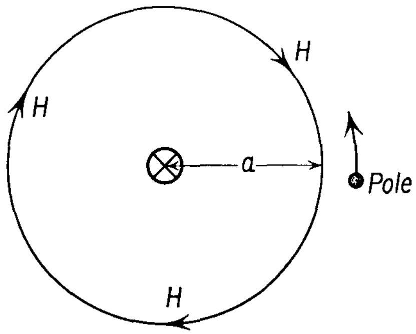


Fig. 33-10 Work done in carrying a magnetic pole of strength p around a wire carrying current. The direction of the current is into the paper.

is carried a distance $2\pi a$ against the force exerted by the magnetic field, so that the work done $\mathscr{W}$ is given by

$$\mathscr{W} = \int pH\,ds = Ip;$$

therefore

$$\frac{\mathscr{W}}{p} = \int H\,ds = I. \tag{33-4}$$

The work per unit pole in carrying a pole around a wire carrying current does not depend upon the radius of the circle but only upon the current in the wire. Any arbitrary path followed in carrying the pole around the wire may be approximated by a combination of radial paths, in which no work is done, and circular paths, in which the work done does not depend upon the radius but only upon the fraction of the circle traversed, as shown in Figure 33-11. By analogy with electromotive force, the work per unit pole done in carrying a north pole around a closed path is called the *magnetomotive force*, abbreviated mmf, and represented by $\mathscr{F}$.

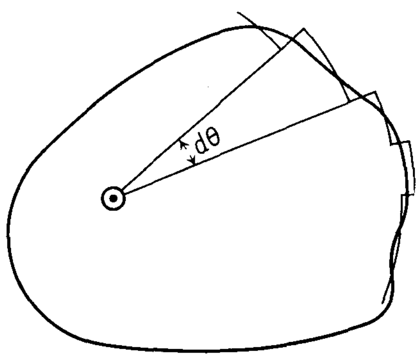


Fig. 33-11 Approximation of any path around a current by radial and circular displacements.

If the closed path encircles a number of wires carrying current, the work done in carrying a unit pole about each wire is given by Equation

(33-4) and, in, mks units, the total work per unit pole done is equal to the sum of the currents. This is the second fundamental principle for the design of magnetic circuits. The first principle relates to the magnetic induction B, while the second is concerned with the magnetic intensity H. In order to relate these, we must make use of the concept of permeability, imposing a third condition upon the magnetic circuit. These three conditions may be expressed in the following equations.

$$B = \mu H = \kappa_m \mu_0 H. \tag{33-5a}$$

$$\mathcal{J} = \int H\, ds = \sum I. \tag{33-5b}$$

$$\Phi = BA = \text{const.} \tag{33-5c}$$

Let us consider the case of a magnetic circuit made up of a ring-shaped core around which a uniform toroid has been wound. We shall imagine that the core has been cut so that a small section of the core may be removed. We shall first compute the induction in the core when the core is intact with a current I in the winding. Next we shall calculate the induction in the core and gap when the cut section is removed. This is the usual case of an electromagnet with an air gap.

Case 1. A uniform toroid of N turns is wound on a core of length s and cross-sectional area A. Find the induction within the core when there is a steady current I in the toroid.

We may assume that all of the magnetic flux is confined to the volume of the toroid, so that the core itself may be considered as a tube of induction. The magnetic intensity is constant around the core. Applying Equation (33-5b) and integrating around the mean circumference s of the core, we find

$$\mathcal{J} = Hs = NI,$$

or

$$H = \frac{NI}{s},$$

a result we have previously stated as Equation (30-5). The magnetic induction within the core may be obtained from Equation (33-5a) as

$$B = \mu H = \frac{\mu NI}{s}\cdot$$

The permeability of iron varies greatly from specimen to specimen, depending upon heat treatment, purity, internal stress, and magnetic intensity. In Figure 33-12 two graphs of the relative permeability of Armco iron are shown displaying the great variation in relative permeability with H and with heat treatment. In order to determine B, it is necessary to have such a curve for the particular iron being used. Alter-

natively, if the iron was initially unmagnetized and the magnetization curve is available, as in Figure 33-3, H may be computed from the formula above and B read from the curve. Appropriate conversion factors must be applied to Figure 33-3 if mks units are used, as in the formulas developed in the text of this chapter. We recall that

$$B: \quad 1 \text{ weber/m}^2 = 10^4 \text{ gausses,}$$

and

$$H: \quad 1 \text{ amp/m} = 4\pi \times 10^{-3} \text{ oersted.}$$

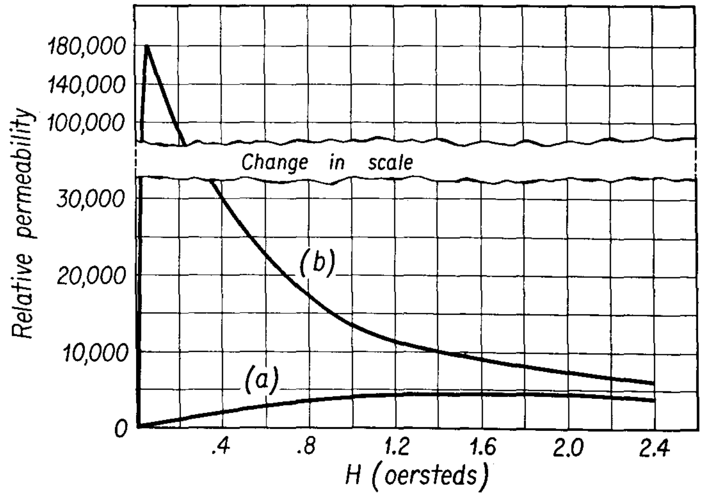


Fig. 33-12 Relative permeability of soft magnet iron (a) with standard annealing and (b) specially annealed. (Courtesy of Armco Steel Corporation.)

Case 2. A uniform toroid of N turns is wound on a core of mean circumference l_c having an air gap of length l_a. The core and gap have cross-sectional area A. Find the induction within the core and the gap when there is a current I in the toroid.

Once again we shall neglect any fringing field (although here this is a more serious approximation which, in practical problems, may give rise to considerable error) and shall consider that the boundary of the toroid is a tube of induction. We denote the magnetic intensity in the core by H_c and the magnetic intensity in the air gap by H_a. Similarly, the magnetic induction in the core is B_c, and the induction in the air gap is B_a. Applying Equation (33-5b), we find

$$\mathcal{J} = H_c l_c + H_a l_a = NI.$$

From Equation (33-5c) we find

$$B_c A_c = B_a A_a = \Phi.$$

If the permeability of the core is μ_c and that of the gap is μ_a, we may combine these two equations through Equation (33-5c) to find

$$\Phi = \frac{NI}{\dfrac{l_a}{\mu_a A_a} + \dfrac{l_c}{\mu_c A_c}}. \tag{33-6}$$

Equation (33-6) is often compared to Ohm's law for electric circuits. The flux Φ is thought to be analogous to the current, the quantity NI is the *magnetomotive force* $\mathcal{F}$, in analogy with the electromotive force, and the quantity $l/\mu A$ is called the *reluctance* $\mathcal{R}$ in analogy with the resistance. This is generally written as

$$\Phi = \frac{\mathcal{F}}{\mathcal{R}}, \tag{33-7}$$

where these quantities are given as

$$\mathcal{F} = NI \qquad \text{(ampere turns)},$$

$$\mathcal{R} = \sum \frac{l}{\mu A} \qquad \text{(ampere turns/weber)},$$

$$\Phi = BA \qquad \text{(webers)}.$$

In general, the permeability is not known as a function of the current in the coil; in such a case, if the desired induction in the gap is known, the induction in the core may first be computed from the formulas. The magnetic intensity in the core appropriate to this value of induction may then be obtained from the magnetization curve, giving the permeability, and the current required may then be obtained from the formulas.

Since the current literature of magnetism is largely written in terms of gausses and oersteds, we will calculate an example in Gaussian units, for which appropriate equations are listed in Table 33-1 and units are listed in Table 33-2.

Illustrative Example. The average circumference of a Rowland ring of soft iron is 50.1 cm. The ring is cut by an air gap 0.1 cm wide. The cross-sectional area of the ring (and gap) is 5 cm^2. The ring is wound with 2,000 turns of wire. Find the current required to produce an induction of 8,000 gausses in the gap.

Assuming that all of the flux in the core passes through the gap, and noting that the area of the gap is the same as the area of the core, the induction in the core is also 8,000 gausses. From Figure 33-3 a magnetic intensity of $H = 4$ oersteds produces this induction in soft iron. Thus, at an induction of 8,000 gausses, the relative permeability κ_m is given by Gaussian units

$$\kappa_m = \frac{B}{H}$$

$$= \frac{8{,}000 \text{ gausses}}{4 \text{ oersteds}} = 2{,}000.$$

The relative permeability of the air gap is 1. Thus the reluctance $\mathcal{R}$ of the magnetic circuit is

$$\mathcal{R} = \sum \frac{l}{\kappa_m a},$$

$$\mathcal{R} = \frac{50 \text{ cm}}{2{,}000 \times 5 \text{ cm}^2} + \frac{0.1 \text{ cm}}{1 \times 5 \text{ cm}^2},$$

$$\mathcal{R} = 0.025 \text{ gilbert.}$$

The magnetomotive force $\mathcal{F}$ is given by

$$\mathcal{F} = \frac{4\pi NI}{c},$$

$$\mathcal{F} = \frac{4\pi \times 2{,}000 \times I}{c}.$$

The required flux Φ is given by

$$\Phi = BA = 8{,}000 \text{ gausses} \times 5 \text{ cm}^2,$$

$$\Phi = 40{,}000 \text{ maxwells.}$$

We relate these quantities by the equation

$$\Phi = \frac{\mathcal{F}}{\mathcal{R}},$$

$$40{,}000 = \frac{4\pi \times 2{,}000 \times I}{c \times 0.025}.$$

Thus

$$I = \frac{c}{8\pi} \text{ statamperes.}$$

Remembering that 1 amp $= 3 \times 10^9$ statamperes, the current through the coil needed to produce the required induction in the gap is

$$I = \frac{10}{8\pi} \text{ amp}$$

$$= 0.399 \text{ amp.}$$

Note that the reluctance of the air gap is much higher than the reluctance of the iron path. In practical magnet problems it is necessary to take the fringing field into account. In general, something less than one half of the flux which passes through the iron also passes through the air gap.

Problems

33-1. Find the self-inductance of a toroid of 500 turns wound over an iron ring whose relative permeability is given by Figure 33-12, (a) when the current through the coil is 0.1 amp, (b) when the current through the coil is 0.2 amp, and

TABLE 33-1 PRINCIPAL EQUATIONS IN MKS AND GAUSSIAN UNITS

Equation	MKS	Gaussian	
(33-1)	$\frac{L}{L_0} = \kappa_m$	Same as mks	Relative permeability
(33-2)	$L = \frac{\mu N^2 A}{s}$	$L = \frac{\kappa_m 4\pi N^2 A}{sc^2}$	Long solenoid or toroid
(29-14)	$\mathbf{B} = \mu_0 \mathbf{H} + \mathbf{M}$	$\mathbf{B} = \mathbf{H} + 4\pi\mathbf{M}$	
(33-5a)	$B = \mu H = \kappa_m \mu_0 H$	$B = \kappa_m H$	
(33-5b)	$\mathcal{F} = \int H\,ds = \sum I$	$\mathcal{F} = \int H\,ds = \sum \frac{4\pi I}{c}$	Mmf
(33-5c)	$\Phi = BA$	Same as mks	Flux
(33-7)	$\Phi = \frac{\mathcal{F}}{\mathcal{R}}$	Same as mks	
	$\mathcal{F} = NI$	$\mathcal{F} = \frac{4\pi NI}{c}$	Mmf
	$\mathcal{R} = \sum \frac{l}{\mu a}$	$\mathcal{R} = \sum \frac{l}{\kappa_m a}$	Reluctance
(32-14)	$\mathscr{W}_V = \frac{1}{2}BH$	$\mathscr{W}_V = \frac{1}{8\pi} BH$	Magnetic energy density

TABLE 33-2 CONVERSION FACTORS RELATING MKS AND GAUSSIAN UNITS

Quantity	Symbol	MKS Unit	Gaussian Unit	
Magnetic intensity	H	$1\,\frac{\text{amp}}{\text{m}}$	$= 4\pi \times 10^{-3}$ oersted	(emu)
Flux density	B	$1\,\frac{\text{weber}}{\text{m}^2}$	$= 10^4$ gausses	(emu)
Flux	Φ	1 weber	$= 10^8$ maxwells	(emu)
Magnetomotive force	$\mathcal{F}$	1 amp turn	$= \frac{4\pi}{10}$ gilbert	(emu)
Reluctance	$\mathcal{R}$	$1\,\frac{\text{amp turn}}{\text{weber}}$	$= 4\pi \times 10^{-9}$ emu	(emu)
Magnetization	M	$1\,\frac{\text{weber}}{\text{m}^2}$	$= \frac{10^4}{4\pi}\,\frac{\text{unit poles}}{\text{cm}^2}$	(emu)

$$\mu_0 = 4\pi \times 10^{-7}\,\frac{\text{henry}}{\text{m}} = 4\pi \times 10^{-7}\,\frac{\text{weber}}{\text{nt m}^2} = 4\pi \times 10^{-7}\,\frac{\text{nt}}{\text{amp}^2} = 4\pi \times 10^{-7}\,\frac{\text{weber}}{\text{amp m}}$$

(c) when the current through the coil is 0.4 amp. The cross-sectional area of the ring is 1 cm^2, and its mean circumference is 125 cm.

33-2. A coil of wire contains 8 turns and has a resistance of 24 ohms. The coil is connected to a ballistic galvanometer which has a resistance of 60 ohms. If the magnetic flux through the coil is changed from 0 to 0.003 weber, (a) determine the charge which flows through the circuit. (b) If the sensitivity of the galvanometer is 25 μcoul/cm, determine the galvanometer deflection in centimeters.

33-3. A small coil of 200 turns of wire, having a circular area of 5 cm^2 and a resistance of 12 ohms, is used as an exploring coil to measure the magnetic field between the poles of a magnet. The coil is connected to a ballistic galvanometer whose resistance is 36 ohms and whose sensitivity is 0.18 μcoul/cm. (a) The coil is thrust into the magnetic field with the plane of the coil perpendicular to the lines of induction. The observed galvanometer deflection is 6.30 cm. Determine the magnetic induction. (b) While the coil is in this field, it is rotated through 180° about a diameter as axis. Determine the deflection of the galvanometer.

33-4. A solenoid 80 cm long has 500 turns and a cross-sectional area of 3.0 cm^2. A short secondary coil of 20 turns is wound around the middle of the first solenoid. The secondary coil, of resistance 1.3 ohms, is connected to a ballistic galvanometer of resistance 26.2 ohms. Determine the charge which flows through the secondary coil (a) when the switch is closed and the current through the primary rises to 2 amp and (b) when the current in the primary is increased from 2 amp to 3 amp.

33-5. A Rowland ring, wound with 1,000 turns of wire and having a mean circumference of 50 cm, carries a current of 4 amp. The relative permeability of the core is 800. (a) What is the magnetic intensity in the core? (b) What is the induction in the core? (c) What is the magnetization of the core?

33-6. Repeat Problem 33-5 in the case that the ring has been cut so that it has a gap 1 mm wide.

33-7. A ring of magnet iron, whose magnetization curve is shown in Figure 33-3, is wound with a toroidal coil of 250 turns. The mean circumference of the ring is 15 cm, and its cross-sectional area is 5 cm^2. Find the current in the coil required to produce an induction of 2,000 gausses.

33-8. A piece of magnet iron is used as the core of a solenoid. The magnetic field intensity inside the solenoid is 5 oersteds, and the induction within the core is 2,000 gausses. (a) What is the magnetic energy per unit volume within the iron? (b) What is the relative permeability of the iron at this induction? (c) What is the magnetization of the iron?

33-9. A long, straight, hollow tubular conductor of radius r carries a current I uniformly distributed around the conductor. Find the magnetic field intensity (a) at a point P_1 inside the tube at a distance a less than r, and (b) at a point P_2 outside at a distance b greater than r. [HINT: Carry a unit pole around a circular path concentric with the tube and apply a symmetry argument.]

33-10. Repeat the calculation in the illustrative example of Section 33-6 in mks units.

34

Alternating Currents

34-1 Resistance in A-C Circuits

If the current in a resistor varies sinusoidally, as shown in Figure 34-1, the potential difference across the terminals of the resistor will also vary sinusoidally in the same manner, *in phase* with the current, in accordance

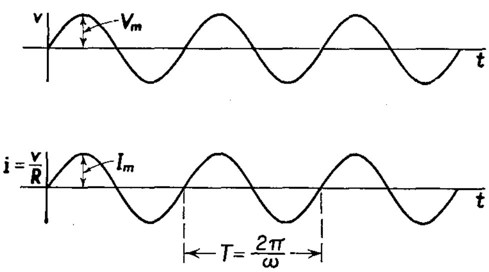


Fig. 34-1 Instantaneous values of the alternating current and voltage in a resistor.

with Ohm's law. Writing i for the instantaneous value of the current and v for the instantaneous value of the potential difference, we have

$$v = iR. \tag{34-1}$$

The current and the potential difference reverse direction at the same time and also reach their maximum values and their minimum values simultaneously. We may describe the sinusoidal variation of the current with time by writing

$$i = I_m \sin \omega t, \tag{34-2}$$

where I_m is the *maximum* value of the current and ω is the *angular frequency* expressed in radians per second. The angular frequency ω is related to the *frequency* f of the current through the equation

$$\omega = 2\pi f. \tag{34-3}$$

The potential difference across the terminals of the resistor is obtained by substituting from Equation (34-2) into Equation (34-1) to obtain

$$v = I_m R \sin \omega t. \tag{34-4}$$

By analogy with Equation (34-2), we might describe the instantaneous value of the potential difference by the equation

$$v = V_m \sin \omega t, \tag{34-5}$$

where V_m represents the maximum value of the potential difference across the resistor. Comparing Equations (34-4) and (34-5), we find

$$V_m = I_m R, \tag{34-6}$$

that is, Equation (34-1) is true for the maximum values of the current and potential difference as well as every other instantaneous value.

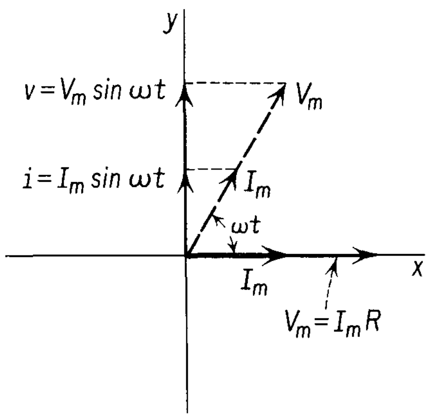


Fig. 34-2

It is convenient to think of the current and potential difference in a-c circuits in terms of the idea of the reference circle developed in our description of simple harmonic motion in Chapter 12. Let us draw a vector of length I_m along the x axis and imagine this vector to rotate in the x-y plane with an angular velocity ω, in the counterclockwise direction. The instantaneous value of the current is given by the projection of this vector onto the y axis. Thus, if the current vector is directed along the x axis at time $t = 0$, the instantaneous value of the current at that time is $i = 0$, as shown in Figure 34-2. This value is in agreement with Equation (34-2). At some subsequent time t the current vector will have rotated through an angle ωt, and the projection of the current vector onto the y axis will have a value given by

$$i = I_m \sin \omega t.$$

The potential difference between the terminals of a resistor may be represented by a vector of length V_m which is rotating with the same angular velocity. Since the current and potential difference are in phase with each other, the voltage vector is also drawn along the x axis and may be imagined to rotate along with the current vector.

One way to keep the relationships of current and voltage in mind is to think of these vectors as drawn on a transparent card which is rotated with constant angular velocity. The instantaneous voltage and current

are then the projection of the voltage and current vectors onto the y axis.

34-2 Effective Values of Current and Voltage

One of the important effects of a current is the production of *heat* in its passage through a resistor. The heating effect is used to define the *effective* value of a given alternating current as compared to a steady *direct* current. *The effective value of an alternating current* I_{eff} *is equal to the direct current which would develop the same heat in a resistor in the same period of time.* To find the effective value of the current, we must find the heat liberated in a resistor in one complete cycle. The rate of development of heat is the instantaneous power $\mathcal{P}$ given by

$$\mathcal{P} = i^2R = I_m^2 R \sin^2 \omega t,$$

and the heat $\mathcal{W}$ developed in a complete cycle of period T $(= 2\pi/\omega)$ is given by

$$\mathcal{W} = \int_0^T I_m^2 R \sin^2 \omega t \, dt.$$

Now

$$\int \sin^2 x \, dx = \tfrac{1}{2}x - \tfrac{1}{4} \sin 2x,$$

so that

$$\mathcal{W} = \tfrac{1}{2} I_m^2 RT.$$

The heat which would be developed by a direct current I_{eff} in the same time would be

$$\mathcal{W} = I_{\text{eff}}^2 RT.$$

Equating these two values, we find

$$I_{\text{eff}}^2 = \tfrac{1}{2} I_m^2,$$

or

$$I_{\text{eff}} = \frac{I_m}{\sqrt{2}} = 0.707 I_m. \tag{34-7}$$

The value of I_{eff} calculated above is the square root of the average or mean of the square of the instantaneous current. The effective value of the current is therefore called the root mean square or rms current, I_{rms}.

The effective value of the potential difference between the terminals of the resistor may be found in the same way, for the instantaneous power may be expressed in terms of the voltage as

$$\mathcal{P} = \frac{v^2}{R} = \frac{V_m^2}{R} \sin^2 \omega t.$$

Thus we may write

$$V_{\text{eff}} = \frac{V_m}{\sqrt{2}} = 0.707 V_m. \tag{34-8}$$

Following convention, we shall write the symbols I and V without

subscripts to designate effective values of current and potential difference in a-c circuits. It must be emphasized that the effective values given by Equations (34-7) and (34-8) are correct only when the current and voltage are varying sinusoidally. The customary a-c power supplied by utility companies is designated as 110 volt, 60 cycle. This means that the company endeavors to operate its generators so that the frequency of the alternating voltage delivered to the user is 60 cycles/sec, and that the *effective* voltage is 110 volts. The maximum or *peak* voltage available at the receptacle is therefore $V_m = 156$ volts.

In a-c circuits composed of resistors, inductors, and capacitors, the only circuit element which consumes electrical power and converts it into heat is the resistor. The power consumed in an a-c circuit in which V_R is the effective voltage across the terminals of the resistor is therefore

$$\mathcal{P} = V_R I_R, \tag{34-9a}$$

$$\mathcal{P} = I_R^2 R, \tag{34-9b}$$

$$\mathcal{P} = \frac{V_R^2}{R} \cdot \tag{34-9c}$$

In Equations (34-9) the symbol $\mathcal{P}$ is understood to represent the average power, while the symbols V_R and I_R are understood to represent the *effective* values of the current through the resistor and the voltage across the resistor.

34-3 Inductance in an A-C Circuit

We have already seen that if a potential difference v is supplied from an external source to a circuit containing inductance L and resistance R, then

$$v = Ri + L\frac{di}{dt} \cdot \tag{32-10}$$

If the resistance of the circuit is negligible, then

$$v = L\frac{di}{dt} \cdot$$

When the current in the inductor is sinusoidal,

$$i = I_m \sin \omega t,$$

and

$$\frac{di}{dt} = I_m \omega \cos \omega t.$$

To find the phase relationship between the emf induced in the inductor

and the current in it, we apply the trigonometric indentity

$$\sin\left(x \pm \frac{\pi}{2}\right) = \pm \cos x.$$

Thus

$$v = LI_m\omega \sin\left(\omega t + \frac{\pi}{2}\right). \tag{34-10}$$

This equation may be written as

$$v = V_m \sin\left(\omega t + \frac{\pi}{2}\right), \tag{34-11}$$

where V_m is the maximum value of the voltage across the inductor. If we wish to relate the *maximum* value of the voltage drop across the inductor to the *maximum* value of the current in it, we find, comparing Equations (34-10) and (34-11),

$$V_m = I_m\omega L. \tag{34-12}$$

The maximum value of the current in the inductor and the maximum value of the potential difference between its terminals do not occur at the same time, for the potential difference is greatest when the rate of change

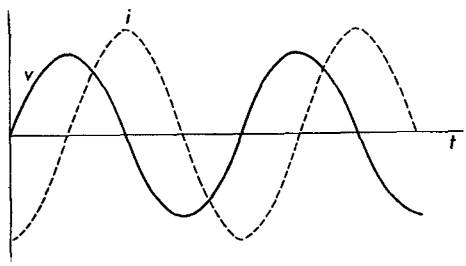


Fig. 34-3 The voltage leads the current by 90° in an inductor.

of current is a maximum; that is, when the current itself is zero. The current and voltage relationships for an inductor are drawn as a function of time in Figure 34-3. We describe these phase relationships by saying that *the voltage across an inductor leads the current through it by* 90°. The word *lead* is associated with the fact that at a time t, when the phase angle of the current is given by Equation (34-2) as ωt, the phase angle of the voltage is given by Equation (34-11) as $\omega t + \pi/2$. This statement does not imply that there is a voltage across the terminals of the inductor before any current is flowing through it, but is rather to be applied to the steady state when an alternating current has been established.

These phase relationships may be described with the aid of appropriate vectors. If the maximum value of the current is drawn as a vector in the $+x$ direction, the maximum value of the voltage across the inductor may be drawn as a vector in the $+y$ direction, as shown in Figure 34-4. If we

think of these vectors as drawn onto a transparent card which is rotated in the counterclockwise direction with angular velocity ω, the projection of these vectors on the y axis at any instant of time t may then be thought of as the instantaneous values of the current and voltage. At a time t, when the phase angle of the current vector is ωt, the phase angle of the voltage vector is $\omega t + \pi/2$. The instantaneous value of the current is, from Figure 34-4,

$$i = I_m \sin \omega t, \tag{34-2}$$

while the instantaneous value of the voltage is

$$v = V_m \sin\left(\omega t + \frac{\pi}{2}\right), \tag{34-11}$$

so that, from Equation (34-12),

$$v = I_m \omega L \sin\left(\omega t + \frac{\pi}{2}\right). \tag{34-10}$$

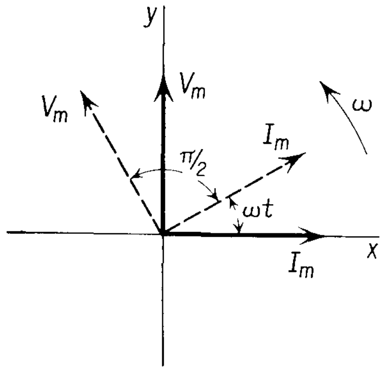


Fig. 34-4

It is customary to use the symbol X_L, called the *inductive reactance*, defined by means of the equation

$$X_L = \omega L = 2\pi f L, \tag{34-13}$$

to describe the behavior of an inductor in an a-c circuit. In these terms Equation (34-12) becomes

$$V_m = I_m X_L. \tag{34-14}$$

The inductive reactance is expressed in *ohms* when the inductance is expressed in *henrys* and the frequency f is expressed in cycles per second, or the angular frequency ω is expressed in radians per second.

34-4 Capacitance in an A-C Circuit

When a steady potential difference is applied to a capacitor, current flows only while the capacitor is being charged. There is a transient rise of the potential difference across the terminals of the capacitor which occurs in an exponential manner, similar to the transient increase in current through an inductor, described in Section 32-6. If an alternating voltage is applied to its terminals, the capacitor is charged and discharged periodically, and we say that current flows *through* the capacitor, even though no electrons actually pass through the dielectric medium separating the capacitor plates. While a steady direct current may flow through a resistor or an inductor, it is clear that the average direct current in a capacitor in a

sufficiently long time interval must be zero, for as much charge must flow in one direction as in the other.

If the charge on one plate of the capacitor at any instant is q, the potential difference between the plates of the capacitor at that instant is v, given by

$$v = \frac{q}{C}.$$

The charge on the capacitor plate is equal to the integral of the current over the time during which the charge was flowing into the capacitor, so that

$$Cv = \int i\,dt.$$

If the current is sinusoidal,

$$i = I_m \sin \omega t, \tag{34-2}$$

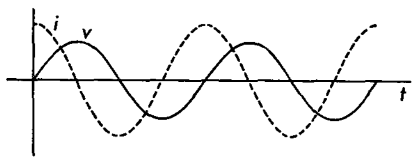
and we have

$$Cv = \int I_m \sin \omega t\,dt,$$

$$Cv = \frac{-I_m}{\omega} \cos \omega t,$$

$$v = \frac{I_m}{\omega C} \sin\left(\omega t - \frac{\pi}{2}\right). \tag{34-15}$$

Fig. 34-5 The current in a capacitor leads the voltage across the capacitor by 90°, or the voltage lags by 90°.



In the above development, the constant of integration, representing the initial charge on the capacitor, has been set equal to zero. The potential difference between the plates of the capacitor may be expressed as

$$v = V_m \sin\left(\omega t - \frac{\pi}{2}\right), \tag{34-16}$$

where

$$V_m = \frac{I_m}{\omega C}. \tag{34-17}$$

Thus we see that the potential difference between the plates of the capacitor lags behind the current by 90°, as shown in Figure 34-5. If the maximum value of the current is drawn as a vector in the $+x$ direction, the maximum voltage across the capacitor may be represented as a vector drawn in the

$-y$ direction, as shown in Figure 34-6. Once again the instantaneous values of the current and voltage are to be found by examining the projection of these two vectors on the y axis, and imagined to rotate in the counterclockwise direction with angular velocity ω.

It is customary to use the symbol X_c, called the *capacitive reactance*, defined by means of the equation

$$X_c = \frac{1}{\omega C} = \frac{1}{2\pi f C}, \tag{34-18}$$

to describe the behavior of a capacitor in an a-c circuit. In these terms Equation (34-17) becomes

$$V_m = I_m X_c. \tag{34-19}$$

When alternating current flows through an ideal capacitor, no electrical energy is consumed by the capacitor. Instead, the capacitor stores energy in the electric field between its plates while being charged, and returns that energy to the circuit when it is discharging.

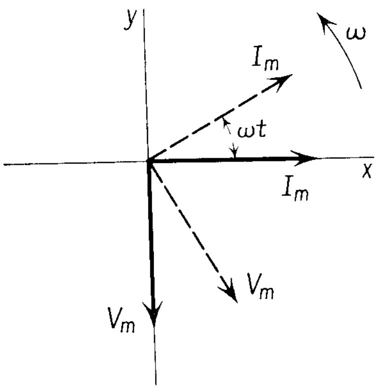


Fig. 34-6

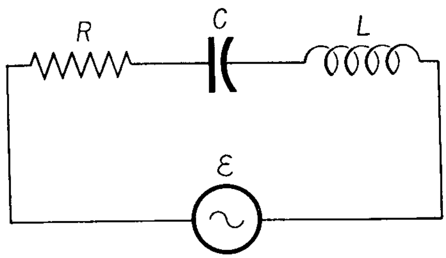


Fig. 34-7

In our initial discussion of capacitance in Chapter 25, we saw that one application of capacitors was to store electrical energy. By far the largest application of capacitors is as a circuit element in a-c circuits. One important use of capacitors is associated with the fact that direct current cannot flow through a capacitor, but that alternating current does. The capacitor may be used to block the flow of direct current.

34-5 Series Circuits

Let us suppose that an alternating current is flowing in a series circuit made up of a resistor R, an inductor L, and a capacitor C, and that these are connected to an a-c generator, as shown in Figure 34-7. At any instant of

time the current must be the same in all parts of the circuit. If the current varies sinusoidally with time, as given by

$$i = I_m \sin \omega t, \tag{34-2}$$

the potential difference across the circuit elements will be

$$v_R = I_m R \sin \omega t, \tag{34-4}$$

$$v_L = I_m \omega L \sin\left(\omega t + \frac{\pi}{2}\right), \tag{34-10}$$

$$v_C = \frac{I_m}{\omega C} \sin\left(\omega t - \frac{\pi}{2}\right); \tag{34-15}$$

and the value of the emf across the terminals of the generator at any instant will be the sum of these potential differences at the same instant of time, according to Kirchhoff's second law (Section 27-6). To find the maximum value of the emf from these equations, we should have to find the instant t at which the sum of these equations is a maximum, and then add the separate potential differences at that time. The problem may be greatly simplified by use of the *vector diagram.*

If we draw a vector representing the maximum value of the current I_m along the $+x$ axis, the vectors representing the maximum values of the potential differences across the terminals of the various circuit elements are shown in Figure 34-8. We imagine the diagram to be drawn on a transparent card which is rotating in the counterclockwise direction with angular velocity ω. The instantaneous value of the current and of the voltage across each element at a given time t is to be determined by finding the projection of the appropriate vector onto the y axis. Thus the instantaneous emf of the generator is given by adding the y components of the three voltage vectors.

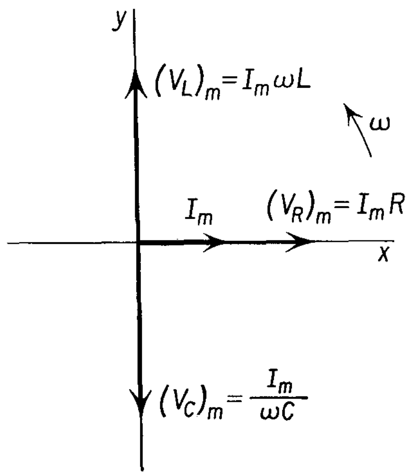


Fig. 34-8

The concepts of vector addition tell us that the sum of the y components of three vectors is equal to the y component of their resultant. Thus we may add the three voltages vectorially and draw a new diagram, shown in Figure 34-9, in which the maximum value of the current I_m is one

vector, represented along the x axis, and the vector sum of the separate voltages, represented as V_m, is the other. The vector V_m is clearly equal to the maximum value of the potential difference across the three circuit elements and is therefore the maximum value of the emf of the generator. The angle ϕ is the angle by which the voltage vector leads the current vector

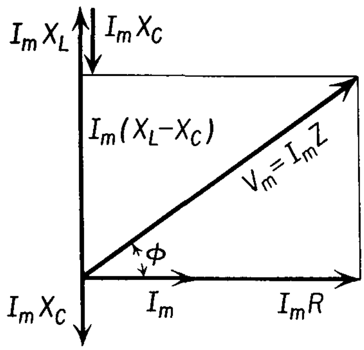


Fig. 34-9 Vector diagram of the voltages and currents in a series circuit containing resistance, inductance, and capacitance.

in the figure and is therefore the phase difference between voltage and current.

From the figure we find that

$$\mathcal{E}_m = V_m = \{[I_m(X_L - X_C)]^2 + (I_mR)^2\}^{1/2}, \tag{34-20}$$

$$\mathcal{E}_m = V_m = I_m [(X_L - X_C)^2 + R^2]^{1/2}. \tag{34-21}$$

The quantity in brackets in Equation (34-21) is called the *impedance* Z of the series circuit. Thus

$$Z = [(X_L - X_C)^2 + R^2]^{1/2}. \tag{34-22}$$

The maximum value of the voltage across the terminals of an a-c circuit may be expressed in terms of the impedance as

$$V_m = I_mZ. \tag{34-23}$$

This equation may be taken as the definition of the impedance of any a-c circuit.

The instantaneous value of the voltage may be seen from Figure 34-9 to be

$$v = V_m \sin (\omega t + \phi), \tag{34-24}$$

where the phase angle ϕ is given by

$$\phi = \arc\tan \frac{X_L - X_C}{R}. \tag{34-25}$$

The power delivered to the circuit is consumed only in the resistor. The power consumed is therefore the product of the *effective* voltage across

the resistor by the *effective* current, according to Section 34-2:

$$\mathcal{P} = V_R I. \tag{34-9a}$$

The effective values of voltage and current used in Equation (34-9a) are equal to the maximum values of these quantities divided by the square root of 2. Thus

$$\mathcal{P} = \frac{(V_R)_m I_m}{2}.$$

From Figure 34-9 the maximum voltage across the resistor is related to the maximum voltage across the entire circuit through the cosine of the phase angle

$$(V_R)_m = V_m \cos \phi,$$

so that the power consumed is

$$\mathcal{P} = \frac{V_m I_m \cos \phi}{2}.$$

In terms of the effective voltage across the circuit and the effective current through the circuit

$$\mathcal{P} = VI \cos \phi. \tag{34-26}$$

The factor $\cos \phi$ is called the *power factor.*

It is customary to analyze a-c circuits in terms of effective current and voltage rather than maximum current and voltage. Since the effective values are related to the maximum values by a constant factor, the square root of 2, a vector diagram may be drawn to relate effective values just like Figure 34-9, without altering any of the phase relationships between voltage and current.

In general, the voltage and current indicated by a-c measuring instruments are effective values rather than maximum values.

In terms of the effective values we have for each element:

$$V_R = IR, \tag{34-27a}$$

$$V_L = IX_L, \tag{34-27b}$$

$$V_C = IX_C; \tag{34-27c}$$

and for the entire circuit,

$$V = IZ. \tag{34-27d}$$

Illustrative Example. A series a-c circuit consists of a capacitor of 8 μfd, an inductor of 600 millihenrys, and a resistor of 48 ohms resistance. Determine (a) the current in the circuit, (b) the voltage across each element in the circuit, (c) the phase angle, and (d) the power supplied to the circuit when the terminal voltage of the 60-cycle generator is 220 volts.

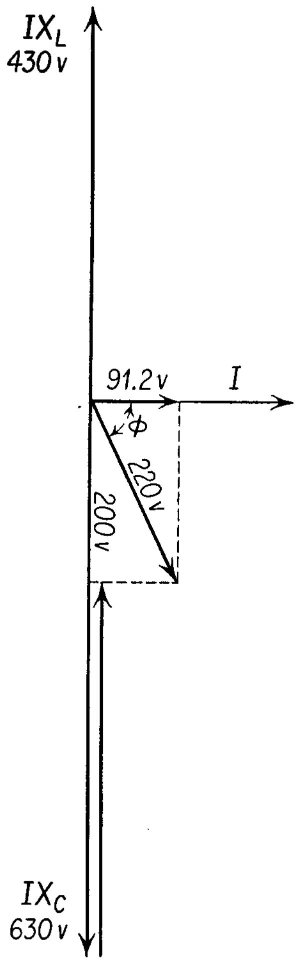


Fig. 34-10

The capacitive reactance is

$$X_C = \frac{1}{2\pi f C} = \frac{1}{2\pi \times 60 \times 8 \times 10^{-6}} \text{ ohms}$$
$$= 332 \text{ ohms}.$$

The inductive reactance is

$$X_L = 2\pi f L = 2\pi \times 60 \times 0.60 \text{ ohms} = 226 \text{ ohms}.$$

The total impedance is

$$Z = \sqrt{R^2 + (X_L - X_C)^2}$$
$$= \sqrt{(48)^2 + (226 - 332)^2} \text{ ohms} = 116 \text{ ohms}.$$

(a) The current in the circuit is

$$I = \frac{V}{Z} = \frac{220}{116} = 1.90 \text{ amp}.$$

(b) Constructing the vector diagram as shown in Figure 34-10, with the current I along the x axis, we find that the voltage across the resistor is

$$V_R = IR = 1.90 \times 48 \text{ volts} = 91.2 \text{ volts}$$

in phase with the current and therefore drawn along the x axis; the voltage across the capacitance is

$$V_C = IX_C = 1.9 \times 332 \text{ volts} = 630 \text{ volts}$$

in the negative y direction, since the curent leads the voltage by 90°; the voltage across the inductance is

$$V_L = IX_L = 1.9 \times 226 \text{ volts} = 430 \text{ volts}$$

in the positive y direction, since the current lags behind the voltage by 90°.

The impressed voltage V is the vector sum of these individual voltages and is given by

$$V = \sqrt{V_R^2 + (V_L - V_C)^2} = \sqrt{(91.2)^2 + (630 - 430)^2} = 220 \text{ volts}.$$

This is a convenient way of checking the calculations.

(c) The phase angle is determined from the equation

$$\tan \phi = \frac{X_L - X_C}{R} = \frac{226 - 332}{48},$$

so that
$$\tan \phi = \frac{-106}{48} = -2.21,$$

and
$$\phi = -65°40';$$

hence the current leads the voltage by 65°40′.

It will be noted that the voltage across a capacitor or an inductor in an a-c series circuit may be much greater than the voltage across the terminals of the circuit.

(d) The power supplied by the generator is

$$\begin{aligned} P &= VI \cos \phi \\ &= 220 \times 1.9 \times \cos 65°40' \\ &= 173 \text{ watts.} \end{aligned}$$

34-6 Resonance in an A-C Series Circuit

A case of very great interest is one in which the current and the voltage are in phase in an a-c series circuit containing resistance, inductance, and capacitance. This can be established by adjusting the values of C and L so that

$$X_L = X_C, \tag{34-28}$$

in which case the phase angle ϕ will be zero, since

$$\tan \phi = \frac{X_L - X_C}{R}. \tag{34-25}$$

Putting in the values for X_L and X_C in Equation (34-28), we get

$$2\pi f L = \frac{1}{2\pi f C},$$

from which

$$f = \frac{1}{2\pi\sqrt{LC}}. \tag{34-29}$$

When this condition is fulfilled, the current in the circuit will be a maximum, since the impedance is a minimum, and will be given simply by

$$I = \frac{V}{R},$$

since for this case

$$Z = R.$$

A circuit for which Equation (34-29) holds is said to be in *resonance* at the frequency f. When the frequency of the a-c supply is that given by Equation (34-29), there is a maximum transfer of energy from the generator to the circuit, since the phase angle ϕ is zero and the power factor is 1. For example, if, in the illustrative example of the previous section, sufficient inductance is added to the circuit either by inserting more iron in the inductance coils or by adding additional inductance coils so that the circuit is in resonance at 60 cycles/sec then the current in the circuit would be

increased to its maximum value given by

$$I = \frac{V}{R} = \frac{220}{48}\text{ amp} = 4.58\text{ amp}.$$

To determine the new value of the inductance, we can solve Equation (34-29) for L and get

$$L = \frac{1}{4\pi^2 f^2 C} = \frac{1}{4\pi^2 \times 3{,}600 \times 8 \times 10^{-6}}\text{ henry}$$
$$= 0.88\text{ henry} = 880\text{ millihenrys}.$$

34-7 The Transformer

The electric energy which is transmitted from the generating station to the consumer is transmitted during a time interval t, and, in designing a transmission system, it is the power or the rate at which the energy is transmitted that is of importance. If the terminal voltage of a d-c generator is V, the power delivered to the transmission line is

$$\mathcal{P} = VI,$$

where I is the current in the line. If the transmission line has a resistance R, then the rate at which heat is developed in the line is I^2R, and hence the power $\mathcal{P}$ delivered to the consumer is

$$\mathcal{P} = VI - I^2R. \tag{34-30}$$

A greater amount of power can be transmitted to the consumer by reducing the resistance of the power line, that is, by using wires of larger diameters, or else by transmitting the power at smaller currents. The latter method means stepping up the voltage at the generating station.

It has been found difficult to build d-c generators which will develop emf's greater than about 3,000 volts. Hence to transmit direct current at higher voltages, it would be necessary to connect several generators in series. This practice is not commonly followed in this country. Another difficulty is that, for safe handling, the voltage at the consumer's end of the line must be comparatively low—not more than a few hundred volts— and no efficient methods have been developed for stepping down the voltage of a d-c line. For a-c generating stations, however, the problem is entirely different. With the aid of a device known as a *transformer*, it is possible to step up the voltage at the transmission line to any desired value, and then to use another transformer at the consumer's end of the line to step down the voltage to a safe, usable value, the power having meanwhile been transmitted at a high voltage and low current. Modern transmission lines are operated at voltages as high as 250,000 volts.

A transformer consists of two coils near each other. In most transformers these coils are wound on closed iron cores such as that shown in Figure 34-11(a). The conventional diagram of an iron-core transformer is shown in Figure 34-11(b). For special uses, particularly in some radio circuits, transformers are made without iron cores; these are usually called air-core transformers.

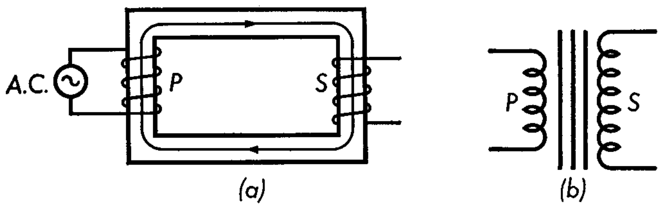


Fig. 34-11 Iron-core transformer.

Suppose that the primary coil P of an iron-core transformer is connected to an a-c source and that the effective voltage across its terminals is $\mathcal{E}$. Let us suppose initially that the terminals of the secondary coil are left open. The current that now flows through the primary coil sets up a magnetic field in the iron core. Because of the high permeability of the iron, practically the entire magnetic flux is inside the iron core. Since the current is alternating, the magnetic flux in the iron core is also alternating. This alternating magnetic flux induces an emf in each turn of the secondary coil, and hence the induced emf $\mathcal{E}_S$ in the secondary coil is proportional to the number of turns of wire n_S in the secondary coil. Since the same magnetic flux goes through each turn of the primary coil, an emf will be induced in each turn of the primary coil so that the total self-induced emf in the primary coil $\mathcal{E}_P$ will be proportional to the number of turns n_P in the primary coil. Since the magnetic flux is changing at the same rate inside each coil, we can write that

$$\frac{\mathcal{E}_S}{\mathcal{E}_P} = \frac{n_S}{n_P}. \tag{34-31}$$

Equation (34-31) holds for both the instantaneous values of the emf's and for their effective values.

In a well-designed transformer, $\mathcal{E}_P$ will differ very slightly from the impressed voltage $\mathcal{E}$. Thus, to a very close approximation, Equation (34-31) may be written as

$$\frac{\mathcal{E}_S}{\mathcal{E}} = \frac{n_S}{n_P}. \tag{34-32}$$

If the number of turns n_S in the secondary coil is greater than the number of turns n_P in the primary coil, the transformer is called a *step-up* transformer;

if the reverse is the case, it is a *step-down* transformer. For example, if the secondary coil has 1,000 times as many turns as the primary coil, the emf $\mathscr{E}_S$ induced in the secondary will be 1,000 times the voltage impressed across the primary.

When a load is connected to the terminals of the secondary coil, a current will flow in the secondary circuit, and power will be supplied by it. This power must, of course, come from the source of power connected to the primary coil. This transfer of power takes place through the interactions of the magnetic fields because of the current in the primary coil and that in the secondary coil. In well-designed transformers the efficiency is as high as 98 or 99 per cent. Neglecting the slight loss of power in heating the coils and the iron core, we find that the power drawn from the secondary coil must equal the power supplied to the primary coil; that is,

$$\mathscr{E} i_P = \mathscr{E}_S i_S, \tag{34-33}$$

where the symbols refer to the instantaneous values of the voltage and current in the primary and secondary coils, respectively. Or

$$\frac{\mathscr{E}}{\mathscr{E}_S} = \frac{i_S}{i_P}. \tag{34-34}$$

Since Equation (34-34) holds at any instant, it also holds for the maximum values and hence for the effective values, so that we can write

$$\frac{\mathscr{E}}{\mathscr{E}_S} = \frac{I_S}{I_P}, \tag{34-35}$$

which, combined with Equation (34-32), yields

$$\frac{I_S}{I_P} = \frac{n_P}{n_S}. \tag{34-36}$$

Thus the effective values of the currents in the primary and secondary circuits are in the inverse ratio of the numbers of turns in the two coils.

If we rewrite Equation (34-36) as

$$n_S I_S = n_P I_P, \tag{34-37}$$

we note that, if the current in the secondary is increased, as is the case when the load on the secondary is increased, the current in the primary is also increased.

In a step-up transformer the emf induced in the secondary is large but the current I_S is small, while the voltage across the primary is small and the current I_P through it is large. Both step-up and step-down transformers are used in transmitting power. A simplified version of a transmission system is shown in Figure 34-12. At the powerhouse, the a-c generator develops electric power at, say, 120 volts; the terminals of this generator

are connected to the terminals of the primary coil of a step-up transformer T_1 which steps up the voltage to 12,000 volts at the terminals of the secondary coil. The two wires of the transmission line, which may be several miles long, connect the terminals of this secondary coil to the primary of a step-

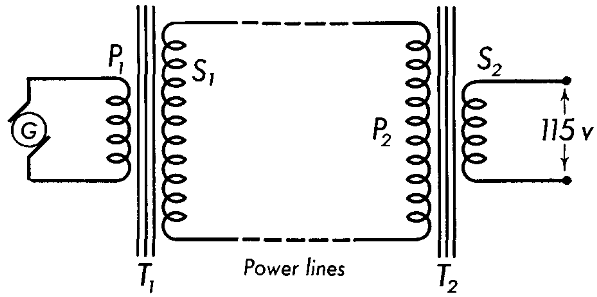


Fig. 34-12 Simple transmission line.

down transformer T_2 at the consumer's end of the line. Because of the voltage drop in the transmission line, the difference of potential at the primary of the step-down transformer may be only 11,500 volts, say. If the ratio of turns between P_2 and S_2 is 100 : 1, the emf at the terminals of S_2 will be about 115 volts, suitable for use with a great many electric appliances.

Problems

34-1. The terminals of a series circuit consisting of a 32-millihenry inductor and a 48-ohm resistor are connected to a source of emf supplying 110-volt, 60-cycle power. Find (a) the reactance of the inductor, (b) the impedance of the circuit, (c) the effective current (d) the maximum current, (e) the phase angle between the current and voltage, (f) the effective voltage across the resistor, (g) the effective voltage across the inductor, and (h) the power dissipated in the circuit.

34-2. A coil with an inductance of 0.020 henry and an internal resistance of 8 ohms is connected in series with a 75-ohm resistor. The series combination is connected to a 110-volt, 60-cycle generator. Find (a) the inductive reactance of the coil, (b) the impedance of the coil, (c) the impedance of the series circuit, (d) the current in the circuit, and (e) the power dissipated in the circuit.

34-3. By setting

$$i = I_m \sin \omega t$$

and

$$\begin{aligned} v &= V_m \sin (\omega t \pm \phi) \\ &= \pm V_m \cos \omega t \end{aligned}$$

and integrating over one cycle, show that the power consumed in an inductor or in a capacitor is equal to zero.

34-4. A 2 μfd capacitor is connected in series with a 36-ohm resistor, and the series combination is connected to a 120-volt, 60-cycle power line. Find (a) the capacitive reactance, (b) the impedance, (c) the current, and (d) the power consumed in the circuit.

34-5. A 40-ohm resistor is connected in series with a 1.8-henry inductor and a 10-μfd capacitor. (a) Determine the inductive reactance, the capacitive reactance, and the impedance of the circuit at 60 cycles. (b) Determine the current in the circuit when it is connected to a 60-cycle, 110-volt source. (c) Determine the effective voltage across the resistor, across the capacitor, and across the inductor. Is the sum of these different from 110 volts? Why?

34-6. A 45-ohm resistor, an 8-μfd capacitor, and an 0.06-henry inductor are connected in series. (a) What is the resonant frequency of this circuit? (b) When connected to a 100-volt power source at this frequency, what is the current in the circuit? (c) Draw a vector diagram of the circuit at resonance. (d) What is the voltage across the inductor at resonance? (e) What is the voltage across the resistor at resonance? (f) What is the power factor at resonance?

34-7. Plot a graph of the impedance of the circuit of Problem 34-6 as a function of frequency, giving particular emphasis to the shape of the curve in the vicinity of resonance.

34-8. An iron-core transformer has 100 turns in the primary winding and 800 turns in the secondary winding and is operated from a 120-volt, 60-cycle generator. Determine (a) the emf induced in the secondary, (b) the current in the secondary circuit when it is taking 2.4 kw of power with a power factor of 0.75, and (c) the current in the primary.

34-9. A transformer is used to step down the voltage of a transmission line from 13,200 volts to 240 volts. (a) What is the ratio of the turns on the two windings? (b) If the secondary supplies 15 amp, determine the current in the primary.

34-10. A capacitor is connected in series with a resistor, and the combination is then connected to a battery whose emf is V. Show that the potential difference across the terminals of the capacitor is given by

$$v = V(1 - e^{-(t/RC)}).$$

[HINT: Follow the development of Section 32-7.]

34-11. An alternator furnishes 80 amp at 240 volts at a frequency of 60 cycles/sec to a shop in which the power consumed is 17.6 kw. What series capacitance must be introduced into the electrical lines to change the power factor to unity?

34-12. A square-wave generator generates current of the wave form shown in Figure 34-13. Find the effective value of the current.

34-13. A series circuit consisting of a 20-ohm resistor and an inductor whose internal resistance is 10 ohms is observed to have an impedance of 50 ohms at a frequency of $100/\pi$ cycles/sec. (a) What is the current in the circuit when the applied emf is 100 volts? (b) What is the inductance in the circuit? (c) What is the impedance of the circuit at a frequency of $300/\pi$ cycles/sec?

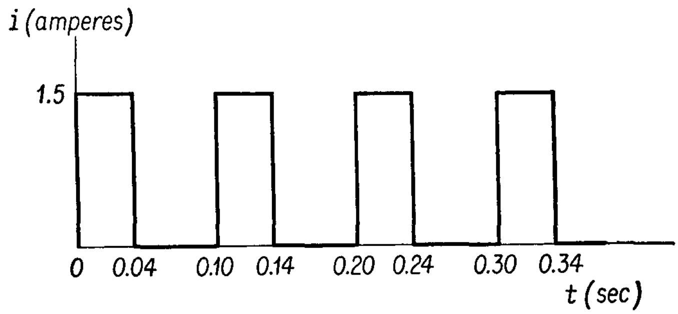


Fig. 34-13

34-14. A series circuit has a power factor of 0.8 with the current lagging the voltage at a frequency of $200/\pi$ cycles/sec. (a) If this circuit has a resistance of 100 ohms, what is the impedance? (b) If this circuit has a capacitance of 10^{-4} farad, what is the inductance? (c) At what frequency will its power factor be zero?

34-15. In a given circuit consisting of a resistor, an inductor, and a capacitor in series, the voltage across the resistor is 100 volts, the voltage across the capacitor is 200 volts, and the voltage across the inductor is 150 volts. The power consumed by the circuit is 150 watts. (a) What is the current in the circuit? (b) What is the power factor? (c) What is the inductive reactance?

35

Electronics

35-1 Thermionic Emission

The electrons of an atom occupy certain energy levels when the atom is far from other atoms. When a large number of atoms are assembled to form a conducting metallic solid, the energy levels of the valence electrons are displaced in position to form a *conduction band,* as discussed in Section 28-4. The valence electrons occupy energy levels within the conduction band and

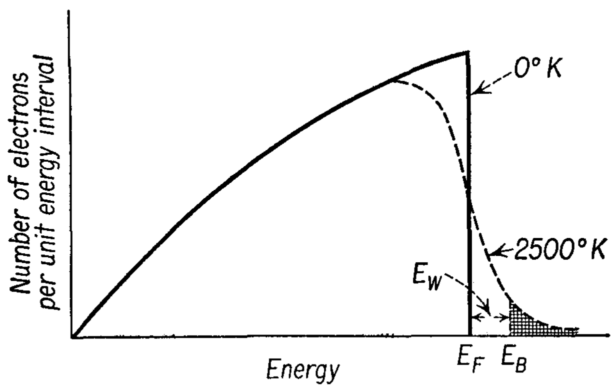


Fig. 35-1 The Fermi-Dirac distribution of electron energies in tungsten at 0°K and 2500°K, showing the position of the Fermi level E_F, the binding energy E_B, and the work function E_W (not to scale). Electrons in the shaded area are thermally emitted at 2500°K if they are traveling toward the surface with sufficient energy.

are relatively free to drift from atom to atom within the metal. In accordance with a fundamental physical principle, known as the *Pauli exclusion principle,* only 1 electron may occupy a particular energy level. The conduction electrons are not all in the state of lowest energy, for only 1 electron can occupy that level. The conduction electrons tend to fill the bottom region of the conduction band. The number of electrons to be found in a

small energy interval centered about a particular energy follows a distribution known as the *Fermi-Dirac* distribution, shown in Figure 35-1. At the absolute zero of temperature, the number of electrons in a given energy interval increases with increasing energy, as shown in the figure. The maximum energy of any conduction electron in the metal is known as the *Fermi* energy E_F. At the absolute zero all energy levels below the Fermi energy are occupied by electrons, and no electron has an energy above the Fermi energy. The particular value of the Fermi energy varies from metal to metal but depends only upon the number of conduction electrons per unit volume of metal.

Another parameter of great interest is the *binding energy*, the energy required to remove an electron from the lowest energy state within the conduction band to a point infinitely distant from the metal. Since electrons normally do not leave the metal, it is apparent that the binding energy E_B is greater than the Fermi energy E_F. The difference between these two energies is called the work function E_W. Thus

$$E_B - E_F = E_W.$$

If an electron at the Fermi level is to be emitted from the body of the metal at the absolute zero, it must be given a quantity of energy E_W. This may be done in a number of ways. When the surface of the metal is struck by a rapidly moving particle, such as a proton or an electron, the kinetic energy of the moving particle is absorbed by the metal and may be given up to some of the electrons in the metal. These electrons are then liberated from the metal by a process called *secondary emission*, which is of great importance in modern *photomultiplier tubes.* A second way for an electron to acquire sufficient energy to be removed from the metal is by the absorption of the energy of a quantum of light; this process is known as *photoelectric emission.*

When the metal is heated to a high temperature, the distribution of electrons within the metal is altered. The energy distribution of electrons in metallic tungsten at 2500°K is shown in dotted lines in Figure 35-1. Those electrons having energies above E_B, shown in the shaded area of the curve, have sufficient energy to escape from the body of the metal and are said to be *thermally emitted.* The fraction of the total number of conduction electrons having sufficient energy to be emitted at a particular temperature depends upon the position of the binding energy E_B with respect to the Fermi energy E_F, hence upon the value of the work function E_W. We must therefore expect the thermionic emission from different metals at a given temperature to vary from metal to metal, according to the differences in work function. From the shape of the distribution curve, we must also expect the thermionic emission to increase rather rapidly with temperature.

The thermionic emission of electrons from a heated filament was discovered by Thomas A. Edison in 1883; it forms the basis of the modern vacuum tube. In such tubes the electron emitter is the negative electrode, or cathode. Because of its high melting point, tungsten is often used for the filaments of tubes which are operated at high cathode temperatures. It is possible to apply low work-function materials to metallic filaments so that these materials act as the source of electrons and the filaments can be operated at lower temperatures for the same emission. In the modern vacuum tube the cathode is often a separate structure surrounding a heater, in order to separate the problem of providing heat from the main function of the cathode as a source of electrons.

35-2 The Diode Rectifier

A thermionic diode is a two-element vacuum tube, in which one element consists of a filament or heated *cathode* which serves as a source of electrons, and the other element, called the *plate*, serves as a collector of electrons. Since the diode is asymmetric in construction, it is also asymmetric in its electrical properties. Electrons can flow in only one direction, from filament to plate.

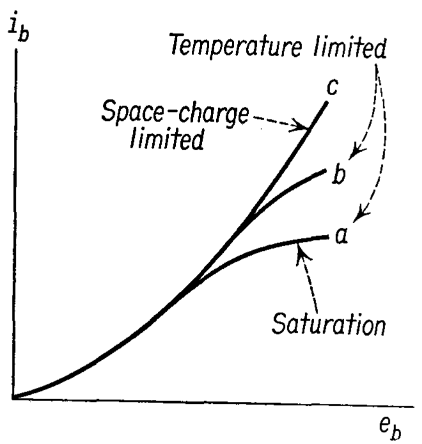


Fig. 35-2 Static characteristics of a typical diode. The filament temperature is increasing from a to c.

When a potential difference is applied between the cathode and plate of the diode, the current through the tube as a function of applied voltage has the form shown in Figure 35-2. The electrons which are emitted from the filament form a space-charge cloud around the filament. When a low positive voltage is applied to the plate, some of the electrons of the space-charge cloud are attracted to the plate; the current is determined by the characteristics of the space-charge cloud. As the plate voltage is increased, more and more of the space charge flows toward the plate, until the current reaches a value appropriate to the cathode emission. The current is then *saturated*, or *temperature limited*. When the temperature of the filament is raised, the current at saturation is greater than previously, as shown in the figure.

Since there is only one direction in which current will flow in a diode, this tube may be used as a rectifier. It is often convenient to think of the

diode as a switch which is closed to permit current to flow when the plate is made positive, with respect to the cathode, and which is opened to stop the flow of current when the plate is made negative.

The basic circuit of a *half-wave* rectifier is shown in Figure 35-3(a). An alternating emf ε is generally supplied by a transformer connected to the a-c line. The transformer is connected to a series combination of the diode and a load, shown here as a resistor R_l. The cathode is shown on the diagram as externally heated. For simplicity, the heater filament and the filament transformer have been omitted. The potential difference between cathode and plate ε_b is equal to the emf of the transformer secondary minus the iR drop in the load resistor. Thus

$$\varepsilon_b = \varepsilon - i_b R_l, \tag{35-1}$$

so that the voltage across the load is given by the difference $\varepsilon - \varepsilon_b$.

We may display the meaning of Equation (35-1) graphically by construction of a *dynamic characteristic* curve of the diode, from its *static characteristic*, by means of the concept of a *load line.* The static characteristic of a tube is the plot of current through the tube as a function of plate voltage, under a particular set of conditions, such as a fixed cathode temperature, as shown in Figure 35-2. The dynamic characteristic of the tube in a particular circuit is a plot of the current through the tube i_b plotted as a function of the emf applied to the circuit ε, when a particular load resistor R_l is used.

One relation between the variables ε, ε_b, and i_b is expressed by Equation (35-1). A second relation between two of these variables is known only as a graph, that is, the static characteristic, which expresses i_b as a function of ε_b. These two functional relationships may be solved simultaneously at specific values of ε for ε_b and i_b, so that we can express i_b as a function of ε.

Let us rewrite Equation (35-1) in the form

$$i_b = -\frac{\varepsilon_b}{R_l} + \frac{\varepsilon}{R_l}. \tag{35-2}$$

On a graph in which i_b is the ordinate and ε_b is the abscissa, as in Figure 35-3(b), this is the equation of a straight line of slope $-1/R_l$ and intercept ε/R_l, from the standard slope-intercept form of the equation of a straight line

$$y = mx + b.$$

Equation (35-2) is the equation of the *load line,* which may be most simply drawn as connecting the points of coordinates $(0, \varepsilon/R_l)$ and $(\varepsilon, 0)$, as shown in Figure 35-3(b).

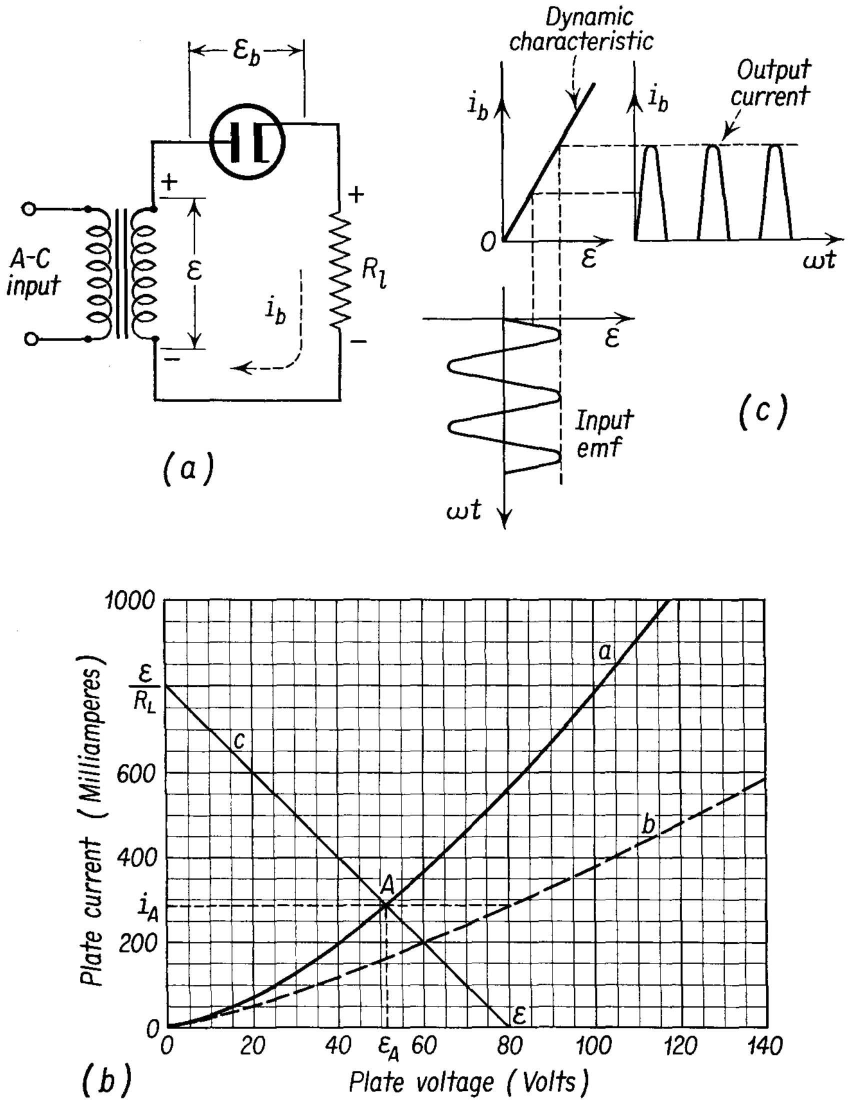


Fig. 35-3 (a) Schematic circuit of a half-wave rectifier. (b) Characteristics of a 5U4–GB diode. Solid line (a) is the static characteristic; dashed line (b) is the dynamic characteristic with 100-ohm load; load line (c) for 100-ohm load resistor drawn for $\mathcal{E} = 80$ volts. When an emf $\mathcal{E}$ is applied to the series circuit of diode and load resistor, the voltage $\mathcal{E}_A$ appears across the tube, and the current i_A flows in the circuit. (Graph courtesy of General Electric Company.) (c) Output current wave form obtained from input voltage wave form from the dynamic characteristic of the diode used in the half-wave rectifier circuit.

The point of intersection of the load line and the static characteristic of the tube represents the solution of the two simultaneous functional relationships for the circuit. For a particular value of the applied emf ε, the intersection of the two curves is at point A, and the current through the tube is i_A. Repeating this operation by displacing the load line parallel to itself for different values of the applied emf ε, we may find the values of ε_b and i_b appropriate to each value of ε.

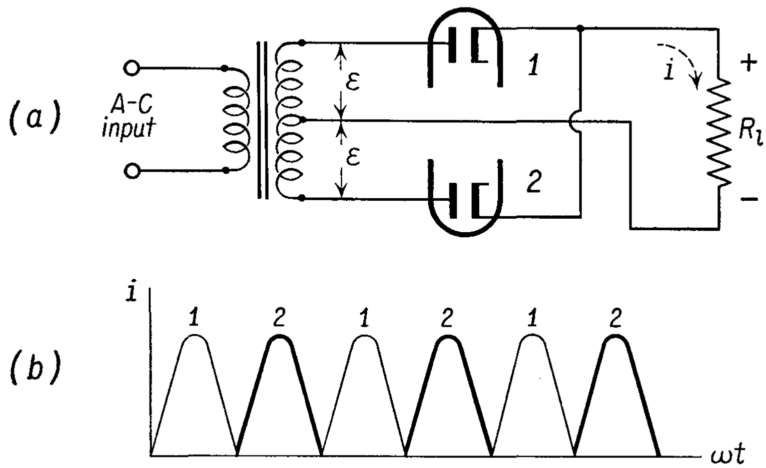


Fig. 35-4 (a) Full-wave rectifier circuit. (b) Current wave form from full-wave rectifier is made up of contributions from the half-wave rectifiers of tubes 1 and 2 which operate on alternate half cycles.

If the applied emf is a sinusoidal function of time, we may imagine time axes to be superimposed upon the dynamic characteristic curve, as shown in Figure 35-3(c), and thus find the wave form of the output current as a function of time for the half-wave rectifier.

When two diodes are so connected that they are conducting during alternate half cycles and yet permit current through the load resistor in a single direction, they constitute a full-wave rectifier. A circuit diagram and wave form resulting from such a rectifier are shown in Figure 35-4.

35-3 Triode Amplifiers

The three-element vacuum tube, or *triode*, has a third element, called a *grid*, often constructed in the form of wire mesh inserted between the cathode and plate. The tube elements sometimes take the form of plane structures and sometimes of concentric cylinders. A potential applied to the grid greatly alters the electric field between cathode and plate. Since the grid is close to the cathode, a small change in grid potential exercises considerable control on the electron current flowing from cathode to plate.

The plate current i_b is a function of two variables, the grid-cathode potential difference ε_c and the cathode-plate potential difference ε_b. The operation of the triode is best analyzed in terms of the static tube characteristics, shown for a typical triode in Figure 35-5.

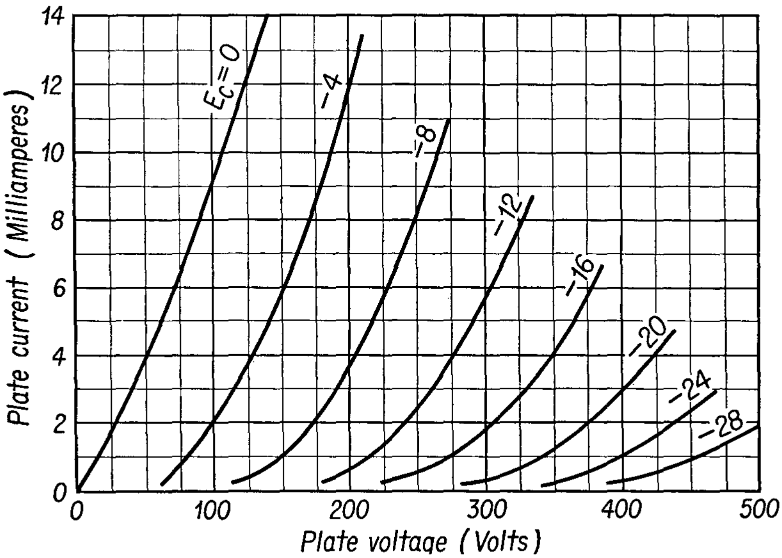


Fig. 35-5 Plate characteristics of the 6C5 triode. (Graph courtesy of General Electric Company.)

In general, we are not interested in the steady or direct current or the fixed electrode potentials except as a means of establishing the operating point of the tube. It is of greater interest to examine the increments in current associated with small increments in the voltage of one or another of the tube elements. Thus we may define the incremental or dynamic *plate resistance* r_p of a tube as the change in the plate-cathode potential difference $\Delta\varepsilon_b$ divided by the resulting change in plate current Δi_b when the other tube element is held at constant potential. Thus

$$r_p = \frac{\Delta\varepsilon_b}{\Delta i_b} \quad ; \quad (\varepsilon_c = \text{const}). \tag{35-3}$$

A second widely used tube characteristic is the *mutual conductance* g_m. The mutual conductance is defined as the change in plate current Δi_b which results from a change in grid potential $\Delta\varepsilon_c$ when the plate potential ε_b is held constant. Thus

$$g_m = \frac{\Delta i_b}{\Delta\varepsilon_c} \quad ; \quad (\varepsilon_b = \text{const}). \tag{35-4}$$

The amplification factor μ is the ratio of the change in plate potential

$\Delta\varepsilon_b$ to the change in grid potential $\Delta\varepsilon_c$ required to produce equal and opposit changes in plate current. If the plate potential is increased by an amount $\Delta\varepsilon_b$ and the grid potential is altered by $-\Delta\varepsilon_c$, there will be no change in the plate current. We have

$$\mu = -\frac{\Delta\varepsilon_b}{\Delta\varepsilon_c} \quad ; \quad (i_b = \text{const}). \tag{35-5}$$

These three tube parameters are related by the equation

$$\mu = r_p g_m. \tag{35-6}$$

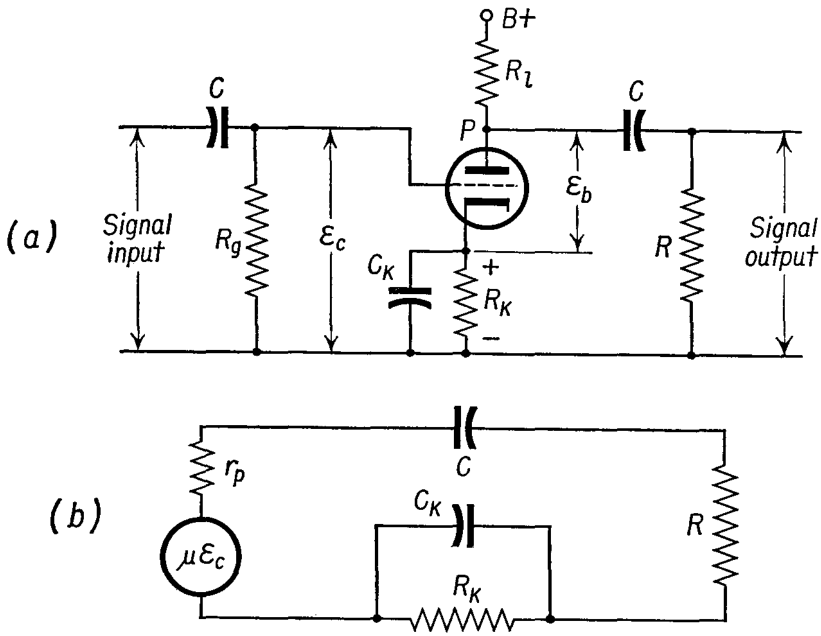


Fig. 35-6 (a) One-stage triode amplifier; grid bias is provided by the voltage drop across the cathode resistor R_{k}. (b) Equivalent circuit of the amplifier. The combination of R_{k} and C_{k} is of negligible impedance at the intermediate signal frequencies, as is the capacitor C.

The values of the plate resistance, the mutual conductance, and the amplification factor are not constant for most tubes and depend upon the operating conditions. Vacuum tubes are nonlinear devices which generally do not obey Ohm's law. Nevertheless, it is often convenient to consider that these parameters are constant for small changes in operating conditions, and to treat the tube as a linear device in discussing vacuum-tube circuits and in preliminary design of such circuits.

A typical schematic circuit of a one-stage triode amplifier is shown in Figure 35-6(a). A power supply consisting of a full-wave rectifier and

suitable filters provides a constant positive potential, shown on the figure only as B^+. By suitable choice of a cathode resistor R_k, the IR_k voltage drop in the cathode resistor establishes the cathode at a suitable positive voltage with respect to the grid, called the *grid bias.*

An input signal is applied to the grid resistor R_g. The variations in grid potential cause the plate current to vary, so that a varying potential is generated at the point P. This varying potential is the output signal of the tube. Since the potential of the point P varies about some high positive potential, it is necessary to connect the point P to a blocking capacitor C, and to take the varying signal output from the opposite terminal of the blocking capacitor, whose function is to permit varying currents to flow while preventing the passage of steady currents. The capacitor C blocks the passage of direct current but permits the passage of alternating current.

One way of analyzing the behavior of such a circuit is to imagine the tube to be replaced by an electrical generator whose emf is $\mu\mathcal{E}_c$ and whose internal resistance is r_p, as shown in Figure 35-6(b). The signal voltage appearing across the load resistor may then be obtained by the application of principles of network analysis.

In this simplified discussion of the operation of a vacuum-tube amplifier, we have neglected the existence of interelectrode capacitances and other limiting effects. Any pair of conductors separated by a nonconductor constitutes a capacitor; there is thus a capacitance between the grid and cathode and between the grid and plate. These interelectrode capacitances are in parallel with the grid resistor. At high frequencies the impedance of the grid network is greatly reduced. This means that the amplification of a vacuum-tube amplifier drops off at high signal frequencies unless special precautions are taken in the design of the tube itself. At low frequencies the coupling capacitor C plays a dominant role in establishing the sensitivity of an amplifier, for it is in series with the resistor R. The voltage division between C and R is frequency dependent, and the amplified signal drops off at low frequencies.

35-4 Other Electron Tubes

Many other electron tubes have been designed for special purposes. In a *tetrode,* or four-element tube, an additional grid, called a *screen grid,* is inserted between the plate and control grid to reduce the interelectrode plate-to-grid capacitance. In a *pentode,* or five-element tube, another grid, called a *suppressor grid,* is installed between the plate and screen grid. Its purpose is to repel secondary electrons back to the plate; these would tend to go to the more positive screen grid in a pentode operating at low plate voltage.

A *photoelectric cell* has a plate coated with a metal of low work function so that it emits photoelectrons when illuminated with light. The photoelectrons are collected by a positive collector electrode; the current through the tube depends upon the intensity of the incident light.

A *cathode-ray* tube, shown in Figure 35-7, is one in which a focused beam of electrons from the cathode is accelerated by a series of electrodes and caused to strike the phosphor on the screen at a point; the phosphor

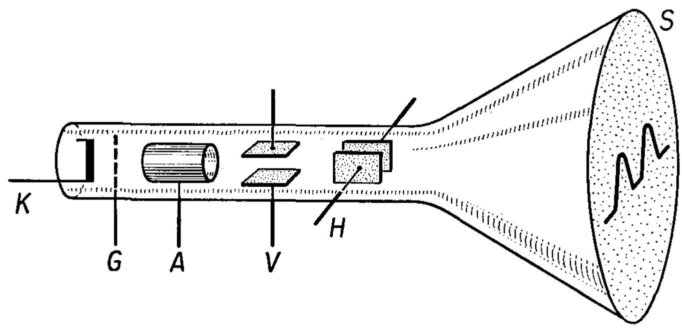


Fig. 35-7 Cathode-ray tube showing cathode K, grid G, anode A, vertical deflection plates V, horizontal deflection plates H, and screen S. Thermally emitted electrons are accelerated and focused by the anode to strike the screen.

fluoresces when struck by rapidly moving electrons. The electron beam is deflected by transverse electric or magnetic fields and can be made to produce a pattern on the screen if the deflecting field is varied. Such a tube is widely used in the laboratory in a device called a cathode-ray oscillograph. In a television set the electron beam is swept across the face of the screen in a series of lines. The intensity of the beam is caused to vary in accordance with the signal transmitted from the television station, producing a pattern of light and dark areas on the screen which we see as a picture.

A *photomultiplier* tube, shown in Figure 35-8, achieves a high output signal by the process of secondary emission. A small number of photoelectrons, emitted from the photoelectric surface, is accelerated by the electric field and caused to strike the first plate or *dynode*. Several secondary electrons are emitted by the dynode for each primary electron which strikes it. These secondary electrons are accelerated and caused to strike a second dynode, and so on, so that after striking some nine dynodes there are about 10^5 to 10^7 electrons collected by the plate for every photoelectron emitted from the photoelectric surface.

In addition to the vacuum tubes in which the electron is the conductor of electricity, there are many tubes filled with gas at low pressure, in which

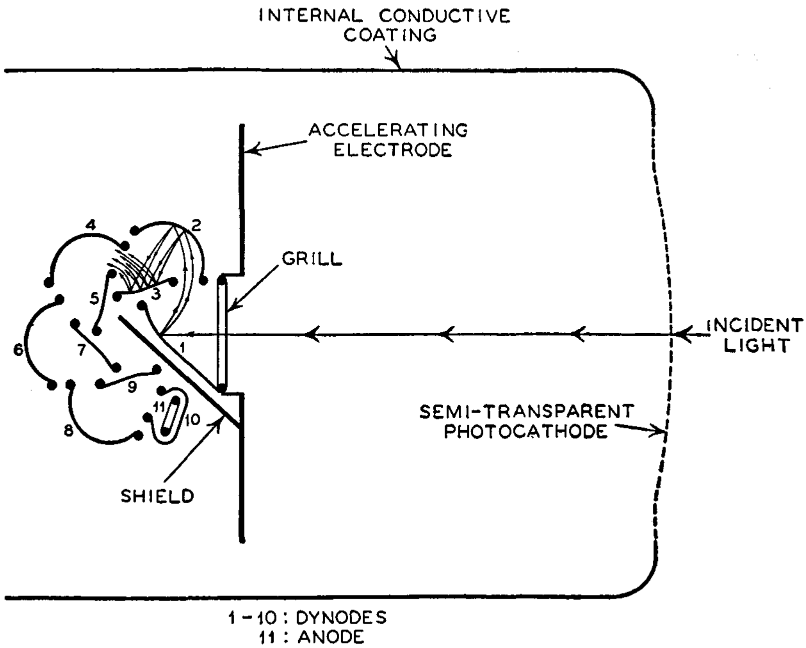


Fig. 35-8 Photomultiplier tube. Schematic diagram showing arrangement of electrode structure. (Courtesy of RCA Victor Division.)

gaseous ions and electrons are the carriers of electricity. In fluorescent lighting units, to be discussed in a subsequent chapter, the gas is used as a source of light. Gaseous rectifier tubes are often used to carry currents larger than can be carried in vacuum-tube diodes. Some gas tubes have the special property that the voltage across the tube is practically independent of the current through the tube. These tubes are used as *voltage regulators*, for, if they are connected across a particular load, the voltage supplied to the load remains constant, regardless of the fluctuation in line voltage.

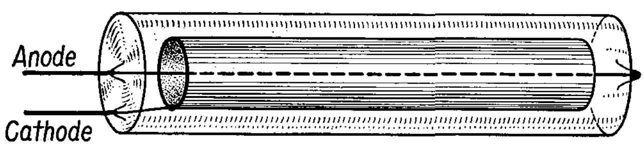


Fig. 35-9 Geiger counter tube.

The *Geiger-counter* tube, shown in Figure 35-9, is a special gas tube which is widely used in nuclear physics. Here a cylindrical anode surrounds a cathode made of a fine wire. The space between the anode and cathode is

filled with a gas at low pressure, approximating 1/10 atm, and a high voltage is applied to the tube. If incident radiations have sufficient energy to ionize a few of the molecules of the gas, the electric field is high enough to accelerate these ions so that they will cause additional ionization when they collide with other gas molecules. In this way an avalanche of gaseous ions is created, and a large burst of current passes through the tube whenever ionizing radiation penetrates the tube.

35-5 Transistors

We have already seen in Section 28-4 that it is possible to construct a rectifier, called a semiconductor diode, from a semiconducting crystal. It is also possible to assemble semiconducting crystals into a device which has properties similar to those of a vacuum-tube triode. Such a device is called a *transistor.*

We will recall that a crystal of germanium, which itself has 4 valence electrons, can be altered by the presence of an antimony impurity, which has 5 valence electrons, so that the excess valence electron can migrate through the crystal. A germanium crystal with this type of impurity is called an n-type semiconductor, for its conductivity is due to negative carriers of electricity.

If indium, having only 3 valence electrons, is introduced as an impurity in germanium, there is a deficiency of electrons at the location of the impurity, and electrons from the germanium may migrate to and from the impurity site. In this case we may think of the hole, or vacancy, as the carrier of electricity, and we designate such semiconductors as p-type semiconductors, for the carrier of electricity may be thought to be the "positive hole."

A transistor is a crystal having two p-type regions separated by an n-type region, in which case it is known as a p-n-p transistor, or having two n-type regions separated by a p-type region, in which case it is called an n-p-n transistor. Figure 35-10 is a schematic diagram of an n-p-n transistor. The electrode connected to the p region is called the *base.* One electrode connected to an n region is called the *emitter*, and the other electrode is connected to the other n region and is called the collector.

The behavior of a transistor can be demonstrated by connecting one battery B_1 so that the base is positive with respect to the emitter, while a second battery B_2 is connected so that the base is negative with respect to the collector. The voltage bias on the collector is positive with respect to the base. As we have seen in Section 28-4, the direction of easy flow in a transistor diode occurs when the n region is connected to the negative terminal of a battery and the p region is connected to the positive terminal. Thus the collector-base interface will have a high resistance when the

emitter-base interface has a low resistance. If the thickness of the intermediate p region is very small, many electrons from the n region on the left will diffuse through the p region without combining with a hole, and will diffuse into the p region on the right toward the collector. Practically all of the current flowing from the emitter will pass to the collector electrode, while a very small current flows from the base to either the emitter or the collector.

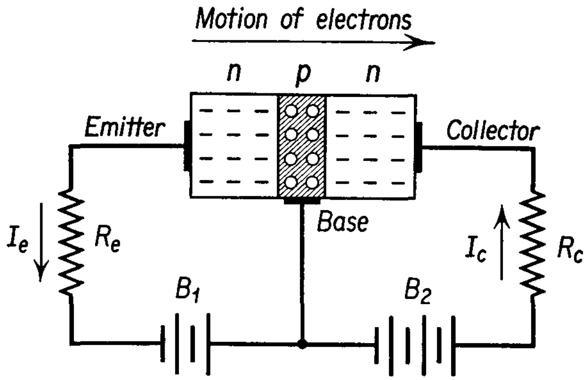


Fig. 35-10 Schematic diagram of the n-p-n transistor.

The base may be likened to the grid of a vacuum triode, the emitter to the cathode, and the collector to the plate. The current flowing from the emitter to the collector depends upon the voltage applied to the base. The signal to be amplified is connected between the emitter and the base. Because of the way the bias voltage is applied, the resistance of the path from emitter to base is small compared to the resistance between emitter and collector.

If we call the current in the emitter branch I_e and the resistance of this branch R_e, the power developed in it is $I_e^2R_e$. Similarly, the power developed in the collector branch is $I_c^2R_c$, where I_c is the current in the collector branch, and R_c is its resistance. The value of R_c is usually much greater than R_e, mainly because of the way the electric fields produced by the batteries B_1 and B_2 are biased in the different sections of the crystal. The ratio of the power delivered to the collector to that in the emitter is

$$\frac{I_c^2R_c}{I_e^2R_e}.$$

Even though I_c may be only slightly larger than I_e, there will still be a large gain in power in the ratio R_c/R_e. The power gain may be as high as 10^5.

The transistor has several advantages over vacuum tubes in many

circuit applications. Since there is no filament to heat, the power consumed by the transistor is very small. The transistor is light in weight and of small physical size. The transistor has good inherent mechanical stability as compared to a vacuum tube, for there are no tube elements whose position may be altered by vibration or shock. At the present time the transistor is limited to the production of small amounts of signal power, of the order of several watts, as compared to a power output of the order of kilowatts which may be obtained from vacuum tubes used in radio transmitters.

35-6 Radiation of Electromagnetic Waves

An electric charge which moves with constant speed in a straight line can radiate no energy. Only when the charge is being accelerated will it radiate electrical energy. A charge, moving at velocities which are small compared

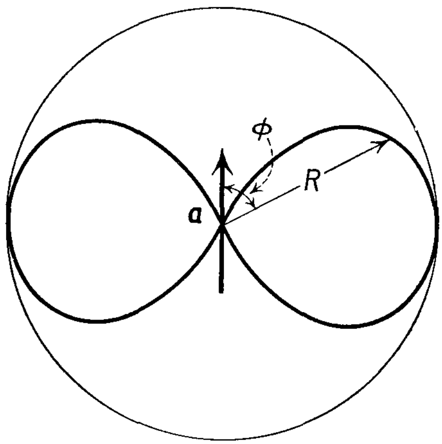


Fig. 35-11 Polar plot of relative rate of radiation from a charged particle whose acceleration vector is **a** in directions making an angle ϕ with the direction of the acceleration; the radius vector R from the position of the particle to the curve is proportional to the rate of radiation in the direction of R.

to the velocity of light in vacuum, does not radiate energy uniformly in all directions. The radiation is a maximum in a direction perpendicular to the direction of the acceleration and is zero in the direction of the acceleration. The rate of radiation of electromagnetic energy from an accelerated charged particle is given by

$$S = \frac{2}{3}\frac{q^2}{c^3}a^2,$$

where S is the rate of radiation in ergs per second from a particle of charge q stcoul moving with an acceleration of a cm/sec^2; $c = 3 \times 10^{10}$ cm/sec. The relative intensity of radiation in directions making an angle ϕ with the direction of the acceleration vector depends upon a factor $\sin^2 \phi$ and is plotted in Figure 35-11.

A steady current in a wire may be considered as being composed of a collection of charges moving with constant velocity. Thus no radiation results from direct current in a wire. Only when the current in the wire is changing is there radiation. Thus, whenever a switch is closed, or whenever there is an impulsive flow of charge, as in an automobile ignition system, we must expect electromagnetic radiation to take place. Indeed, this is the reason for the interference observed on radio and television sets caused by passing automobiles.

When a charge is oscillating in simple harmonic motion with angular frequency ω and amplitude A, the magnitude of the acceleration varies with the displacement in accordance with Equation (12-7).

$$a = \omega^2 A \cos \omega t.$$

Since the rate of radiation depends upon the instantaneous acceleration of the charged particle, the energy radiated from an oscillating charge must display the same time variation, hence must have the same frequency, as the charged particle. In addition we note that the magnitude of the acceleration varies with the amplitude A of the simple harmonic motion and with the square of the frequency ω^2. We must therefore expect a greater rate of radiation from an oscillating charged particle when that particle oscillates at higher frequencies.

A transmitting antenna consists of a wire or some other electrical conductor connected to an oscillator. The antenna may be thought of as a capacitor which becomes charged and discharged as the emf of the oscillator alternates. Electric charges flow to and from the antenna with the frequency of the oscillator. We may relate the behavior of the antenna as a radiator of electromagnetic waves to the radiation from oscillating charged particles.

With the restriction that the antenna is short compared to its wave length (to avoid the complication of standing waves), the longer the antenna, the greater the amplitude of oscillation of the electric charges whose motion constitutes the current. Therefore, a long antenna is a more effective radiator than a short one. Furthermore, the energy radiated depends upon the square of the frequency of oscillation. We must therefore expect the fraction of the energy to be radiated from a 60-cycle power line to be small under the same circumstances that the energy radiated from a 1-megacycle radio antenna is quite large.

The energy which is radiated leaves the antenna in the form of an electromagnetic wave, consisting of an oscillating electric field and an oscillating magnetic field which are coupled together. The changing magnetic field may be thought to generate an electric field in accordance with Faraday's law of induction. The changing electric field may be thought of as though it were generated by a virtual alternating current, called the

displacement current, which generates a magnetic field in accordance with Ampère's law. This exchange of energy between electric and magnetic fields constitutes a wave motion which is propagated with the velocity of light.

Practical antennae are not quite so simple as those we have discussed here, but any element of conductor carrying current may be considered as the source of the wave, and the radiation from the antenna as a whole may be calculated from Huygens' principle. In general, the reciprocal relation between radiation and absorption developed in our discussion of black-body radiation is equally applicable here. Thus an antenna which is an effective transmitter of electromagnetic radiation will also be effective as an absorber of radiation, hence a good receiving antenna.

The mechanism by which an antenna receives a signal transmitted by electromagnetic radiation is the inverse of the mechanism by which it transmits radiation. The passing electromagnetic field induces currents in the antenna which alternate with the frequency of the radiation. These currents flow through a resistor and are subsequently amplified and converted into a useful signal in the receiving set.

Like the transmission of light, the transmission of electromagnetic waves is along the line of sight. Radio communication is received at large distances from the transmitting antenna, at points obscured from the antenna by the curvature of the earth, by reflection from a region of ionized atmosphere known as the *ionosphere,* or the *Kennelly-Heaviside* layer, located at high altitudes. This layer of ionization behaves like an electrical conductor at some frequencies. Radiation at these frequencies may be reflected back and forth between the earth and the ionosphere, so that it is propagated halfway around the earth. The ionosphere is generated by the action of solar radiation upon the atmosphere. Solar flares and other solar disturbances affect the ionization and consequently affect radio transmission and reception at large distances from a transmitting antenna.

Problems

35-1. The peak current in a half-wave rectifier is 100 ma. Find the effective or rms current from this rectifier.

35-2. In the circuit of Figure 35-3(a), the applied 60-cycle alternating emf has a peak or maximum voltage of 150 volts. Find the peak values of (a) the current through the load resistor and (b) the voltage across it when it has a resistance of 500 ohms.

35-3. In the full-wave rectifier of Figure 35-4(a), the applied 60-cycle alternating emf is 300 volts rms across the entire secondary winding of the transformer. The load resistance $R_L = 500$ ohms. Find (a) the peak current through the load resistor and (b) the peak voltage across it.

35-4. Find the plate resistance, the amplification factor, and the mutual conductance of the triode of Figure 35-5 when the plate voltage is 200 volts and the grid bias is (a) -12 volts and (b) -8 volts.

35-5. Find the plate current in a 6C5 triode whose static characteristics are given in Figure 35-5 when its cathode is connected to the negative terminal of a 300-volt battery, its plate is connected to the anode of the battery through a 30,000-ohm resistor, and $\varepsilon_c = -4$ volts.

35-6. A triode amplifier has an amplification factor of 20 and a plate resistance of 10,000 ohms. What is the gain of the amplifier when the plate load resistor is 20,000 ohms? Draw the equivalent circuit of the amplifier.

35-7. The change in plate current of a triode may be expressed as the sum of two terms, for the change in plate current Δi_b may be due to the change in plate potential $\Delta\varepsilon_b$ and to the change in grid potential $\Delta\varepsilon_c$, as

$$\Delta i_b = \frac{1}{r_p}\,\Delta\varepsilon_b + g_m\,\Delta\varepsilon_c.$$

From the above expression prove that

$$g_m r_p = \mu.$$

35-8. A two-stage amplifier has a gain of 10^3. What is the gain of each stage, assuming these to be equal.

35-9. In a television set, a ghost pattern caused by reflection of the signal from a distant building is observed to follow the main pattern by 1 μsec. What is the difference in the length of the path from the transmitter to the tube between the direct signal and the reflected signal?

Part Five

LIGHT

36

Light and Its Measurement

36-1 The Nature of Light

The word *light*, as commonly used, refers to the radiant energy which produces a visual effect. It was first shown by Maxwell that light is an electromagnetic radiation, propagated with a speed whose value was numerically determined by the relationship between electric and magnetic units. As we have seen in Section 31-3, the velocity of light in vacuum, represented by the symbol c, is given by the equation

$$c = (\epsilon_0\mu_0)^{-\frac{1}{2}},$$

where c has the numerical value of 3×10^8 m/sec, or 186,000 mi/sec. Light waves are now understood to be electromagnetic waves of shorter wavelength and higher frequency than the electromagnetic waves used in radio.

If light is propagated as an electromagnetic wave, its origin must ultimately be traced to electric charges. Thus an analysis of the light emitted by a substance may be used to determine the electronic constitution of that substance, and so to give information about the structure of nuclei, of atoms, and of aggregates of atoms. The emission of light by a source involves an interaction between matter and radiant energy. To explain the phenomena associated with the interaction of matter and radiant energy, it has been necessary to supplement Maxwell's electromagnetic theory of light by what is known as the *quantum theory of light.* Radiant energy, in its interaction with matter, behaves as though it consists of bundles of energy. The amount of energy $\mathcal{E}$ in each bundle, or *quantum,* is related to the frequency f by the equation

$$\boxed{\mathcal{E} = hf,} \tag{36-1}$$

where h is a universal constant called *Planck's constant* whose value is

$$h = 6.62 \times 10^{-27} \text{ erg sec.}$$

A quantum of light energy hf is also called a *photon*. We shall return to the quantum theory later in discussing the analysis of light and atomic structure. For the present we shall concern ourselves with the propagation of radiant energy, and shall emphasize particularly that form which produces a visual effect.

The normal human eye is capable of responding to those waves whose wavelengths lie in the range between 3.8×10^{-5} cm and 7.6×10^{-5} cm.

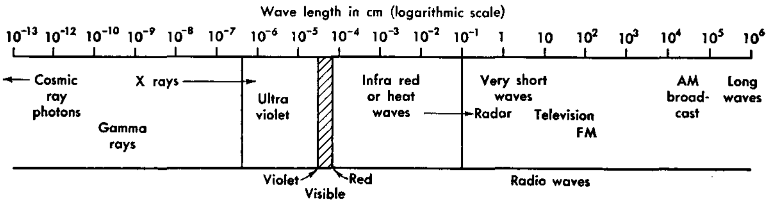


Fig. 36-1 The complete electromagnetic spectrum. Because of the wide range in wavelengths, the latter have been drawn to a logarithmic scale.

This is a very narrow region of the electromagnetic spectrum, corresponding roughly to that region of the solar spectrum which is transmitted by the earth's atmosphere. The electromagnetic spectrum is indicated in Figure 36-1, and ranges from waves of very short wavelength known as *x-rays* or *gamma rays*, to waves of very long wavelength used in radio broadcasting. The distinction between the terms "x-ray" and "gamma ray" lies in the origin of the radiation rather than in its wavelength; the term "gamma ray" is reserved for radiation emitted by the nucleus of an atom.

36-2 Sources of Light

Our principal source of light is the sun, but when sunlight is not available, other sources of light must be used. Before electric power became generally available, the sources of light widely used were the kerosene lamp, the candle, and illuminating gas. Light is emitted from these sources as a result of the chemical reactions in combustion. With the development of large sources of electric power, many new sources of light were developed to convert electric energy into light. Edison's incandescent electric light consisted of a carbon filament sealed into an evacuated glass bulb. The modern incandescent lamp contains a tungsten *filament* which may be heated to a temperature as high as 3000°C. The glass bulb is generally filled with an inert gas, such as argon, to retard evaporation of tungsten from the filament.

The *carbon arc* is frequently used as an intense source of light, as in searchlights or motion-picture projectors. The carbon arc consists of two carbon rods connected to a source of power through a resistor. The arc is started by momentarily touching the two carbon rods together. The rods are heated at the points of contact where the current density is very high. Some of the carbon vaporizes and becomes ionized, acting as the conducting

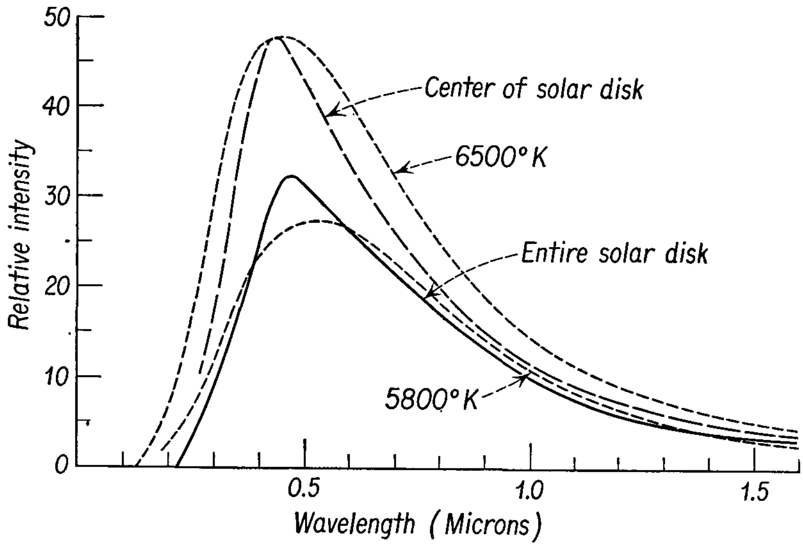


Fig. 36-2 Spectral distribution of solar radiation. The solid line shows the spectrum obtained from light from the entire solar disk, while the dashed line shows the spectrum of light from the center of the solar disk. Dotted lines indicate the spectrum of black body radiation at temperatures of 6500°K and 5800°K. (After R. N. Thomas, *The Scientific Monthly*, September, 1956.)

path for the current between the two electrodes. The positive terminal may reach a temperature of 3700°C, becoming white hot. This terminal is the principal source of light in the arc. Some carbon arcs, used primarily for spectroscopic or therapeutic purposes, have the cores of the positive electrodes filled with some material other than carbon. The light obtained from cored carbons is characteristic of the core material and is superimposed on the white light from the hot carbon.

Sunlight comes to us from a layer of the sun where the transition from opacity to transparency takes place. This layer is known as the *photosphere* and is at a temperature of about 6000°C. Sunlight is whiter than the light from either the tungsten lamp or the carbon arc. The hotter the source, the closer will the light it emits approach the color of sunlight. The mean-

ing of the temperature of the sun or of the photosphere is somewhat vague. One way of assigning a number to the temperature of the sun is to compare the solar spectrum to the radiation spectrum emitted by a black body. If light from the entire solar disk is admitted to the spectroscope, we find that the spectrum is in approximate agreement with a black-body spectrum at a temperature of 5800°K, while the maximum intensity of the light from the center of the solar disk corresponds to a temperature of 6500°K, as shown in Figure 36-2.

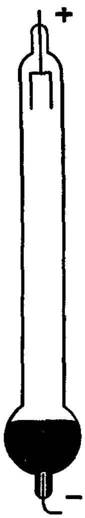


Fig. 36-3 Mercury arc.

Another important source of light is the *mercury arc.* In the low-pressure mercury arc a pool of mercury is placed at the bottom of a glass or quartz tube containing two metal electrodes sealed into the ends, as in Figure 36-3. The mercury is in contact with one electrode, and mercury vapor at a low pressure fills the rest of the space in the tube. To start the arc the tube is tilted so that liquid mercury makes momentary contact with both electrodes, and then the tube is returned to its vertical position. The electric discharge through the ionized mercury vapor is similar to the electric discharge through a gas at low pressure; the light emitted is rich in blue, green, and violet light and is characteristic of the mercury. A mercury arc also emits ultraviolet light; this is readily absorbed by ordinary glass but is transmitted by quartz.

The high-pressure mercury arc contains a small amount of liquid mercury which is completely vaporized by the heat developed by the passage of current through the tube. The pressure increases as the temperature of the vapor increases. Pressures of 50 to 100 atm and temperatures of 5000°K are common in these arcs. The high-pressure mercury arc is an intense source of illumination.

A *fluorescent* lamp is a low-pressure mercury arc in a long glass tube which has a coating of some fluorescent material on the inside surface. Some argon gas is also put into the tube to make it easier to start the lamp. Two filaments are built into the ends of the tube. When heated, these filaments emit electrons which collide with and ionize the gas within the tube, thus starting the arc. The construction of a typical fluorescent lamp is shown in Figure 36-4. A special starting arrangement must be used with fluorescent lamps, similar to the schematic diagram in Figure 36-5. When the lamp is switched on, alternating current flows through the two filaments in series, through the inductor L, and through the switch S. After a short time the switch S is opened, producing a large induced emf across the filaments which now act as electrodes and start the arc in the tube. The

starters built into fluorescent lamps automatically open and close the switch S through the use of a bimetallic strip.

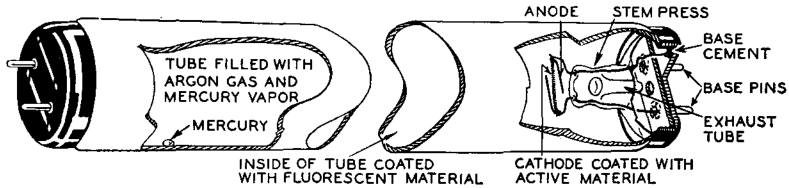


Fig. 36-4 Details of the construction of a fluorescent lamp. (Courtesy of Westinghouse Electric Company.)

There are essentially two sources of light in a fluorescent lamp. The mercury vapor emits green, blue, and violet light and, in addition, a large amount of ultraviolet radiation which is absorbed by the fluorescent coating

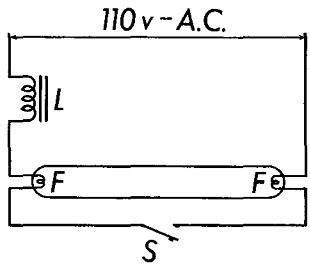


Fig. 36-5 Starting arrangement for a fluorescent lamp. S is the switch, F and F are filaments, and L is the inductance.

on the wall of the tube and is re-emitted as visible light. There are many substances which fluoresce under the action of ultraviolet light. These substances are called *phosphors*. By a suitable choice of phosphor, many different colors, such as red, orange, or yellow, can be added to the light from the mercury, so that the fluorescent lamp can have almost any desired color.

36-3 Electric Discharge through Gases

For the study of the passage of an electric current through a gas, let us put the gas into a very long glass tube which has a circular electrode sealed into each end, and which is provided with a small side tube which can be connected to a pumping system, as sketched in Figure 36-6.

Suppose that there is air in the tube, and that the two electrodes A and C are connected to the terminals of a source of high potential, say 50,000 volts. When the pressure of the air inside the tube is reduced to a few millimeters of mercury, the passage of electric current through the gas will be accompanied by the emission of light from the gas. At this pressure

the entire space between the electrodes will be filled with a pink or reddish glow. When the pressure of the air in the tube is reduced to about 0.1 mm, there is no longer a uniform glow between the electrodes but a series of dark and light regions, as shown in Figure 36-7. A bluish velvety glow, known as the *cathode glow*, covers the entire negative electrode, or cathode

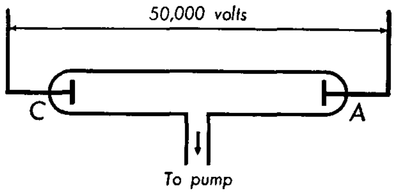


Fig. 36-6 A gas-discharge tube.

C. This is followed by a dark space called the *Crookes dark space* which ends at the *negative glow*. This is separated by the *Faraday dark space* from the luminous column, known as the *positive column*, which extends up to the anode. The positive column usually appears to be striated, consisting of a series of bright and dark regions, equally spaced. Covering the entire anode, or positive terminal *A*, is the *anode glow*.

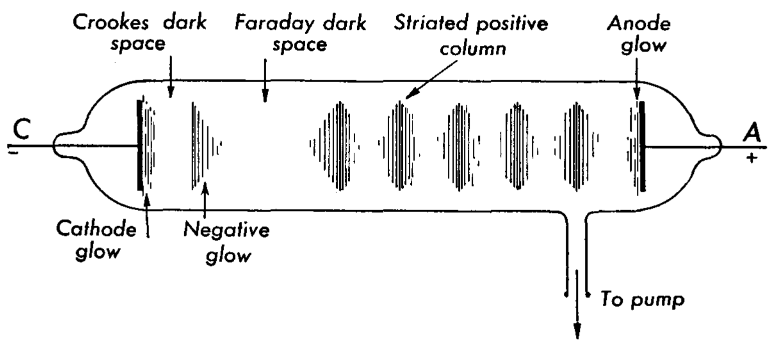


Fig. 36-7 Appearance of the electrical discharge when the pressure of the air in the tube is about 0.1 mm.

Experiments show that there are always some ions present in the air. These ions may have been produced by the action of light, or as the result of collisions between molecules, or by ionizing agents, known as *cosmic rays*, which are always present at the earth's surface. Whatever the original source of these ions may be, when a potential difference of several thousand volts is placed across the electrodes *A* and *C*, the charged ions will be accelerated toward the appropriate electrode. Let us call the average distance between collisions of molecules with each other by the term *mean free path*. If the pressure of the gas is in the neighborhood of 1 atm, the

ions will have very short mean free paths and will acquire very little energy before making a collision. But when the pressure is reduced to about 1 mm of mercury, the mean free path is longer, and the ion may acquire sufficient kinetic energy to ionize the molecule with which it collides. This process is known as *ionization by collision*. As the pressure is lowered, a greater number of collisions result in the production of ions and electrons in the gas. At the same time some of the positive and negative charges recombine to form neutral atoms and molecules. When the pressure of the gas in the tube becomes very low, the mean free path may exceed the length of the tube, so that an ion may travel the entire length of the tube without colliding with a second molecule. At such low pressures, in the neighborhood of 0.001 mm of mercury, the positive ions reaching the cathode have a great deal of energy. One of the results is that the cathode may emit electrons under the bombardment of the positive ions. These electrons are called *cathode rays;* they leave the cathode in a direction at right angles to its surface, since the cathode is an equipotential surface and the electric field is perpendicular to it.

An electric discharge can be maintained through any gas at a low pressure. The light emitted is characteristic of the gas and comes principally from the positive column. The spectrum is a line spectrum, rather than a continuous spectrum, with the energy concentrated in very narrow wavelength intervals which appear as lines in the spectrometer. The color of the discharge is the visual impression of the combination of the intensity and wavelengths of the spectral lines. Thus neon appears red, argon appears blue, and so forth. When examined on a spectroscope, the light yields information about the structure of the atoms and molecules of the gas. As sources of light, these gases are used chiefly for display purposes.

Another interesting aspect of electrical discharge in gases is the formation of *free radicals*, such as the hydroxyl radical OH, the methyl radical CH_3, and so on. Free radicals are highly reactive and play an important part in maintaining chemical chain reactions. Thus, if a chlorine molecule is split in two, the chlorine radical may attach a hydrogen molecule H_2 in its vicinity, forming stable HCl and releasing a hydrogen atom as a free radical. The hydrogen radical then combines with a second chlorine molecule Cl_2, and so on. Most free radicals are short-lived at ordinary temperatures, lasting about 10^{-3} sec, but some free hydrocarbon radicals have lifetimes of many days. Recent research has shown that it is possible to maintain normally short-lived free radicals at low temperatures by trapping them in a chamber cooled with liquid helium. This study is of great interest in the field of low-temperature chemistry. The free radicals liberated in a gaseous discharge may serve to form unusual chemical compounds in the discharge tube.

36-4 Color Sensitivity of the Eye

The problem of color vision embraces three distinct fields—physics, psychology, and physiology. Here we shall touch only upon a few important aspects of the subject. The normal eye is sensitive to a short range of wavelengths. It is customary to express wavelengths in Angstrom units (abbreviated A), where 1 A = 10^{-8} cm, and in microns, where 1 μ = 10^{-6} m = 10^{-4} cm = 10^4 A. The sensitivity of the normal human eye

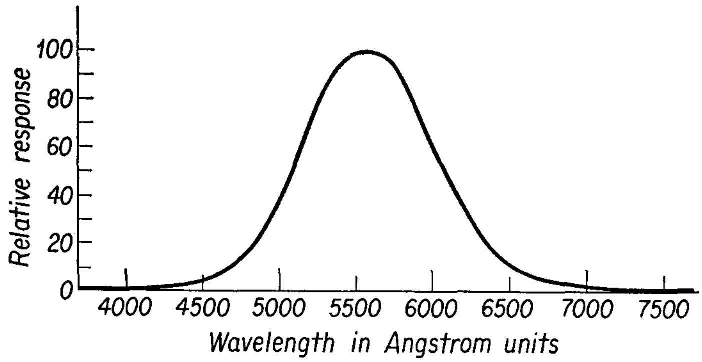


Fig. 36-8 Relative brightness sensitivity of the average human eye for daytime vision.

under normal daylight illumination is not constant over its range of sensitivity, ranging from 3,800 A to 7,600 A, as shown in Figure 36-8, but has a maximum response to light of 5,500 A, in the yellow-green region of the spectrum. Daytime sensitivity coincides approximately with the radiation maximum of the sun; this is an adaptation. In rod vision (see Section 39-3) the most sensitive wavelength is in the blue-green region. Purely nocturnal creatures have only rods; purely diurnal creatures have only cones.

A very narrow band of wavelengths constitutes a spectral color. The color identification ranges from violet for the shortest wavelengths to blue, green, yellow, orange, and red as the wavelength increases toward the maximum wavelengths visible. Light of any given color can be analyzed into its component wavelengths by means of a spectroscope, and the intensity at each wavelength, or within each narrow band of wavelengths, can be measured with an appropriate instrument. It must be emphasized that many problems within the field of vision are outside the realm of physics, for physics deals only with phenomena that may be observed by every observer. While we may agree that the colors on two different cards match, and we may agree to call them orange, it is impossible for any one of us to be certain that the sense impression he receives is the same as the sense impression anyone else receives from the same card.

36-5 Intensity and Brightness of Light Sources

We may express the intensity of a source of light in terms of the rate at which it emits radiation, in units of power, or energy emitted per unit time. Since all sources emit some radiation to which the eye is not sensitive, and since the eye is not equally sensitive to all wavelengths emitted by the source, a special unit must be used to express the rate at which a source emits *visible* radiation. Commercially obtainable lamps are presently rated in terms of the electrical energy consumed by the lamp rather than in terms of light output. A 60-watt lamp is one which consumes 60 joules of electrical energy in each second. Only a small fraction of this energy is converted into light. If the energy radiated by a given lamp is known as a function of wavelength, the energy radiated as light can be determined by applying weighting factors in each wavelength interval from Figure 36-8 to the lamp's spectral distribution. This is a cumbersome procedure and is not generally followed. Instead, a light source of special construction has been adopted as standard, by international agreement, and other sources are visually compared to the standard source.

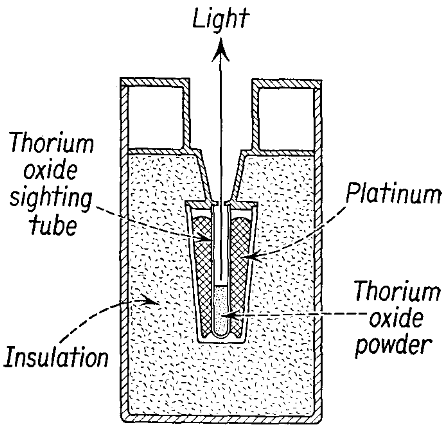


Fig. 36-9 Standard light source.

The present standard of light intensity is a black-body radiator at the temperature of molten platinum, at 2047°K, as shown in Figure 36-9. The unit of *luminous intensity* is called the *candela* (formerly called the *candle*), defined as the intensity of an opening in the standard source 1/60 cm^2 in area, when observed in a direction normal to the plane of the opening, along the axis of the cylindrical tube which makes up the black body. This body is used as a practical approximation to a point source of light.

The rate at which visible radiation is emitted by a source is called the *luminous flux,* expressed in *lumens.* The luminous flux from a source of radiation may not be the same in all directions, so that it is necessary to specify the luminous flux emitted into a narrow cone whose axis is in a particular direction, expressed in *lumens per steradian.* The steradian is a unit of measuring the *solid angle.*

To describe the solid angle subtended by a cone, in steradians, we construct a sphere of arbitrary radius r centered at the apex of the cone.

If the area of the sphere bounded by the cone is ΔA, the solid angle $\Delta\omega$ subtended by the cone is given by

$$\Delta\omega = \frac{\Delta A}{r^2}\text{ steradians.} \tag{36-2}$$

The definition of the steradian is analogous to the definition of the radian. Since the area of a sphere is $4\pi r^2$, the solid angle subtended by a sphere at its center is 4π steradians. As shown in Figure 36-10, if we wish to determine the solid angle subtended by a surface α at the point P, we draw a set of lines from P to the boundary of α, thus generating a conelike figure. From P as center, we describe a spherical surface of radius r and follow the procedure outlined above to find the solid angle subtended by this cone.

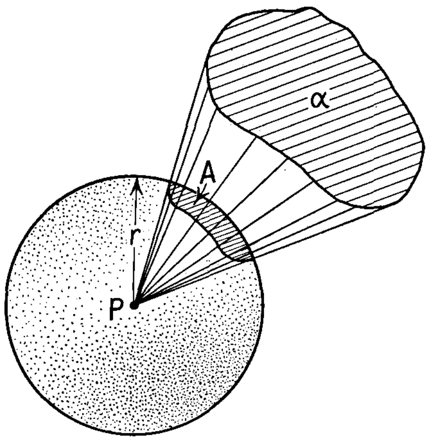


Fig. 36-10

A point source of light of luminous intensity I which radiates uniformly in all directions is said to radiate a luminous flux of $4\pi I$ lumens. The luminous intensity I of an arbitrary source, when viewed from a particular direction, may be determined by measuring the light flux ΔF, expressed in lumens, which passes through an element of solid angle $\Delta\omega$, expressed in steradians, by the equation

$$I = \frac{\Delta F}{\Delta\omega}\cdot \tag{36-3}$$

From Equation (36-3) we see that a luminous intensity of 1 candela is equivalent to a luminous flux of 1 lumen/steradian, and that an isotropic source of luminous intensity 1 candela radiates a total luminous flux of 4π lumens. From the definition of the standard light source, we see that the lumen is the luminous flux emitted from 1/60 cm^2 opening in the standard source, into a solid angle of 1 steradian.

Practical sources of light are not point sources but radiate light from an extended area, as in a fluorescent lamp or a tungsten filament. The *brightness* of an extended source is referred to technically as its *luminance,* and is expressed in units of luminous intensity per unit area, as *candelas per square meter*. Thus a source of unit luminance which is 1 m^2 in area has a luminous intensity of 1 candela. The luminous flux emitted normally from such a source is 1 lumen/steradian, when viewed from afar, so that the

source appears to be a point source. Many varied units of luminance are used. For example, the luminance of cathode-ray-tube screens is quoted by the manufacturer in units of foot lamberts, where a luminance of 1 foot lambert is equal to $1/\pi$ candela/ft^2.

Most diffuse radiating or reflecting objects appear equally bright visually from whatever direction they are viewed. Thus the sun and the full moon are seen as uniformly bright disks, in spite of the fact that they are spheres. This observation is formulated quantitatively as *Lambert's law* which states that the luminance, or brightness, B of an element of area ΔA in a direction making an angle θ with the normal to the element varies with the cosine of the angle θ. In other words, the radiant flux per unit solid angle radiated by the element of area varies with the angle in such a way that

$$\left(\frac{\Delta F}{\Delta\omega}\right)_\theta = \left(\frac{\Delta F}{\Delta\omega}\right)_0 \cos\theta. \tag{36-4}$$

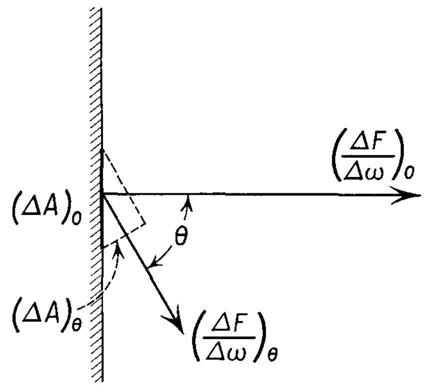


Fig. 36-11

From Figure 36-11 we see that the projected area $(\Delta A)_\theta$ of the source in the new direction is related to the actual area $(\Delta A)_0$ of the element by the equation

$$(\Delta A)_\theta = (\Delta A)_0 \cos\theta.$$

When we view the source from a direction making the angle θ with the normal, we calculate the luminance of the source by dividing the radiant flux per unit solid angle, as we see it, by the projected area of the source, as we see it. Thus the brightness at the angle θ, B_θ is given by

$$B_\theta = \frac{(\Delta F/\Delta\omega)_\theta}{(\Delta A)_\theta} = \frac{(\Delta F/\Delta\omega)_0 \cos\theta}{(\Delta A)_0 \cos\theta},$$

or

$$B_\theta = B_0;$$

that is, the apparent brightness of a source at an angle θ is equal to its brightness when viewed in the normal direction, provided that Lambert's law is obeyed. Thus the brightness of a diffuse source is independent of the direction from which it is viewed. Most commonly encountered light sources, such as the tungsten filament, the carbon arc, and so on, obey Lambert's law.

In optical projection systems the brightness of the source determines the brightness of the image. In efficiently designed optical systems, using a lamp of increased wattage does not appreciably increase the brightness

of the image, for the extra power consumption and increased light output are obtained by increasing the size of the filament rather than its brightness. It is for this reason that carbon arcs are used in theater projection systems.

Sources of illumination are rated as to their *luminous efficiency* in units of lumens per watt. The eye is most sensitive to light of wavelength 5,500 A. At this wavelength a lumen is equivalent to 0.00147 watt; that is, a source radiating 1 lumen of luminous flux at this wavelength radiates energy at the rate of 0.00147 joule/sec. From Figure 36-8 we note that the sensitivity of the eye at 5,000 A is only 0.4 of its sensitivity at 5,500 A, so that a source which radiates 1 lumen of luminous flux at 5,000 A radiates energy at the rate of 0.0037 watt. The brightness, luminous efficiency, and power consumption of a number of light sources are given in Table 36-1.

TABLE 36-1 POWER CONSUMPTION, LUMINANCE, AND EFFICIENCY OF LIGHT SOURCES

Source	Watts	Luminance (candelas/cm^2)	Luminous Efficiency (lumens/watt)
Tungsten lamp, gas filled	100	650	13
	1,000	1,100	19
Tungsten lamp, projection	100	1,800	23
	1,000	2,450	28
Carbon arc	1,500	15,000	16
	14,000	100,000	30
Fluorescent lamp (white)	40	0.6	58
Mercury arc			
10 atm	80	680	38
115 atm	1,000	48,000	60
Sodium vapor	45	9	55
Ideal monochromatic source (5,500 A)	1	—	680

36-6 Illuminance of an Illuminated Surface

The illumination of a surface is called the *illuminance E* and is the amount of luminous flux per unit area of the surface. Let us suppose that a point source of luminous intensity I is located a distance r from a surface of area ΔA, as shown in Figure 36-12, and that the line joining the point P to ΔA makes an angle θ with the normal to the area element. The solid angle $\Delta\omega$ subtended by ΔA at P is given by

$$\Delta\omega = \frac{\Delta A \cos\theta}{r^2}.$$

From Equation (36-3) the luminous flux ΔF radiated into a solid angle $\Delta\omega$

by a source of luminous intensity I is

$$\Delta F = I\,\Delta\omega.$$

Thus the luminous flux incident upon the area ΔA is

$$\Delta F = I\frac{\Delta A\cos\theta}{r^2},$$

and the illuminance of the surface is

$$E = \frac{\Delta F}{\Delta A} = \frac{I\cos\theta}{r^2}. \qquad \textbf{(36-5)}$$

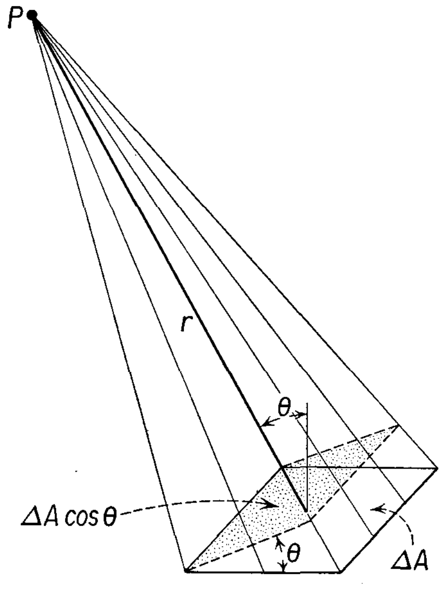


Fig. 36-12

The illuminance of a surface which is perpendicular to the line of sight, that is, a surface for which $\theta = 0°$, will vary inversely with the square of the distance from the point source. The units of illuminance are lumens per square meter, termed *lux*. The British unit of illuminance is the lumen per square foot, formerly called the *foot candle*, for it is the illuminance of a surface located at a distance of 1 ft from a 1-candle source. Similarly, the lux was sometimes called the *meter candle*.

Proper illumination has come to be considered essential not only for comfort and enjoyment but also for safety and efficiency. For example, it is now considered desirable to have an illuminance of 10 lumens/ft^2 on library reading desks, and 75 lumens/ft^2 in hospital operating rooms. Although these values may be considered bright for artificial light, they are quite small when compared to the illumination provided naturally on an overcast day, when the illuminance is approximately 1,000 lumens/ft^2.

36-7 Photometry

The measurement of illuminance and luminous intensity is called *photometry*, and instruments used for such measurements are called *photometers*. A Bunsen photometer is simple in design and illustrates the principle used in many photometers for determining the intensity of a source of light. The Bunsen photometer has a screen S which consists of a white sheet of paper with a translucent grease spot G in its center, as shown in Figure 36-13. This screen is mounted in a box containing two mirrors MM, so arranged that both sides of the screen can be viewed simultaneously. The box is

mounted on an optical bench, and the two sources which are to be compared are mounted on the bench on opposite sides of the photometer. Most of the light which strikes the paper from each source is diffusely reflected, while most of the light which strikes the grease spot is transmitted through

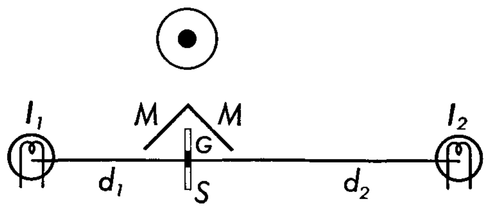


Fig. 36-13 Bunsen grease-spot photometer.

it. The experimental procedure is to move the screen S until the grease spot cannot be distinguished from the rest of the screen. In this case the illuminance of the photometer screen due to the source I_1 located a distance d_1 from the photometer is equal to the illuminance of the screen due to the source I_2 located a distance d_2 from the screen. From Equation (36-5) we have

$$\frac{I_1}{d_1^2} = \frac{I_2}{d_2^2}. \tag{36-6}$$

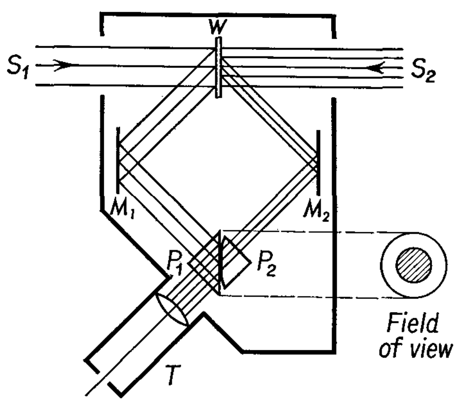


Fig. 36-14 Diagram of a Lummer-Brodhun photometer head.

If I_2 is the luminous intensity of a standard lamp, the luminous intensity of any other lamp can be determined with this photometer.

The Lummer-Brodhun photometer is an improvement on the Bunsen photometer, in which the grease spot is replaced by the photometer head shown in Figure 36-14. Light from the two sources S_1 and S_2, of intensities I_1 and I_2, respectively, falls on a diffuse white surface W. Some of the light which is scattered by this surface is reflected from the mirrors M_1 and M_2 into two right-angled prisms P_1 and P_2. Part of the base of prism P_2 is ground away so that only that portion of the light which strikes the surface of contact passes into the telescope T. The light striking the base of P_1 from M_1 is totally reflected into the telescope from the part of P_1 which is not in contact with P_2. The field of view seen through the telescope is sharply divided into two sections,

one brighter than the other, except when their illuminations are equal, in which case the dividing line disappears.

Difficulties are encountered in visual photometry when two sources of different color, or two lamps at different temperatures, are compared. In this event a device called a *flicker photometer* is used, in which opposite sides of the photometer head are alternately illuminated. At sufficiently high

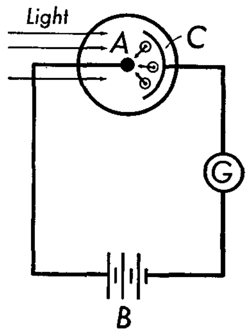


Fig. 36-15 A photoelectric tube used for measuring illuminance.

frequencies of alternation, the difference in appearance of the two sections of the photometer field which are due to color seems to disappear, and differences in intensity may be compared.

Besides the photometers which depend upon the visual matching of illuminated surfaces, other photometers are used which are based on some electrical effect resulting when light is incident upon a surface. In one type of *photoelectric effect,* light incident upon a metallic surface causes the ejection of electrons from the surface. When light falls on the cathode C of the photoelectric tube, shown connected to a simple circuit in Figure 36-15, electrons are ejected from the surface and are attracted to the anode A, because of the difference in potential between anode and cathode. The number of electrons emitted per second, and hence the current through the galvanometer, is proportional to the intensity of the incident light. The relative sensitivity of the surface to various wavelengths is different from the relative sensitivity of the eye, but, with the use of proper filters, such photocells can be used to measure illuminance. Filters have also been developed to match the sensitivity of photographic films for use of photoelectric cells as exposure meters. By use of appropriate filters, photoelectric cells have been used industrially to match color as well as intensity.

Another type of photoelectric effect which is widely used is the *photovoltaic effect,* in which a difference of potential is developed in an illuminated cell. For example, the Weston photronic cell consists of a very thin film of selenium formed on an iron plate. Light incident upon the selenium passes through it to the iron and produces an emf between the two substances, with the iron as positive terminal and the selenium as negative terminal.

The current through a galvanometer of internal resistance less than about 100 ohms is proportional to the illuminance of the surface. Such cells are widely used in photographic exposure meters, and similar cells are used in the newly developed *solar battery* which has recently been placed in service to charge storage batteries which power isolated telephone lines.

36-8 Speed of Light

The earliest determinations of the speed of light were based on astronomical observations. In 1675, Roemer, a Danish astronomer, observed that the time of occurrence of the eclipse of one of Jupiter's moons differed from that calculated on the assumption that light travels with infinite speed. He explained this discrepancy by assuming a finite speed of light and calculated this speed with the data then available. Referring to Figure 36-16, suppose that the time of occurrence of an eclipse of one of Jupiter's moons is observed when the earth is at E_1, and a calculation is then made as to the time of occurrence of an eclipse of this same moon of Jupiter 6 months later, when the earth is at E_2. If this calculation, which takes into consideration the motion of Jupiter as well as the earth, is made on the assumption that the speed of light is infinite, observation will show that the eclipse occurs about 1,000 sec later than the calculated time. The diameter of the earth's orbit around the sun is about 186,000,000 mi. If the difference between the observed time and the calculated time is ascribed to the finite speed of light, we find the speed of light to be about 186,000 mi/sec.

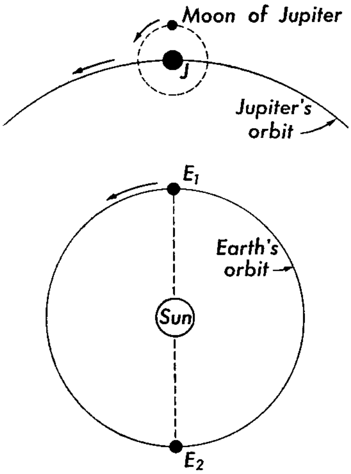


Fig. 36-16 Roemer's method of determining the speed of light.

One of the earliest terrestrial determinations of the speed of light was made in 1849 by Fizeau, who timed the passage of a beam of light a distance of 8.633 km from a source to a mirror and then back to the source. On its way from the source the light passed through the space between two teeth on a wheel whose speed was adjusted so that, on its return, the light failed to pass through this space but hit the adjacent tooth and was eclipsed. In this experiment the wheel had 720 teeth and the light was eclipsed when the speed was 12.6 rps. In 1850, Focault measured the speed of light using a rotating mirror instead of a toothed wheel.

Some of the best determinations of the speed of light were made by Albert A. Michelson (1852–1931), who began his experiments in about 1878 and continued them for about 50 years. In one of Michelson's experiments, a beam of light was sent from Mt. Wilson to Mt. San Antonio and back again, a distance of about 22 miles, measured very accurately by the U. S. Coast and Geodetic Survey. The essential arrangement for this experiment is shown in Figure 36-17. An octagonal mirror M_1 is mounted on the shaft of a variable-speed motor so that it rotates about an axis through its center. Light from a source S strikes mirror M_1 at an angle of

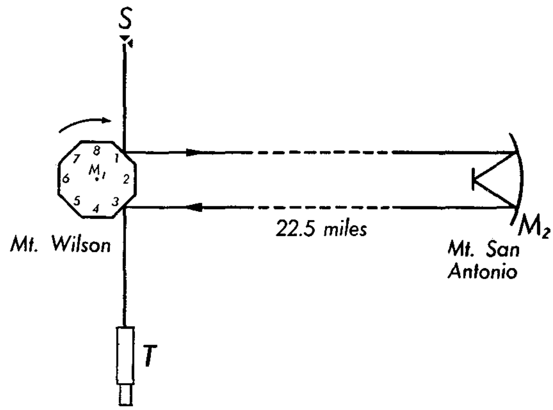


Fig. 36-17 Michelson's method for measuring the speed of light.

45° and is reflected from it to the distant mirror M_2 and back again, in such a way that when the octagonal mirror is stationary, the light is reflected from face 3 into the telescope. When the mirror is rotating, the image of the source is seen in the telescope when the time required for the light to traverse the path is equal to the time for face 2 to rotate into the position formerly occupied by face 3. In these experiments the angular speed of the mirror was about 530 rps. A measurement of the distance and the speed of the motor suffices to determine the velocity of light.

The speed of light in a material medium is less than its speed in vacuum. Focault placed a tube of water between the two mirrors and found that the speed of light in water is less than that in vacuum. Michelson used a tube 3 ft long and found the speed of light in carbon disulphide to be about four sevenths that of the speed of light in vacuum.

The presently accepted value for the velocity of light in vacuum is $c = 2.997923 \pm 0.000008 \times 10^{10}$ cm/sec. This is one of the most important constants of modern physics. Present-day methods for measuring the velocity of light use electronic devices rather than rotating mirrors to produce and detect very short bursts of light, but the basic concept in many of these experiments remains the same as that of Fizeau. In other methods

the velocity of light is inferred from the equation

$$c = \lambda f$$

and the measurement of λ and f of an electromagnetic wave. The precise knowledge of the velocity of light is used in radar, in which the distance of any object such as an airplane from a radar antenna is determined by measuring the time for a pulse of electromagnetic radiation to return to the antenna after reflection from the object. A very precise surveying instrument called the *geodimeter* is capable of measuring distances to an accuracy of 1 part in 1,000,000 by measuring the time interval between the radiation of a light beam modulated at a frequency of 10^7 cycles/sec and its detection after reflection from a distant mirror.

It is one of the cornerstones of Einstein's special theory of relativity that the velocity of light must have a constant value when measured by any observer, regardless of his velocity of translation. Near the turn of the last century, a number of experiments were conducted by Michelson, Morley, and many others in attempts to measure a difference in the velocity of light in the direction of the earth's motion and transverse to that motion. It was reasoned that light was propagated in a universal medium, called the *ether*, and that the velocity of light was a velocity with respect to the ether. In this case the velocity of light with respect to the earth would depend upon the velocity of the earth with respect to the ether. The velocity of light was found to be independent of its direction of propagation with respect to the earth or with respect to the stars. Einstein proposed his theory of relativity in an attempt to explain this result. If it were possible for an observer to find that the velocity of light, as he measured it, depended upon his own velocity, then the observer could determine his velocity simply by carrying out a measurement of the velocity of light, say in an airplane. Such a result would imply that motion was absolute rather than relative. To date, no such result has been obtained, and we must infer that all motion is relative motion.

Problems

36-1. What is the solid angle subtended by a sphere 1 cm in radius at a distance of 5 m?

36-2. A distant object of circular outline subtends an angle of 1°. (a) What is the solid angle subtended by the object? (b) What is the cross-sectional area of the object if it is 3 m distant?

36-3. A sodium lamp emits essentially monochromatic light of wavelength 5,890 A. (a) What is the energy, in ergs, of one quantum of this radiation? (b) If the radiant power emitted by the lamp is 0.1 watt, what is the number of quanta radiated per second?

36-4. A point source of light whose luminous intensity is 2 candelas is placed 3 m above a horizontal table. The table is provided with a square hole of edge length 0.5 cm, whose center is 5 m from the light source. (a) Find the solid angle subtended by the hole at the light source. (b) Find the luminous flux passing through the hole.

36-5. A 25-watt tungsten-filament electric lamp operates at a luminous efficiency of 10.4 lumens/watt. (a) How many lumens of luminous flux does it radiate? (b) What is its efficiency in the production of light? (c) If we are sufficiently far away from the lamp that it may be approximated as a point source, what is its luminous intensity? (d) Assume that the bulb is spherical, 5 cm in diameter, and is perfectly diffusing, and neglect the area occupied by the base of the bulb. What is its luminance?

36-6. The bulb of Problem 36-5 is hung 4 ft above a desk top. What is the illuminance on the desk top at a point (a) directly beneath the lamp and (b) 5 ft from the lamp?

36-7. A lamp emits 100 lumens of luminous flux at a wavelength of 6,500 A. (a) What is the radiant power, in watts, emitted by the lamp? (b) The lamp is supplied with 500 watts of electrical power. What is its luminous efficiency?

36-8. Two sources of light are mounted 2 m apart on an optical bench, and a grease-spot photometer head is placed between them. One source is rated as 41.6 candelas. Determine the luminous intensity of the second source if the grease spot is observed to vanish when at a distance of 120 cm from it.

36-9. A lamp whose intensity is to be measured is placed 200 cm from a standard lamp rated at 450 lumens. Equal illuminances are produced by the two lamps when the Lummer-Brodhun photometer head is 60 cm from the standard source. Determine the luminous intensity of the test lamp.

36-10. Two lamps of luminous intensity 40 candelas and 60 candelas, respectively, are placed at opposite ends of a photometer bench 200 cm long. (a) Where will the photometer have to be placed to show equal illuminance from these two lamps? (b) Determine this illuminance.

36-11. A workbench 6 ft long and 3 ft wide is illuminated by two 100-watt tungsten lamps hung 3 ft above the center line of the bench, each lamp being 2 ft from the end of the bench, at the opposite ends. Compute the illuminance of the bench at a point (a) directly under one lamp, (b) at one corner of the bench, and (c) at the center of the bench.

36-12. The distance between Mt. Wilson and Mt. San Antonio is 35.426 km, and the angular speed of the motor in one of Michelson's experiments was 528.76 rps. Calculate the speed of light from these data.

36-13. If the price of electric power is 8 cents/kw hr, what is the daily cost of illuminating a room 20 ft long by 15 ft wide by 8 ft high with an average illumination of 10 lumens/ft^2 with (a) 100-watt tungsten lamps and (b) 40-watt fluorescent lamps? Assume that the room is illuminated 8 hrs a day, and that there is no reflection of light from the walls.

37

Reflection and Refraction

37-1 Passage of Light through a Medium

In the passage of a beam of light through a medium, some of the radiant energy is absorbed and is transformed into internal energy, while some of it is scattered in all directions. The oscillating electric field associated with the light wave sets some of the electrons of the medium into oscillation, thus giving up some of its energy, and these oscillating electrons subsequently reradiate energy as scattered electromagnetic radiation. Scattering therefore takes place only in the presence of matter. The color of the sky is due to the small amount of scattering of sunlight by the molecules of the air. At high altitudes the number of scattering particles diminishes, and the sky is darker. In the stratosphere the sky appears almost black. The molecules of the air are more effective in scattering short wavelengths, so that the scattered light appears blue. Since the shorter wavelengths are scattered from the direct beam, the beam appears redder as it passes through larger thicknesses of air. It is for this reason that the setting sun looks redder than the noonday sun. The scattering of light by the air is responsible for twilight. In microwave communications the distance between transmitting and receiving antennae is limited by the curvature of the earth, because of the straight-line propagation of electromagnetic waves. By transmitting a stronger initial signal and detecting the scattered microwaves, it has been possible to reduce the number of transmitting and receiving stations in the microwave communications network by increasing the distance between relay stations beyond the line of sight.

If a beam of light of luminous flux F is incident upon a slab of matter of thickness Δx, some of the incident flux is absorbed in the substance as heat, and some is scattered out of the incident beam, as shown in Figure 37-1. The luminous flux reaching the detector is diminished by an amount ΔF in such a way that the fractional loss in luminous flux is proportional

to the thickness of the slab. Thus we have

$$\frac{\Delta F}{F} = -\mu\, \Delta x, \tag{37-1}$$

where μ is a constant of proportionality called the *absorption coefficient.* The minus sign appearing in the equation is associated with the fact that a positive increment in thickness is accompanied by a negative increment

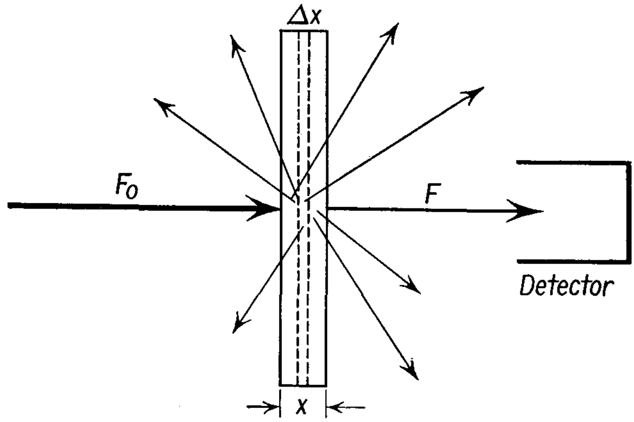


Fig. 37-1

in luminous flux. When a beam of incident flux F_0 is incident upon a slab of thickness x, the emergent beam is of flux F, whose value may be obtained by integrating Equation (37-1). We have

$$\int \frac{dF}{F} = -\int_0^x \mu\, dx,$$

or

$$\ln \frac{F}{F_0} = -\mu x,$$

yielding

$$F = F_0 e^{-\mu x}. \tag{37-2}$$

Thus the luminous flux in a beam of radiation is exponentially absorbed on its passage through matter. The absorption coefficient μ depends upon the wavelength λ and upon the nature of the material through which the radiation is passing.

When a beam of light strikes the surface separating one medium from another—for example, the surface between air and glass—some of the light is reflected back into the first medium at the surface of separation, and the remainder enters the second medium. The light which passes from one medium to another is said to be *refracted.* If the surface of separation

between the two media is smooth, the light which is reflected back into the first medium is said to be *regularly reflected* or *specularly reflected.* If the surface is rough, the light is *diffusely reflected.* Unless otherwise stated, we shall assume that the surface between two media is smooth.

In general, smooth polished metal surfaces will reflect about 90 per cent of the incident light, while smooth polished glass surfaces will reflect about 4 to 10 per cent for angles of incidence from 0° to 60°. In the case of a metal, the light which is refracted through the surface is absorbed in a very small thickness of the surface; that is, the absorption coefficient of a metal is very high. Glass, on the other hand, has a very low absorption coefficient for visible light; that is, very little light is absorbed in its passage through reasonable thicknesses of glass.

The electrical conductivity, the permittivity, and the permeability of a substance all affect the transmission of light through that substance. These properties depend upon the frequency, so that substances which are good conductors at one frequency may be poor conductors at a much higher frequency. As a general statement, electrical conductors transmit light poorly. An optically transparent substance cannot be a good conductor of electricity. Because of the variation of electrical properties with frequency, we find that materials which are transparent to visible light and to radio waves may be opaque to infrared or ultraviolet radiation. While all metals are opaque to visible light, it is possible to transmit a beam of x-rays through them. Sea water, which is a relatively poor conductor, reflects about 80 per cent of the normally incident light. It is possible to coat the surface of glass with a thin layer of silver to make a mirror which will reflect about half the incident light and will transmit the other half. In such cases the layer is very thin, and its electrical resistance is high.

Although reflection and refraction can be studied by considering the detailed interaction between the light and the individual particles of the medium through which it travels, we may obtain useful results by assuming that light travels along straight lines through optically homogeneous media, the direction of motion being indicated by *rays* of light. These rays are drawn at right angles to the *wave front.* The wave front is the envelope of the waves emitted by the particles of the medium; *all points on a wave front are in the same phase.* This method of treating reflection and refraction is a first approximation only. Later, in the discussion of *interference* and *diffraction,* we shall consider some of the modifications that must be made because of the wave character of light. The simpler treatment is called *geometric optics.*

37-2 Laws of Reflection

When a narrow beam of light strikes a smooth surface separating two media, the angles of incidence, reflection, and refraction are all measured

from a *normal* to the surface; a normal to a surface at a given point is a line drawn perpendicular to the surface at that point. In Figure 37-2 NP is the normal to the surface at P, AP is the incident ray, PB is the reflected ray, and PC is the refracted ray. The angle of incidence i is the angle between the incident ray and the normal; the angle of reflection r is the angle between the reflected ray and the normal; the angle of refraction r' is the angle between the refracted ray and the normal.

The two *laws of reflection* may be stated as follows: (a) *The incident ray, the normal, and the reflected ray all lie in one plane,* and (b) *the angle of incidence is equal to the angle of reflection.* The laws of reflection are empirical laws and have been known since the tenth century. As we

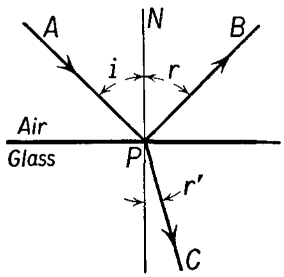


Fig. 37-2 Light reflected and refracted at a surface separating two media.

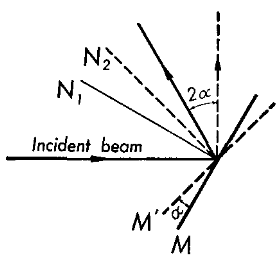


Fig. 37-3 The angle through which the reflected beam rotates is twice the angle through which the mirror rotates.

have seen in Section 20-4, the equality of the angles of incidence and reflection may be derived from Huygens' principle for an incident plane wave. If a wave front is spherical, we can take a sufficiently small portion of the wave and treat it as a plane wave.

Rotating mirrors are often used in physical apparatus to provide an instrument with a long weightless pointer, as in the wall galvanometer. When the mirror turns through a small angle α, the angle of incidence is increased (or decreased) by α so that the angle of reflection is also increased (or decreased) by α. As shown in Figure 37-3, we observe that when a plane mirror, originally in position M, is rotated through an angle α, the reflected beam will rotate through an angle 2α.

An observer whose eye is directed toward the reflecting surface of a mirror, Figure 37-4, sees rays of light reflected from the mirror. The eye observes the light rays QC and $Q'C'$, and the observer interprets these rays as coming from the point P' behind the mirror. If the observer's head is turned toward the point P, the rays QP and $Q'P$ enter his eye, and he interprets these rays as originating at the point P. If the mirror is perfectly reflecting, the observer sees the real source at P and the virtual

source at P' as sources of equal brightness. The point P' is called the image point of P in the mirror. Since the rays QC and $Q'C'$ do not actually come from P' but only appear to do so, the point P' is called a *virtual image* of the point P in the plane mirror. *A plane mirror forms a virtual image of a real object, and the image is located as far behind the mirror as the object is in front of the mirror.*

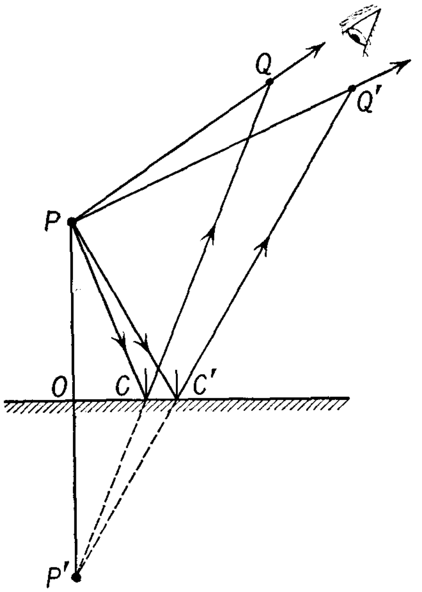


Fig. 37-4 Determining the position of the image formed by a plane mirror.

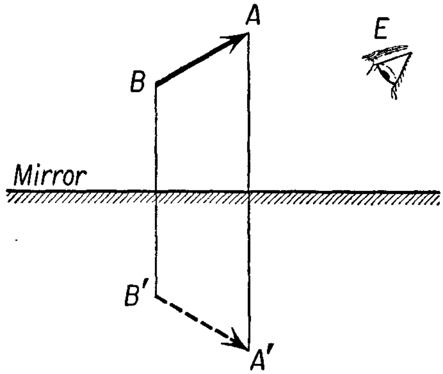


Fig. 37-5 The image formed by a plane mirror appears reversed to an observer at E.

If an object of finite size, such as the arrow BA of Figure 37-5, is placed in front of a plane mirror, each point on the object is imaged behind the mirror. Thus the point B is imaged in the point B', the point A is imaged in the point A', and so on. An observer looking into the mirror sees the virtual image $B'A'$ of the arrow BA, in the position indicated on the figure, no matter from what position he observes the image. An observer at E sees the head of the real arrow AB on his right, while the head of the imaged arrow $A'B'$ appears on his left. The image is said to be *reversed* from right to left. *The size of the image formed in a plane mirror is the same as the size of the object.* The image is formed in a region behind the mirror in which there is no light, so that it is impossible to form the image on a piece of paper or a ground glass. Nevertheless, the eye has interpreted the rays of light reflected from the mirror as though there really were a source of light behind the mirror.

37-3 Refraction of Light

Refraction occurs whenever light passes from one medium to another and is due to the fact that light travels with different speeds in different media. Refraction was studied, and the laws of refraction were derived empirically, long before the cause of refraction was known. Experimental observations on the refraction of light at a plane interface may be summarized in the following *laws of refraction:* (a) *The incident ray, the normal, and the refracted ray are in the same plane, and* (b) *for monochromatic light, the ratio of the sine of the angle of incidence to the sine of the angle of refraction is a constant.* In the form of an equation we have

$$\frac{\sin i}{\sin r'} = n_r, \tag{37-3}$$

in which i is the angle of incidence, r' is the angle of refraction, and n_r is a constant known as the *relative index of refraction* of the second medium with respect to the first medium. Equation (37-3) is often referred to as *Snell's law,* after Willebrord Snell (1591–1626), who is thought to have discovered this empirical relationship in 1621.

We have shown in Section 20-7 that the existence of a relative index of refraction can be derived from Huygens' principle, a derivation first presented by Huygens to the Royal Academy of Sciences at Paris in 1678, and that the relative index of refraction is equal to the ratio of the velocities of light in the two media, as in Equation (20-5)

$$n_r = \frac{v_1}{v_2}, \tag{37-4}$$

in which v_1 is the velocity of light in the first medium, and v_2 is the velocity of light in the second medium.

We may define the *absolute index of refraction* n of a medium as the index obtained when light passes from vacuum into the medium. From Equation (37-4) we find that

$$\boxed{n = \frac{c}{v},} \tag{37-5}$$

where c is the velocity of light in vacuum, and v is the velocity of light in the medium. The absolute index of refraction n is customarily found in tables. Since the velocity of light in a medium is dependent upon its frequency, and therefore upon its wavelength, the absolute index of refraction varies with the wavelength, as shown in Table 37-1 for several

TABLE 37-1 ABSOLUTE INDICES OF REFRACTION

Substance	Absolute Indices of Refraction				
Wavelength	7,682 A	6,563 A	5,893 A	4,861 A	4,047 A
Borosilicate crown glass	1.5191	1.5219	1.5243	1.5301	1.5382
Dense flint glass	1.6441	1.6501	1.6555	1.6691	1.6901
Water 20°C	1.3289	1.3311	1.3330	1.3371	
Carbon disulphide 18°C		1.6198	1.6255	1.6541	
Diamond			2.417		

substances. From Equation (37-5) the absolute index of refraction for a medium 1 is given by $n_1 = c/v_1$, while for a second medium, 2, the absolute index of refraction is given by $n_2 = c/v_2$. Dividing the second of these equations by the first, we find

$$\frac{n_2}{n_1} = \frac{c/v_2}{c/v_1} = \frac{v_1}{v_2},$$

so that

$$n_r = \frac{n_2}{n_1}. \tag{37-6}$$

If we relabel the angle of incidence ϕ_1 and the angle of refraction ϕ_2, we may rewrite Snell's law, as given in Equation (37-3), in a more convenient form as

$$\boxed{n_1 \sin \phi_1 = n_2 \sin \phi_2} \tag{37-7}$$

where we may interpret ϕ_1 as the angle the ray makes with the normal in medium 1, and ϕ_2 as the angle the ray makes with the normal in medium 2. From the symmetry of Equation (37-7), we see that the direction of the light beam may be reversed without altering the path taken by the light beam. Thus a beam which makes an angle ϕ_1 with the normal in medium 1 will make an angle ϕ_2 with the normal in medium 2, regardless of whether it passes from the first to the second medium or from the second medium to the first.

A medium whose absolute index of refraction is high is sometimes called an *optically dense medium*, while a medium whose index of refraction is low is called an *optically rare medium*. Optical density bears no relation to actual physical density for different substances. The velocity v of light in a medium is related to its electromagnetic properties through the equation

$$v = (\epsilon\mu)^{-1/2}. \tag{37-8}$$

As we have already seen in Sections 25-6 and 29-8, both the permittivity ϵ and the permeability μ of a medium depend upon the number of molecules per unit volume of a medium. If there are no molecules present, there can be no induced dipole moment. The induced electric and magnetic dipole moments per unit volume vary directly with the number of molecules per unit volume, provided that the molecules are far enough apart so that there is no appreciable interaction between them, as in the case of a gas. Thus we must expect both ϵ and μ to increase with the density of a gas, and consequently we must expect the velocity of light in a gas to decrease with increasing density. A gas whose physical density is relatively high has a correspondingly high index of refraction and is therefore an optically denser medium than the same gas at low density.

37-4 Refraction Effects

When a beam of light from a distant source passes through a transparent medium with parallel surfaces, the beam is displaced parallel to itself, but its direction is unaltered. The beam is *displaced* but is undeviated, as

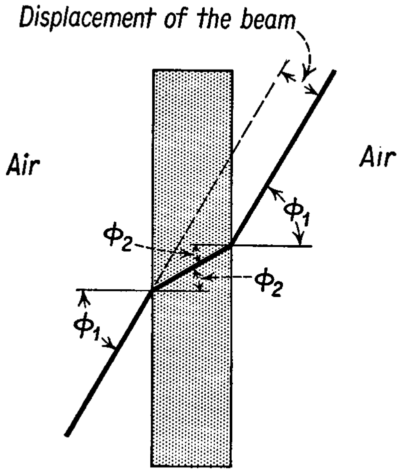


Fig. 37-6 Displacement of a ray of light by a transparent plate with parallel surfaces.

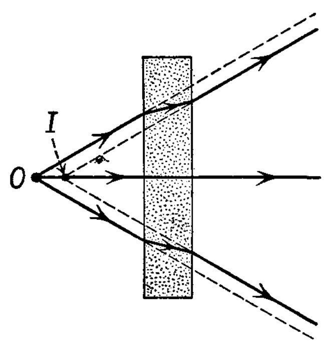


Fig. 37-7 The point of divergence of a beam of light appears displaced after passing through a plate of glass with parallel faces.

shown in Figure 37-6. A beam of light from a source close to the transparent plate must be treated as a divergent beam, made up of spherical waves. Each ray striking the plate is displaced a different amount. If the divergence is not too large, the rays seem to come from a point I, Figure 37-7, displaced from the source of light O by an amount which

depends upon the thickness of the glass. By reversing the directions of the rays in this diagram, we note that the effect of a parallel plate of glass on a bundle of rays converging to a point, or to a focus, is to shift the point of convergence away from the plate of glass. This method has been used to change the focal length of camera lenses.

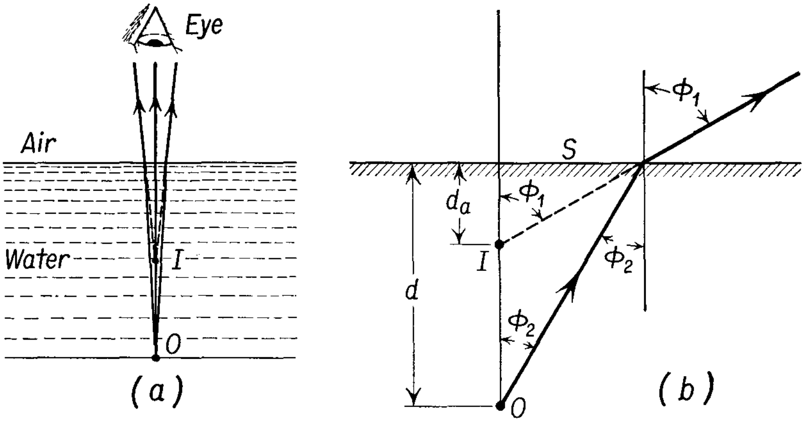


Fig. 37-8 Shallowing effect produced by refraction.

Another interesting effect produced by the refraction of a narrow bundle of rays is illustrated in Figure 37-8(a). Rays from some point O at the bottom of a pool of water are viewed by an observer whose eye is vertically above this point in the air. The narrow bundle of rays entering the eye appears to come from a point I above point O. The pool seems shallower. Redrawing the path of one ray on an exaggerated scale in Figure 37-8(b), we note that

$$n_1 \sin \phi_1 = n_2 \sin \phi_2,$$

and, if the angles ϕ_1 and ϕ_2 are sufficiently small, we have

$$n_1 \tan \phi_1 = n_2 \tan \phi_2.$$

From the figure we note that $\tan \phi_1 = s/d_a$, and $\tan \phi_2 = s/d$, so that

$$n_1/d_a = n_2/d,$$

and since medium 1 is air with $n_1 = 1$, we may write n for n_2, dropping the subscript 2 to find

$$d_a = d/n. \tag{37-9}$$

Thus the apparent depth d_a of a transparent medium is its actual depth divided by the index of refraction of the medium. This method may be used to measure the refractive index of a glass microscope slide whose thickness is measured with a caliper, and whose apparent thickness is

measured by determining the distance the microscope must be lowered when it is first focused on the top surface of the slide and them focused on the bottom surface of the slide.

37-5 Critical Angle. Total Reflection

A ray of light traveling from a region of high refractive index n_2 to a region of low refractive index n_1 may penetrate the surface of separation and enter the medium 1 only if the angle of incidence ϕ_2 is less than a certain angle ϕ_c, called the *critical angle.* If the angle of incidence is greater than the critical angle, the light will not be able to penetrate the surface but will be totally reflected back into medium 2. From Snell's law

$$n_1 \sin \phi_1 = n_2 \sin \phi_2,$$

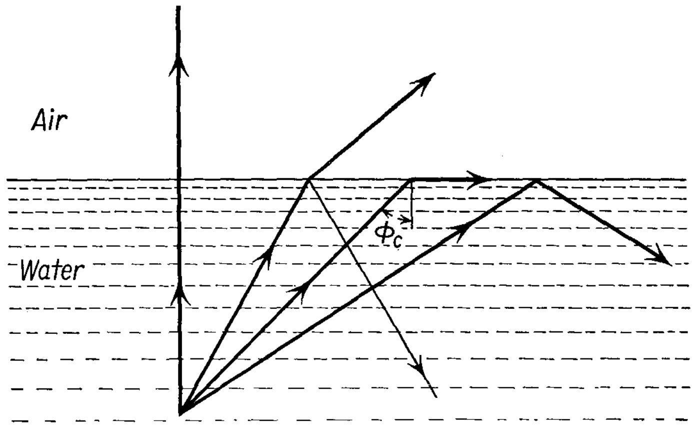


Fig. 37-9 Critical angle of refraction; total reflection.

and since n_2 is greater than n_1, the angle ϕ_1 must be greater than ϕ_2; that is, the ray is refracted away from the normal. The greatest value ϕ_1 may have is 90°. Thus we find that

$$\sin \phi_2 = n_1/n_2$$

for the greatest angle of incidence at which a ray will emerge in medium 1. This angle is the critical angle ϕ_c. In case air is the rarer medium, $n_1 = 1$, and, dropping the subscript 2, we find that

$$\sin \phi_c = 1/n. \tag{37-10}$$

At incident angles greater than the critical angle, Snell's law yields an imaginary solution, for which the sine of the angle of refraction is required to be greater than 1.

As illustrated in Figure 37-9, light incident upon a plane interface between an optically dense and an optically rare medium is partially re-

flected and partially transmitted as long as the angle of incidence is less than the critical angle. At angles greater than the critical angle, the light is totally reflected.

The property of total internal reflection is applied in optical instruments where glass prisms rather than silvered surfaces, are used to reflect light, for even the best silvered surface absorbs several per cent of the incident light. In prism binoculars a pair of 45° right-angle prisms, called Porro prisms, are used in each monocular to provide two reflections in each prism so that the resulting image is not reversed as shown in

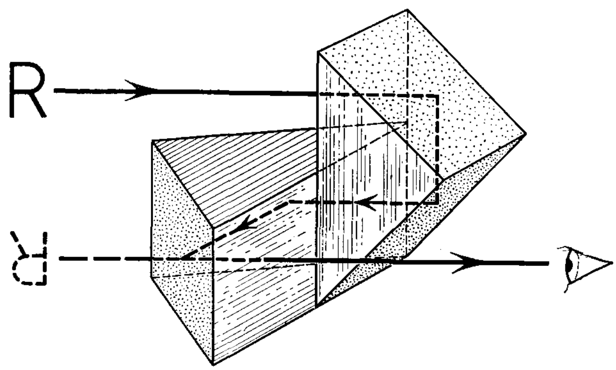


Fig. 37-10 Porro prisms, as used in a prism binocular.

Figure 37-10. The prisms are used to shorten the distance between the objective or front lens of the binocular and the eyepiece. A long piece of plastic rod may be used as a *light pipe*, and the light may be made to bend around corners by the property of total internal reflection. The rod is illuminated at one end and emerges from the other end of the polished rod. This concept is presently being extended in the field of *fiber optics*, where a bundle of fine glass fibers transmits the light incident at one end of the bundle down to the other end. Thus the image focused on one end of the bundle may be made to appear at the other end of a flexible bundle of fibers. This appears to have very promising application in medical research, permitting the visual examination of the interior of the stomach, for example. Display signs of glass or plastic are sometimes illuminated at the edge, and light emerges from the sheet of transparent material where the sheet has been engraved or lettered, so that the lettering is luminous, on a transparent background.

37-6 Refraction and Dispersion

As we have seen, the speed of light in a medium depends upon the nature of the medium and upon the wavelength of the light. A medium in which

the velocity of propagation of a wave depends upon the wavelength is called a *dispersive medium.* The index of refraction of a dispersive medium also depends upon the wavelength. If a narrow beam of white light is incident upon the face AB of a triangular glass prism, as shown in Figure 37-11, the emergent beam is not white but consists of an array of colors extending from red through orange, yellow, green, blue, and violet. The

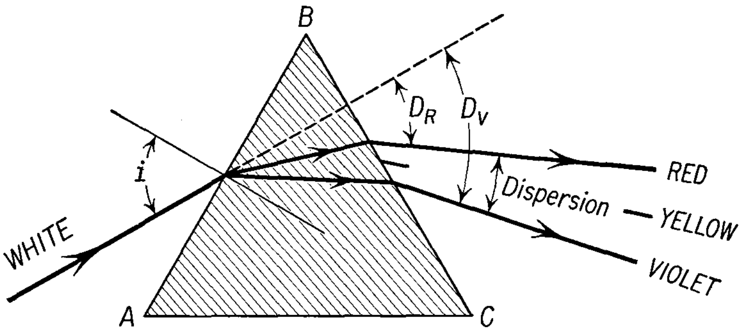


Fig. 37-11 Deviation and dispersion of a beam of white light by a triangular prism. The angle D_R is the deviation of the red light and D_V is the deviation of the violet light.

prism is said to *disperse* the incoming radiation into its spectrum. The white light incident upon the surface AB at some angle i is refracted and dispersed. The different wavelengths, or colors, which constitute the white light are bent toward the normal at the surface AB by different amounts, depending upon the index of refraction at the particular wavelength. The dispersed rays travel through the prism at different speeds and are refracted again at the surface AC, this time being refracted away from the normal to the surface AC, thus deviating the rays still more. The greater the index of refraction, the greater the angle of deviation. Thus red light, for which the index of refraction of glass is smallest, is deviated least by a glass prism, while violet light is deviated the greatest amount, corresponding to its high index of refraction.

If a very narrow beam of *monochromatic* light, that is, light of a single wavelength such as the yellow light from a sodium flame or a sodium arc, is sent through a triangular prism ABC which is made of some transparent substance, the beam will be deviated from its original direction through an angle D, the angle of deviation. If the angle of incidence is changed by rotating the prism with respect to the incident beam, the angle of deviation will also change. It may be shown by a rather tedious computation that the deviation is a minimum when the angle of incidence i and the angle of emergence e are equal, and when the ray within the prism is parallel to the base BC of the prism, as shown in Figure 37-12. There is a simple relationship connecting the *angle of minimum deviation* D_m, the angle A between

the two refracting faces of the prism, and the index of refraction n of the material of the prism, for the particular wavelength at which the angle of minimum deviation is observed. In Figure 37-12 we note that D_m is an exterior angle to the isosceles triangle whose base angles are each $i - r'$, and we get

$$D_m = (i - r') + (i - r') = 2(i - r').$$

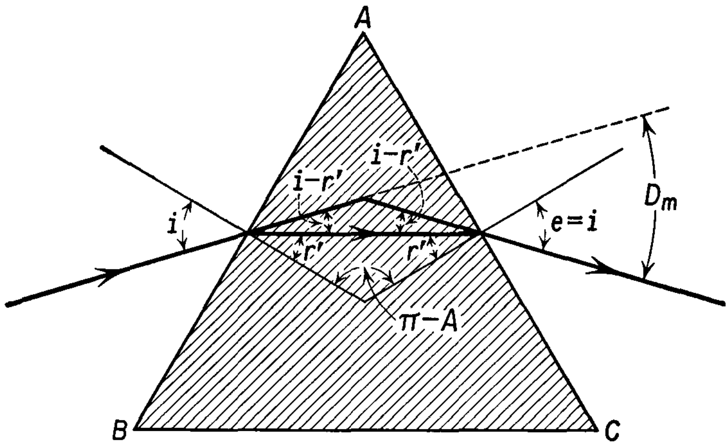


Fig. 37-12 Angle of minimum deviation.

The angle opposite A in the quadrilateral formed by the sides of the prism and the normals is equal to $\pi - A$. Since the sum of the angles of a triangle is equal to π, we may write

$$r' + r' + (\pi - A) = \pi,$$

so that

$$2r' = A.$$

Hence

$$D_m = 2i - A,$$

or

$$i = \frac{D_m + A}{2}.$$

Using Snell's law,

$$n = \frac{\sin i}{\sin r'} = \frac{\sin \frac{1}{2}(A + D_m)}{\sin \dfrac{A}{2}}. \tag{37-11}$$

If a transparent substance is fashioned in the form of a prism of refracting angle A, its index of refraction for any desired wavelength can be determined by measuring the angle of minimum deviation for this particular wavelength.

Problems

37-1. What is the velocity in water (a) of red light of wavelength 6,563 A and (b) of blue light of wavelength 4,861 A? Use the data of Table 37-1.

37-2. A narrow beam of light, of wavelength 7,600 A, is incident on the surface of water at an angle of 40°. Determine the angle of refraction in the water.

37-3. A narrow beam of white light is incident upon one face of a sheet of crown glass of thickness 5 cm at an angle of 45°. (a) Determine the displacement of the red light of the beam of wavelength 7,600 A. (b) Determine the displacement of the blue light of the beam for which $\lambda = 4{,}000$ A.

37-4. Water is placed in a flat-bottomed jar made of crown glass. The depth of the water is 5 cm, and the thickness of the glass bottom of the jar is 1 cm. Trace the path of a ray of light of wavelength 5,893 A which is incident on the top surface of the water at an angle of 45° and passes through both the water and the glass. (a) What is the angle of refraction in the water? (b) What is the angle of refraction at the glass-water interface? (c) What is the displacement of the beam when it emerges from the glass bottom of the jar? (d) What is the deviation of the beam when it emerges from the glass bottom of the jar?

37-5. The angle between the two refracting surfaces of a crown-glass prism is 60°. Trace the path of a ray of yellow sodium light ($\lambda = 5{,}893$ A) through the prism if the angle of incidence is 55°. (a) What is the angle of refraction at the first surface? (b) What is the angle of emergence at the second surface? (c) What is the angle of deviation of the emergent ray?

37-6. The absorption coefficient of a filter for a particular wavelength of light is 0.2 per centimeter. (a) What thickness of filter is required to reduce the incident flux upon the filter to $1/e$ of its original value? (b) What fraction of the incident flux will pass through a filter 0.5 cm thick? (c) If a thickness t of filter reduces the intensity of the beam by one half, what thickness will reduce the beam by one fourth?

37-7. A sheet of aluminum 0.4 cm thick reduces the flux of an incident beam of x-rays by 30 per cent. What is the absorption coefficient of aluminum for this radiation?

37-8. (a) Determine the angle of minimum deviation for a beam of light of wavelength 5,893 A passing through a 60° glass prism if its index of refraction is 1.65. (b) Determine the angle of incidence of this beam when its deviation is a minimum.

37-9. A low-power microscope is used to measure the index of refraction of a sample of glass which has parallel surfaces and is 1.20 cm thick. The microscope is first focused on a small dot on the lower surface; the microscope tube is raised until a small dot on the upper surface is in focus. The distance the microscope tube has been raised is found to be 0.80 cm. Determine the index of refraction of the glass.

37-10. When the refracting angle A of a prism is small, the angle of minimum deviation is also small. Show that, for a prism of small angle, the angle of minimum deviation is given by $D_m = A(n - 1)$. [HINT: For small angles the sine of θ may be approximated by θ, in radians.]

37-11. Using the equation derived in Problem 37-10, calculate the deviation of a ray of light going through a glass prism having a 5° refracting angle if the index of refraction of the glass is 1.65.

37-12. A drop of water of refractive index 1.33 is placed on the top face of a cube of glass of refractive index 1.66. At what angle should a ray of light be directed into the side face of the cube in order that the ray is reflected from the glass-water interface at the critical angle?

37-13. Show that the path taken by a light ray in reflection from a plane mirror is the path which requires the shortest time. [HINT: Show that the actual path length is a minimum.]

37-14. Show that the path taken by a light ray in its travel between a point A in medium α to a point B in medium β, which is separated from medium α by a plane interface, is the path of shortest time which can be taken by the light ray which passes from A to B.

38

Mirrors and Lenses

38-1 Spherical Mirrors

A spherical mirror consists of a small section of the surface of a sphere with one side of the surface covered with a polished reflecting material, usually silver or aluminum. If the outside, or convex surface, is silvered, we have a *convex* mirror; if the inside, or concave surface, is silvered, we have a

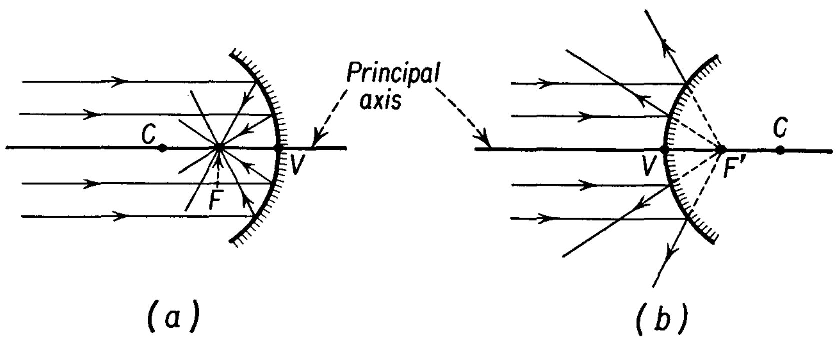


Fig. 38-1 Light parallel to the principal axis (a) converges toward F after reflection from a concave mirror, and (b) diverges from F' after reflection from a convex mirror.

concave mirror, as shown in Figure 38-1. Most mirrors used commercially are made of glass, with the rear surface silvered and then coated with a layer of paint or lacquer for protection. Mirrors for astronomical telescopes or other accurate scientific work are provided with a reflective coating on the front surface, for back-silvered mirrors give rise to two images, one from the glass-air interface and one from the glass-silver interface. This results in a loss of light from the primary reflection from the silvered surface. In the following discussion only front-surface mirrors will be considered. As a convention we will draw our diagrams in such a way that light incident upon the optical system is traveling from left to right.

The paths of light rays incident upon the surface of a mirror may be determined by application of the laws of reflection. The *principal axis* of a mirror is a line through the *center of curvature* C of the mirror, and the vertex V, as shown in Figure 38-1. A bundle of rays parallel to the principal axis of a concave mirror will be reflected from the mirror to converge to a point on the principal axis, called the *principal focus* F of the mirror, provided that the diameter of the mirror is small compared to its radius of curvature R. The principal focus lies midway between the vertex V and the center of curvature C of the mirror. A beam of light parallel to the principal axis of a convex mirror appears to diverge from a virtual focus F' after reflection from the mirror. As in the case of the concave mirror, the point F' lies midway between the vertex of the mirror and its center of curvature.

38-2 Image Formation in Spherical Mirrors

Let us examine the path of a ray of light PA emitted by a source at P and reflected from a concave mirror at A, as shown in Figure 38-2. The vertex of the mirror is at V, and the line PV is the principal axis of the mirror.

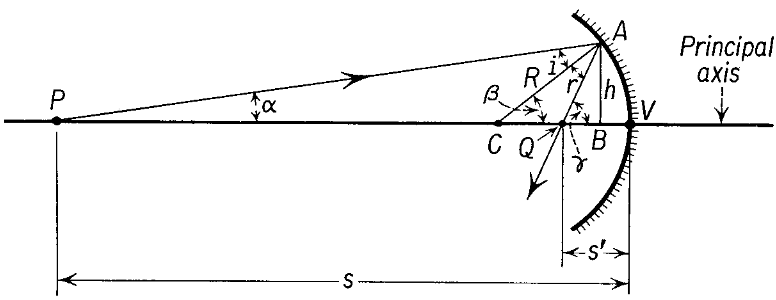


Fig. 38-2

The center of curvature of the mirror is at point C. To find the direction of the reflected ray at A, we draw the radius CA, of length R from the center of curvature C to the point A. The reflected ray AQ is drawn so that the angle of incidence PAC (or i) is equal to the angle of reflection CAQ (or r). The reflected ray intersects the principal axis of the mirror at point Q. Let us label the angle APC by α, the angle ACQ by β, and the angle AQB by γ. Since the exterior angle of a triangle is equal to the sum of the interior angles, we find from triangle APC that

$$\beta = \alpha + i,$$

while from triangle ACQ we find that

$$\gamma = \beta + r.$$

Making use of the fact that $i = r$, we find that

$$\alpha + \gamma = 2\beta.$$

If the mirror represents a small section of a spherical surface, the angles α, β, and γ are small and may be taken as equal to their tangents. We designate the length AB of the perpendicular from A to the principal axis as h, the distance PV of the source to the vertex of the mirror as s, the distance QV as s', and the distance CV as R. In addition we shall take the distance BV as negligibly small. We then have

$$\alpha = \tan \alpha = h/s,$$

$$\beta = \tan \beta = h/R,$$

and

$$\gamma = \tan \gamma = h/s'.$$

Substituting these results into the preceding equation, we find

$$\frac{1}{s} + \frac{1}{s'} = \frac{2}{R}. \qquad (38\text{-}1)$$

The above equation was derived by examination of a single reflected ray, under the restriction that the ray made a small angle with the principal axis of the mirror. Such a ray is called a *paraxial ray*. Since the ray chosen was arbitrary, all paraxial rays emitted from P, at an object distance s from the vertex of a concave mirror of radius R, will intersect at the point Q, called the image point, located an image distance s' from the vertex of the mirror. Since angle i equals angle r, a paraxial ray from Q will pass through P after reflection from the mirror; that is, the path of a ray of light that has undergone reflection, or refraction, can be reversed. Since the path of a light ray is reversible, all paraxial rays emitted from a source of light at Q will be brought to a focus at P. The points P and Q are therefore called *conjugate points* of the mirror.

From Equation (38-1) we note that when the source P is infinitely distant from the mirror, the image Q is located at a distance $R/2$ from the vertex. The light from P striking the mirror consists of a parallel beam. From the discussion of Section 38-1, the point where a parallel beam, parallel to the principal axis, is brought to a focus is called the *principal focus* of the mirror. The image distance of an infinitely distant object, located on the principal axis, is called the *focal length* f of the mirror. Thus we find

$$f = R/2 \qquad (38\text{-}2)$$

for the concave mirror, as indicated in Section 38-1.

It may be shown that Equations (38-1) and (38-2) are correct for all positions of the image and object of both concave and convex mirrors, provided that the following *sign conventions* are employed:

(a) Light is assumed to come to the mirror from the left.

(b) The distance s is positive when the object is to the left of the mirror.

(c) The distance s' is positive when the image is to the left of the mirror.

(d) The radius of curvature R and the focal length f are to be taken as positive for a concave mirror and negative for a convex mirror.

(e) The positive and negative signs are to be used *only* when numerical values are substituted for s, s', f, and R.

We may examine the images of a finite object formed by a mirror by considering four particular rays from the object, called *principal rays*, chosen for their simplicity in graphical construction. In the imaging process, all rays from the object intersect at the image point. Consequently, the image point may be located by finding the intersection of any two rays after these have been reflected from the mirror. The principal rays generally chosen are: (a) a ray which is parallel to the principal axis, and which

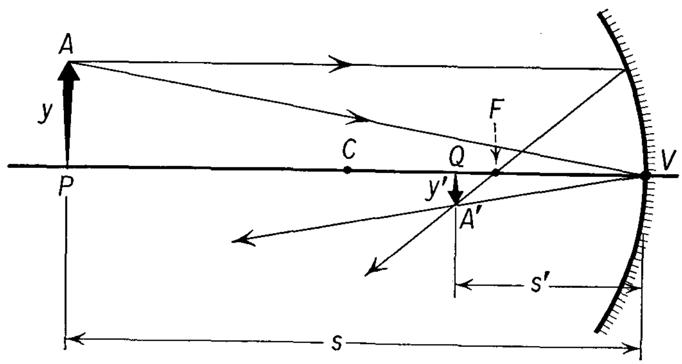


Fig. 38-3 A real, inverted, smaller image is formed by a concave mirror when the object distance is greater than $2f$.

therefore must pass through the principal focus of the mirror after reflection (or, in the case of the convex mirror, which appears to diverge from the principal focus); (b) a ray which is directed toward the vertex of the mirror and is reflected as though the mirror is plane; (c) a ray which is directed toward the principal focus and which is reflected from the mirror parallel to the principal axis; (d) a ray which is directed to or from the center of curvature of the mirror, and hence strikes the mirror at normal incidence; this ray is reflected back along its incident path.

From Equation (38-1) we note that

$$\frac{1}{s'} = \frac{1}{f} - \frac{1}{s}.$$

As long as the object distance s is greater than the focal length f of a *concave* mirror, the image distance is positive. In the event that the object distance

is greater than $2f$, the image distance is less than $2f$, and the image is reduced in size, as shown in Figure 38-3. The position of the image A' of the head of the arrow at A is located by the use of a principal ray parallel to the principal axis and a second principal ray reflected from the vertex of the mirror. The image is smaller then the object and inverted. From the similar triangles AVP and $A'VQ$ we note that the size of the image y' is to the image distance s' as the size of the object y is to the object distance s'. We may define the linear or transverse *magnification* m of the system as

$$m = \frac{y'}{y}, \tag{38-3}$$

where we adopt the convention that the *distance* y (or y') *is positive above the axis and is negative below the axis.* From the figure, and the sign conventions on s and s', we note that

$$m = -\frac{s'}{s}. \tag{38-4}$$

As before, the positive and negative signs associated with the sign conventions are to be used only when numerical values are substituted into the above equations. In the case illustrated in Figure 38-3, the magnification is negative, which means that the image is inverted, and the value of m is less than 1, which means that the image is smaller than the object.

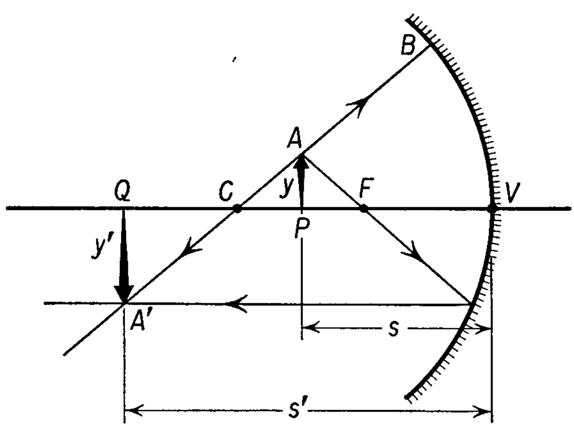


Fig. 38-4 A real, inverted, enlarged image is formed by a concave mirror when the object is between C and F.

In Figure 38-4 we examine the case of an object which lies between the center of curvature and the focal point of the mirror. The principal rays used are the ray AB, which strikes the mirror normally and is reflected

back along its own path, and the ray AF, which passes through the focus F and is reflected parallel to the principal axis. These rays intersect in the point A', the image of A. The image is real; it can be seen on a ground glass as the point where light rays actually intersect.' The image $A'Q$ is inverted and enlarged. The magnification is negative and greater than 1.

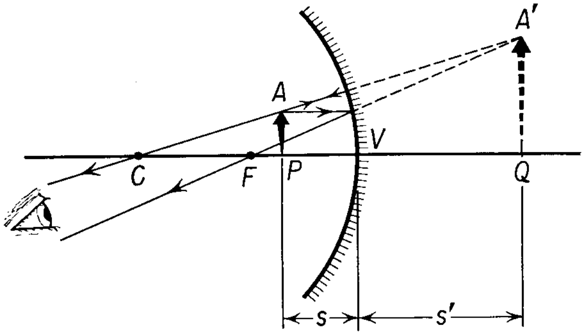


Fig. 38-5 A virtual, enlarged image is formed by a concave mirror when the object distance is less than f.

Concave mirrors are often used as shaving mirrors to produce an erect, enlarged image. In Figure 38-5 we see that the image A' of a real object A, placed between the focal point and the vertex of the mirror, is enlarged and erect. The magnification is positive and greater than 1. Since the rays reflected from the mirror do not actually intersect but only appear to do so, as shown by the dotted lines, the image is virtual. An observer looking at the image A' sees rays which appear to diverge from A' and interprets this point as the origin of the rays. While the image can be seen, it cannot be focused on a ground glass, for rays from A do not converge to any point to the left of the mirror, which is the only region of space in which there is actually any light.

These relationships may all be calculated analytically by use of Equations (38-1) through (38-4) and by use of the sign conventions given above.

Illustrative Example. An object is placed 7.5 cm from a concave mirror whose focal length is 15 cm. Determine the position and character of the image.

The graphical solution is shown in Figure 38-5. We imagine the object to be to the left of the mirror. The object distance is $s = 7.5$ cm, and the focal length is $f = 15$ cm. We have

$$\frac{1}{s} + \frac{1}{s'} = \frac{1}{f},$$

$$\frac{1}{s'} = \frac{1}{15} - \frac{1}{7.5} = -\frac{1}{15},$$

$$s' = -15 \text{ cm}.$$

Thus the image is formed a distance of 15 cm to the right of the mirror. Since the image appears where there is no reflected light, the image is virtual. The magnification produced is

$$m = -\frac{s'}{s} = -\frac{-15}{7.5} = +2.$$

The image is erect and is twice the size of the object.

No matter where the object is placed in front of a *convex* mirror, its image will always be virtual, erect, and smaller than the object, as shown in Figure 38-6. A ray from A parallel to the axis is reflected back as though

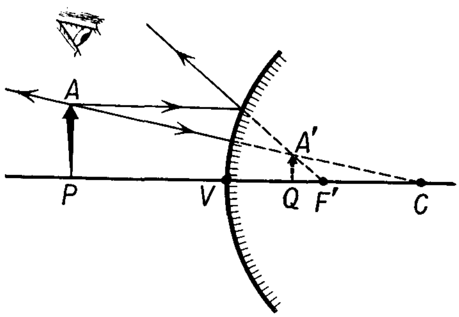


Fig. 38-6 A virtual, erect, smaller image is formed by a convex mirror for any position of a real object.

it came from the virtual focus F'. A second principal ray, taken normal to the mirror, is reflected back on itself. These two rays do not meet but appear to have originated from the point A', the virtual image of A. The image $A'Q$ is virtual, erect, and smaller than the object.

Illustrative Example. An object is placed 45 cm in front of a convex mirror whose radius of curvature is 30 cm. Determine the position and character of the image.

Following the sign conventions indicated above, we have $s = +45$ cm, $f = R/2 = -15$ cm, and, substituting into the appropriate equation, we find

$$\frac{1}{s} + \frac{1}{s'} = \frac{1}{f},$$

$$\frac{1}{s'} = -\frac{1}{15} - \frac{1}{45} = -\frac{4}{45},$$

$$s' = -11.25 \text{ cm}.$$

The magnification is

$$m = -\frac{s'}{s} = -\frac{-11.25}{45} = +0.25.$$

The image is virtual and is situated 11.25 cm behind the mirror. Since the magnification is positive, the image is upright and, in this case, is one fourth the size of the object.

38-3 Lenses

Lenses, either singly, or in combination, are used for the formation of images and are made of a transparent material provided with spherical or plane surfaces. For certain special cases, other surfaces may be used to eliminate certain defects of spherical lenses, or to provide certain desired effects, but the great advantage of spherical surfaces is that they are easiest

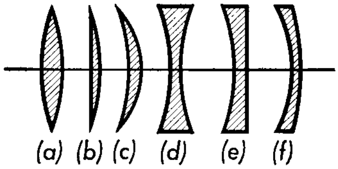
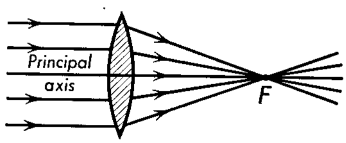


Fig. 38-7 Spherical lenses: (a), (b), and (c) are converging lenses; (d), (e), and (f) are diverging lenses.

to make. Spherical lenses are classified as either *converging* or *diverging* lenses. Some common forms of lenses are shown in Figure 38-7, where we note that a converging lens is thicker at the center than at the edges, while a diverging lens is thinner at the center than at the edges.

Fig. 38-8 Rays parallel to the principal axis are converged toward the principal focus F by the converging lens.



A beam of parallel light incident on a converging lens will be converged toward a point F, known as the *principal focus* of the lens, as shown in Figure 38-8. A beam of parallel light incident upon a diverging lens will

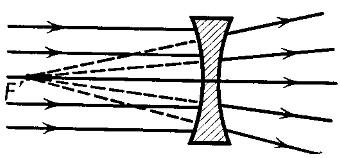


Fig. 38-9 Rays parallel to the principal axis are diverged by the diverging lens. After passing through the lens the rays appear to come from the virtual focus at F'.

diverge after passing through the lens as though it came from F', the principal focus, as shown in Figure 38-9. The point F' is called a virtual focus, since the rays do not actually pass through it.

The action of the lens is due to the refraction of the light as it enters and leaves the spherical surfaces bounding the lens. The effect of a converging lens on the plane wave fronts of a beam of parallel light is shown in Figure 38-10. The part of the wave front which passes through the center of the lens is retarded more than the part which passes through the

outer part of the lens, so that the emerging wave front is spherical, with its center at the principal focus F. Parallel wave fronts incident upon a

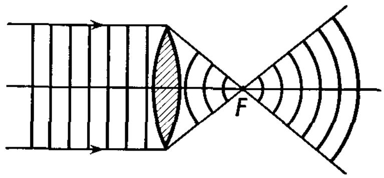


Fig. 38-10 Change in wave front produced by a converging lens on a parallel beam of light incident upon the lens.

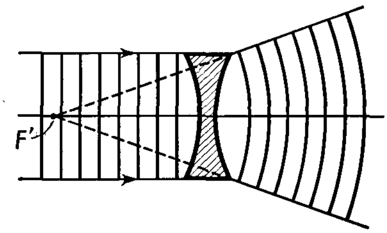


Fig. 38-11 Change in wave front produced by a diverging lens on a parallel beam of light incident upon the lens.

diverging lens are retarded more at the thicker edge of the lens than at its center, and again the emergent wave front is spherical, with its center at F', as shown in Figure 38-11.

38-4 Lensmaker's Formula

Let us consider the process of image formation in a thin, converging lens, as shown in Figure 38-12. We shall neglect the thickness of the lens and shall assume that the diameter of the lens is small in comparison with the radius of curvature of either surface, so that all angles with which we must deal may be approximated by their sines or tangents. For definiteness we shall study the image of an object point located on the principal axis of the lens, but none of the approximations made will depend upon this fact. Thus the object and image relationships that will be derived will be equally valid for pairs of conjugate points lying on any line parallel to the principal axis, and, just as in the case of mirrors, we shall find that points lying on a plane perpendicular to the principal axis, called the *object plane*, will be imaged on a second plane perpendicular to the principal axis, called the *image plane*. We shall assume that the lens is constructed of a medium of refractive index n and that the lens is immersed in air or vacuum.

Let us choose an arbitrary ray PA from the object P which strikes the lens at A. The line AC in Figure 38-12(b) is the normal to the front surface of the lens, and the refracted ray is in the direction AB. By making use of the theorem that the external angle of a triangle is equal to the sum of the two interior angles, we see from triangle ACP that

$$i_1 = \alpha_1 + \beta_1,$$

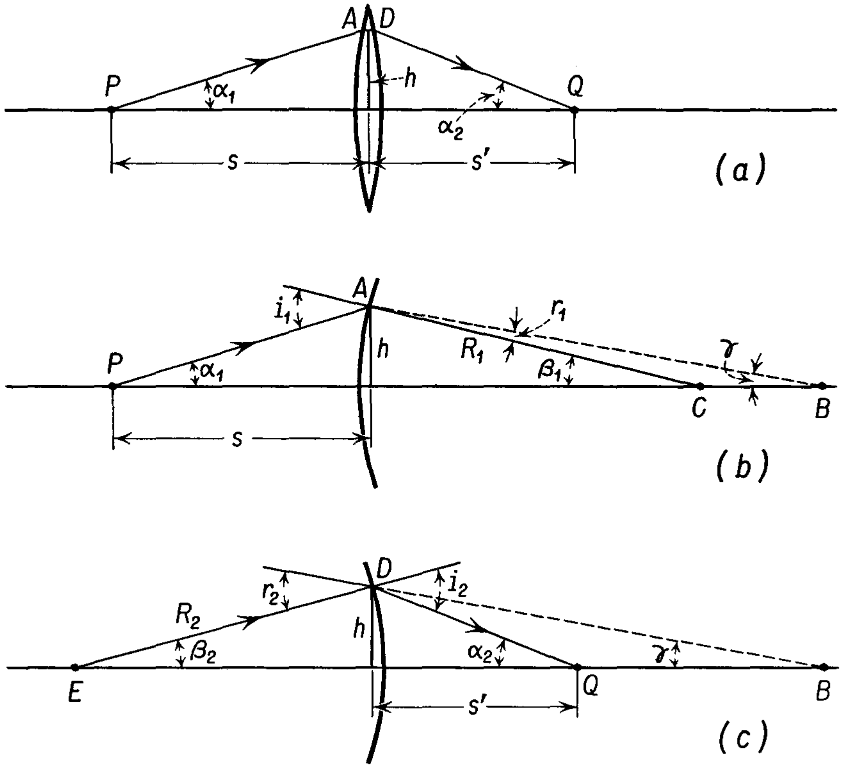


Fig. 38-12 (a) Path of a ray of light through a thin lens. (b) Refraction at first surface. (c) Refraction at second surface.

and, from triangle ABC we find

$$r_1 = \beta_1 - \gamma.$$

For small angles Snell's law may be stated as

$$n = i/r,$$

so that

$$\alpha_1 + \beta_1 = n(\beta_1 - \gamma). \tag{38-5}$$

The refracted ray AB never really reaches the point B on the principal axis but, instead, is refracted at the second surface of the lens, at the point D, Figure 38-12(c), and intersects the principal axis at the point Q, the image of point P. The normal to the second surface of the lens is DE. From the figure we find from triangle DEQ that

$$i_2 = \beta_2 + \alpha_2,$$

while from triangle DEB we find that

$$r_2 = \beta_2 + \gamma.$$

For small angles we have $n = i_2/r_2$, so that

$$\beta_2 + \alpha_2 = n(\beta_2 + \gamma). \tag{38-6}$$

By eliminating γ from Equations (38-5) and (38-6), we find

$$\alpha_1 + \alpha_2 = (n - 1)(\beta_1 + \beta_2).$$

We make a number of approximations consistent with the idea of a thin lens of small diameter. Thus we say that the distance from A to the principal axis is very closely equal to the distance from D to the principal axis, and we label this distance h. The object distance s is the distance from P to the lens, and the image distance s' is the distance from Q to the lens, neglecting the thickness of the lens. We may approximate a small angle by either its sine or its tangent, so that $\alpha_1 = h/s$, $\alpha_2 = h/s'$, $\beta_1 = h/R_1$, $\beta_2 = h/R_2$, and, substituting these relations into the above equation, we find

$$\frac{1}{s} + \frac{1}{s'} = (n - 1)\left(\frac{1}{R_1} + \frac{1}{R_2}\right). \tag{38-7}$$

We may find the principal focal length f of a thin lens by finding the image distance of an infinitely distant object. When s is infinity, we find $s' = f$, and from Equation (38-7)

$$\frac{1}{f} = (n - 1)\left(\frac{1}{R_1} + \frac{1}{R_2}\right). \tag{38-8}$$

Equation (38-8) is called the *lensmaker's equation.* We may write Equation (38-7) as

$$\boxed{\frac{1}{s} + \frac{1}{s'} = \frac{1}{f}.} \tag{38-9}$$

It may be shown that Equations (38-7), (38-8), and (38-9) are applicable to all thin lenses, through the use of the following *sign conventions:*

(a) The radius of curvature R is positive for a convex surface and negative for a concave surface.

(b) Light is assumed to come to the lens from the left.

(c) The object distance s is positive when the object is to the left of the lens.

(d) The image distance s' is positive when the image is to the right of the lens.

(e) The focal length f is positive for a converging lens and negative for a diverging lens.

(f) The positive and negative signs are to be used only when numerical values are substituted for the symbols.

The above equations were derived on the assumption that the lens was in air or vacuum and that n was the absolute index of refraction of the material of the lens. The equations are also correct when the lens is immersed in any medium, provided that n is interpreted as the relative index of refraction of the material of the lens with respect to the medium.

Just as in the case of mirrors, we may define the transverse magnification m of a lens by the equation

$$m = \frac{y'}{y}, \tag{38-10}$$

where y is the transverse dimension of the object, and y' is the transverse dimension of the image, taken positive as above the axis and negative below the axis. In terms of the object and image distances, this equation may be rewritten as

$$m = -\frac{s'}{s}. \tag{38-11}$$

38-5 Image Formation by Thin Lenses

We may examine the images formed in a lens by considering three *principal rays*, chosen for their simplicity in construction. In the imaging process, all rays from the object intersect at the image point, so that the image point may be located by finding the intersection of any two rays after they have passed through the lens. The principal rays generally chosen are: (a) a ray which is parallel to the principal axis, which must pass through the principal focus of a converging lens after refraction (or, in the case of a diverging lens the ray must diverge from the principal focus); (b) a ray which is directed toward the principal focus of the lens, and which is refracted by the lens so as to emerge parallel to the principal axis; and (c) a ray which is directed toward the center of the lens and passes through the lens undeviated, for the section of the lens at its center may be approximated as a plane parallel plate of negligible thickness.

The types of images formed by a *converging* lens may be determined with the aid of Equation (38-9),

$$\frac{1}{s} + \frac{1}{s'} = \frac{1}{f}.$$

We note that as long as s is greater than f, the image distance will be positive, and the image will be formed to the right of the lens. Figure 38-13 shows an object AB located a distance from a converging lens greater than $2f$. The image A' of the head A of the arrow is located by means of a ray parallel to the principal axis and a ray passing through the center of the lens. It is observed that the image is smaller than the object, inverted, and

real. The image can be seen by looking along the axis from a point beyond the image, or the image may be formed on a screen at the location of $A'B'$ and then viewed from any convenient position as a result of the light scattered by the diffuse screen.

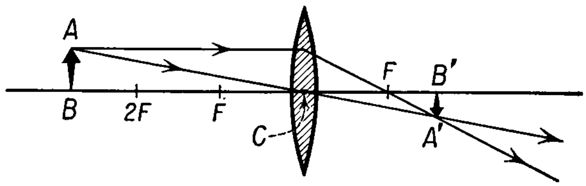


Fig. 38-13 Image formed by a converging lens when the object distance is greater than $2f$.

In the event that the object distance s lies between f and $2f$, the image distance s' is greater than the object distance. The image is real, inverted, and is larger than the object, as shown in Figure 38-14.

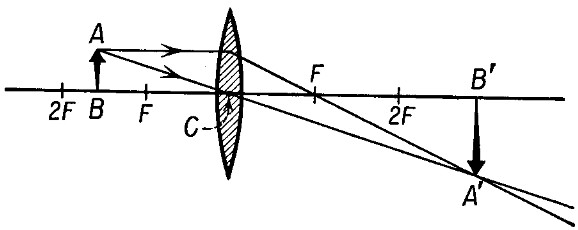


Fig. 38-14 Real, enlarged image formed by a converging lens when the object distance is greater than f and less than $2f$.

When the object distance is smaller than f, we see that the image distance is negative. As shown in Figure 38-15, the rays from A do not

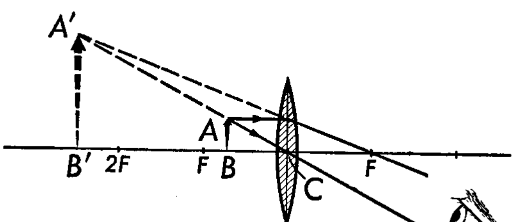


Fig. 38-15 Virtual, enlarged image formed by a converging lens when the object distance is less than f.

meet after passing through the lens, but when projected back through the lens, the refracted rays, when seen by the eye, appear to have originated

from the point A', on the same side of the lens as the object. The image in this case is virtual, erect, and is larger than the object. The virtual image cannot be formed on a ground glass but can be seen by looking through the lens.

Illustrative Example. An object 6 cm tall is situated 45 cm from a converging lens of 15 cm focal length. Determine the position, size, and nature of the image.

This problem is illustrated in the ray diagram of Figure 38-13. The object distance is $s = +45$ cm, and the focal length is $+15$ cm. We find

$$\frac{1}{s} + \frac{1}{s'} = \frac{1}{f},$$

and, substituting numerical values,

$$\frac{1}{s'} = \frac{1}{15} - \frac{1}{45} = \frac{2}{45},$$

$$s' = 22.5 \text{ cm}.$$

To find the size of the image

$$m = \frac{y'}{y} = -\frac{s'}{s}.$$

Since $y = 6$ cm, we find

$$y' = -y\frac{s'}{s} = -6 \text{ cm} \times \frac{22.5 \text{ cm}}{45 \text{ cm}} = -3 \text{ cm}.$$

The image is real, for s' is positive, and is inverted and half the size of the object.

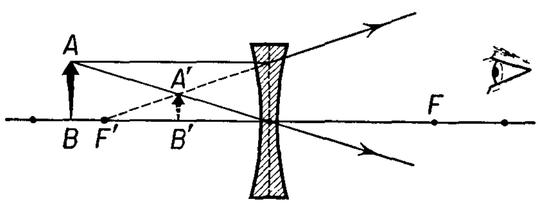


Fig. 38-16 Virtual image formed by a diverging lens.

The formation of an image by a *diverging* lens is illustrated in Figure 38-16. The rays from A diverge after passing through the lens and do not meet on the right-hand side of the lens. The rays appear to originate in the point A', the virtual image of A. A diverging lens cannot form a real image of a real object. In Equation (38-9) we note that the focal length of a diverging lens is negative, so that the image distance s' is negative for all positive values of the object distance s.

Illustrative Example. An object is placed 25 cm in front of a diverging lens of 15 cm focal length. Find the position and character of the image.

An optical diagram of the system is shown in Figure 38-16. We have $s = +25$ cm, $f = -15$ cm, and

$$\frac{1}{s'} = -\frac{1}{15} - \frac{1}{25},$$

$$s' = -9.4 \text{ cm}.$$

The magnification produced is

$$m = -\frac{s'}{s} = -\frac{-9.4}{25} = +0.38.$$

The image is virtual, erect, and reduced with respect to the object.

38-6 Combinations of Lenses

Two or more lenses are frequently used in combination to produce a desired result. We may treat the effect of the lenses sequentially, first determining the image formed by the first lens of the combination, then using the image of the first lens as the object of the second lens, and so on. If two thin lenses are placed in contact, so that their principal axes coincide, the combination may be treated as a single lens of focal length f such that

$$\frac{1}{f} = \frac{1}{f_1} + \frac{1}{f_2}. \tag{38-12}$$

We may easily obtain this result by writing Equation (38-9) twice, once for lens 1 and once for lens 2, and adding these equations, as

$$\frac{1}{s_1} + \frac{1}{s_2} + \frac{1}{s_1'} + \frac{1}{s_2'} = \frac{1}{f_1} + \frac{1}{f_2}.$$

If the lenses are sufficiently thin so that we may neglect their thickness, we note that the numerical value of the image distance s_1' is the same as the numerical value of the object distance of the second lens s_1, but that the image distance of the first lens has an opposite sign from the object distance of the second lens. These two terms cancel, and we have

$$\frac{1}{s_1} + \frac{1}{s_2'} = \frac{1}{f_1} + \frac{1}{f_2}.$$

Since the lenses are thin, there is no longer any need to use subscripts to distinguish the object and image distances, and, comparing the above equation with Equation (38-9), we note that the quantity on the right is the effective focal length of the combination of lenses, as given in Equation (38-12), and that the two-lens combination may be replaced by a single

lens of focal length f, in accordance with this equation. This result is often used by oculists in fitting eyeglasses. The focal length of the required correction lens is determined by placing a variety of thin lenses in contact before the patient's eye and computing the focal length of the combination from Equation (38-12).

Illustrative Example. An object is placed 14 cm in front of a converging lens of 10 cm focal length. Another converging lens of 7 cm focal length is placed at a distance of 40 cm to the right of the first lens. Determine the position and character of the final image.

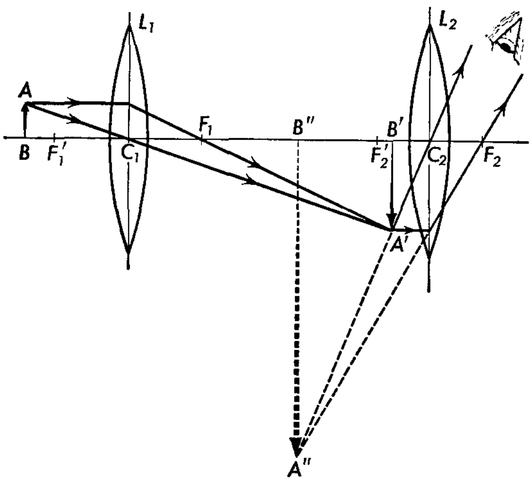


Fig. 38-17

The graphical solution is shown in Figure 38-17. The image produced by the first lens is $A'B'$ which is real, inverted, and larger than the object. This is used as the object of the second lens. To find the image in the second lens, we again use a principal-ray construction, with $A'B'$ as an object for the second lens. This is only a convenience in locating the image, for we do not infer that the light rays are suddenly deviated at the point A'. The final image is $A''B''$, which is virtual, larger than $A'B'$, and erect with respect to it. Thus the final image is inverted, with respect to the original object, and is virtual.

We can solve the problem by two successive applications of the lens equation. For the first lens L_1, $f_1 = 10$ cm, and $s_1 = 14$ cm, so that

$$\frac{1}{s_1'} = \frac{1}{10} - \frac{1}{14},$$

$$s_1' = 35 \text{ cm}.$$

Since the two lenses are 40 cm apart, the image $A'B'$ is 5 cm from the second

lens L_2. Using $A'B'$ as the object for the second lens, $s_2' = 5$ cm, $f_2 = 7$ cm, and we find

$$s_2' = -17.5 \text{ cm};$$

that is, the final image is 17.5 cm to the left of the second lens and is virtual.

Since the second lens magnifies the image produced by the first lens, the total magnification of the system is the product of the magnification m_1 of the first lens by the magnification m_2 of the second lens; that is,

$$m = m_1 m_2 = \frac{s_1'}{s_1} \times \frac{s_2'}{s_2},$$

or

$$m = -\frac{35}{14} \times \frac{17.5}{5} = -8.75.$$

The negative sign shows that the final image is inverted with respect to the original object AB.

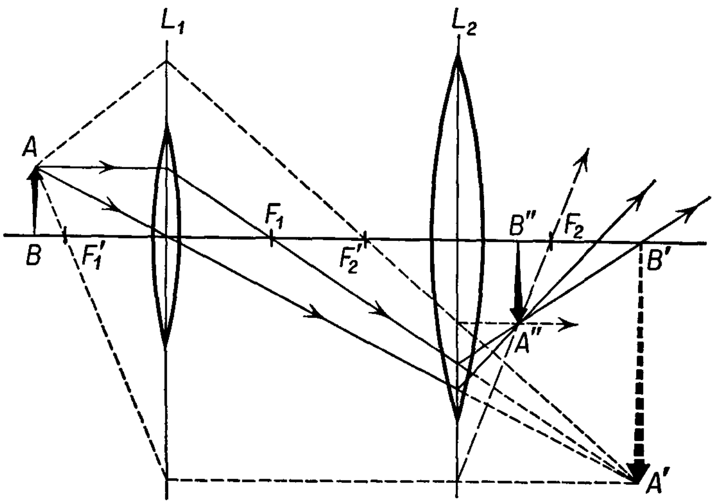


Fig. 38-18

Illustrative Example. Two converging lenses L_1 and L_2 are placed 40 cm apart, their principal axes coinciding. An object AB is placed 15 cm in front of L_1. The focal length of L_1 is 12 cm, and the focal length of L_2 is 10 cm. Determine the position and character of the final image.

The graphical solution is shown in Figure 38-18. Principal rays from the object AB would have intersected at the image $A'B'$ had they not been intercepted by the lens L_2. For purposes of construction, we may imagine the lenses L_1 and L_2 to be of any diameter. If the lens L_1 were sufficiently large, there would have been rays through L_1 to the point A' in many different directions. Thus there would have been a ray from A through L_1 to A' parallel to the principal axis. Another possible ray from A through L_1 to A' might pass

through the center of lens L_2. The second of these rays passes undeviated through L_2, while the first must pass through the principal focus F_2 of L_2 after passing through this lens. A third possible ray from A through L_1 to A' might have passed through F_2' and, upon passage through the lens L_2, would have been deflected parallel to the principal axis. We draw two of these rays and locate the final image by their point of intersection as $A''B''$.

The object for lens L_2 is the image $A'B'$ formed by lens L_1; it is a *virtual object* for lens L_2, for the rays striking L_2 are converging toward $A'B'$ rather than diverging from it.

Analytically, we find

$$\frac{1}{s_1'} = \frac{1}{12} - \frac{1}{15},$$

$$s_1' = 60 \text{ cm}.$$

Thus $A'B'$ is 20 cm to the right of lens L_2, and, by our sign convention, the object distance for the second lens is $s_2 = -20$ cm. The position of the image formed by L_2 may be found from

$$\frac{1}{s_2'} = \frac{1}{10} - \frac{1}{-20},$$

$$s_2' = 6.7 \text{ cm};$$

that is, the final image is 6.7 cm to the right of L_2. The magnification produced by this combination of lenses is

$$m = \frac{s_1'}{s_1} \times \frac{s_2'}{s_2} = \frac{60}{15} \times \frac{6.7}{-20} = -1.34.$$

The negative sign shows that the final image is inverted with respect to the original object.

38-7 Spherical Aberration of Lenses and Mirrors

In the discussion of spherical mirrors, it was stated that the mirror should be a small portion of the sphere. If the mirror surface is a large portion of the sphere, the images are blurred. This blurring of the image is due to *spherical aberration;* that is, rays from one point of the object reflected by different portions of the spherical surface do not meet in a point but cover a sizable area. When a beam of parallel rays is incident upon a large mirror, the reflected rays do not all pass through the principal focus, as shown in Figure 38-19. The rays which are reflected from the outer portion of the mirror cross the principal axis closer to the mirror than do those which are reflected from the central portion. Instead of a sharp point, there is a sizable focal spot.

A sharp focus of light parallel to the principal axis can be obtained with a large mirror if the surface is parabolic in shape instead of spherical. It is a property of a parabola that any ray from the focus which strikes the

parabolic surface is reflected parallel to the principal axis, as shown in Figure 38-20. If a very small source of light is placed at the focus, a parallel beam is reflected from parabolic mirrors; this is another illustration of the reversibility of the paths of rays of light. Parabolic mirrors are commonly used in automobile head lamps and in searchlights. The very

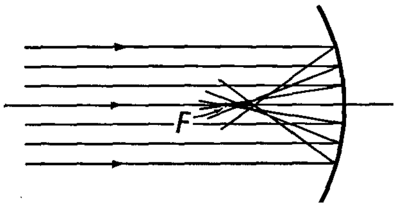


Fig. 38-19 Spherical aberration produced by a mirror.

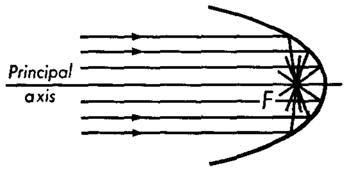
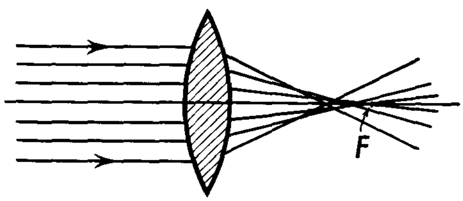


Fig. 38-20 A beam of light parallel to the principal axis is focused in a point by a parabolic mirror.

large astronomical telescopes are reflectors, using parabolic mirrors. Parabolic mirrors are best used in cases in which the incident or emergent light is parallel to the principal axis of the mirror, for these mirrors show aberrations when the image and object are off the principal axis.

The images formed by lenses also show spherical aberration. These aberrations are most clearly indicated by the use of monochromatic light, such as the yellow light from a sodium lamp. A set of parallel rays of

Fig. 38-21 Spherical aberration of a lens.



monochromatic light parallel to the principal axis of a lens is brought to a focus in a blurred spot, as shown in Figure 38-21, the amount of blurring depending upon the diameter of the lens. The rays passing through the outer portion of the lens are deviated so that they cross the principal axis closer to the lens than do the rays passing through the center portion. One way of reducing spherical aberration is to use a diaphragm to limit the light from the object to the central portion of the lens.

It must be emphasized that accurate lenses are not designed by use of the lensmaker's formula. Practical lenses are complex combinations of several different lens elements, combined to minimize spherical aberration and other aberrations which will not be discussed here. In many cases the lens designer finds it necessary to trace rays through the optical system

by use of Snell's law, and must consider skew rays as well as paraxial rays in order to assure a sharp image.

38-8 Chromatic Aberration

From the lensmaker's equation we note that the focal length of a lens depends upon its refractive index. Most transparent substances are dispersive media and thus have different refractive indices at different wavelengths. Thus, when white light is incident upon a lens, the red component

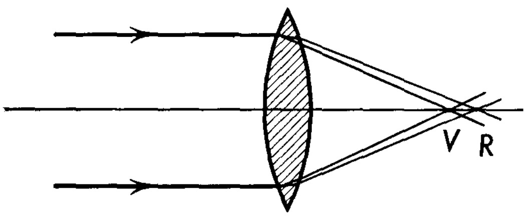


Fig. 38-22 Chromatic aberration of a lens.

of the light will be focused farthest from the lens, while violet light will be focused closest to the lens, as shown in Figure 38-22. The image of a lamp focused by a lens on a screen will be observed to have a colored halo.

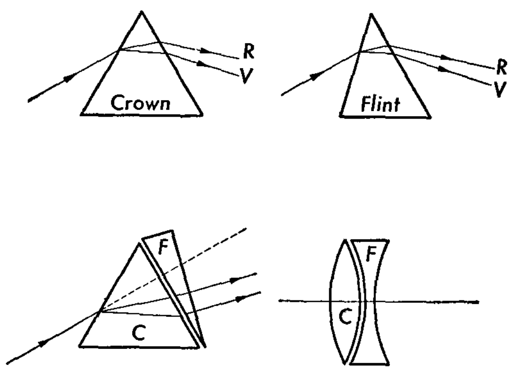


Fig. 38-23 The design of an achromatic lens.

Lenses used for visual or photographic purposes must be corrected so as to minimize this *chromatic aberration.*

In some special cases lenses may be achromatized by use of two converging lenses of the same kind of glass, as in the case that either the incident or emergent light is a parallel beam. In this case the separation between the lenses must be half the sum of the focal lengths of the two lenses. Ramsden and Huygens eyepieces used with telescopes and microscopes are built in this way (see Section 39-4). In general, lenses are achromatized by use of elements constructed from different kinds of glass. The basic idea in the design of such an achromatic lens is illustrated in Figure 38-23, which shows light passing through a prism of crown glass and

a prism of flint glass. The prism angles have been so chosen that the prisms deviate the light differently, but the angular width of the spectrum, or the dispersion, is the same in both cases. When these two prisms are placed in contact with their vertex angles opposed, the prism combination produces a deviated beam of white light, which is not dispersed, for the dispersion of one prism has been compensated by the dispersion of the second prism. A converging achromatic lens combination can therefore be made by combining a crown-glass lens of short positive focal length with a flint-glass lens of longer negative focal length. These may be cemented together to form a single achromatic lens.

38-9 Illuminance of an Image

In all optical systems the amount of light reaching the image is an extremely important consideration. We have already indicated in Section 36-5 that the luminance, or brightness, of a source, rather than its luminous flux

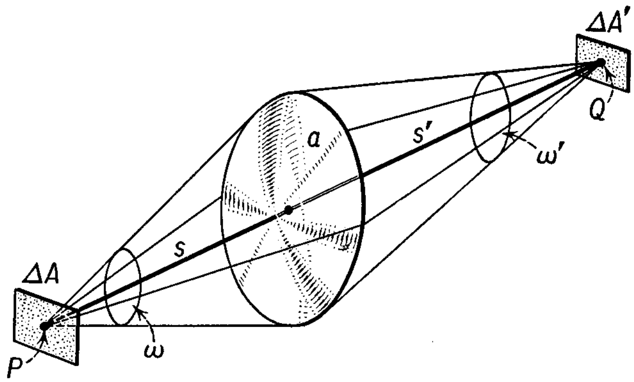


Fig. 38-24

output, determines the illuminance. Let us consider the formation of an image by a simple lens, as shown in Figure 38-24. Let us suppose that the magnification of the system is m. Then an element of area ΔA of the source is imaged in an element of area $\Delta A'$ on the image, such that

$$\frac{\Delta A'}{\Delta A} = m^2. \tag{38-13}$$

Let us assume that all the light emitted by the source which strikes the lens from ΔA is imaged upon $\Delta A'$; that is, there is no absorption or reflection from the lens surfaces. If ω is the solid angle subtended by the lens from the source, the luminous flux F which passes into the lens from ΔA is given by

$$F = B\,\Delta A\,\omega,$$

where B is the luminance or brightness of the source, expressed in lumens per square centimeter per steradian. The illuminance E of the image is the incident flux per unit area, or

$$E = \frac{F}{\Delta A'} = \frac{B\,\Delta A\,\omega}{\Delta A'},$$

or, substituting from Equation (38-13),

$$E = \frac{B\omega}{m^2}. \tag{38-14}$$

Thus the frontal area of the lens (as it influences ω) and the brightness of the source determine the brightness of the image. For a given object distance, the greater the magnification, the smaller is the illuminance of the image. If a represents the frontal area of the lens, the solid angle subtended by the lens at the source is $\omega = a/s^2$, while the solid angle ω' subtended by the lens at the position of the image is $\omega' = a/(s')^2$. Thus

$$\omega = \omega' \frac{(s')^2}{s^2} = \omega' m^2.$$

Substituting this relationship into Equation (38-14), we find

$$E = B\omega'. \tag{38-15}$$

Thus the illuminance of the image depends upon the brightness of the source and upon the solid angle subtended by the lens at the position of the image. The brightness of the source rather than its luminous output determines the brightness of the image.

It is a common misconception that searchlights produce a bundle of parallel rays extending to infinity. Such might be the case with point sources of light at the focal point, but practical sources are extended sources rather than point sources. Portions of the source which do not lie at the focal point, or on the axis, produce bundles of rays which are not parallel to the axis. The illumination of a point on the axis of a searchlight varies inversely with the square of the distance from the source, in accordance with Equation (38-15). The size of the source does not affect the illumination produced on the axis of the searchlight. It is for this reason that carbon arcs are used in high-intensity searchlights as well as in commercial motion-picture projection systems.

Problems

38-1. An object is placed 20 cm from a concave spherical mirror whose radius is 24 cm. Determine algebraically and graphically the position of the image. Describe the image.

38-2. An object 4 cm long is placed 60 cm from a concave spherical mirror whose radius of curvature is 40 cm. Determine (a) the position and (b) the size of the image. (c) Describe the image. (d) Draw a ray diagram.

38-3. An object 3 cm long is placed 4 cm from a concave spherical mirror whose focal length is 12 cm. Determine (a) the position of the image and (b) the magnification. (c) Describe the image. (d) Draw a ray diagram.

38-4. An object 6 cm long is placed 20 cm in front of a convex spherical mirror whose focal length is 24 cm. Determine (a) the position and (b) the size of the image. (c) Describe the image. (d) Draw a ray diagram.

38-5. A concave spherical mirror has a radius of curvature of 50 cm. A square object 3 cm on an edge is placed 10 cm in front of the mirror. Determine (a) the position of the image and (b) the area of the image.

38-6. A plane mirror is separated from a concave mirror of 50 cm radius by a distance of 100 cm. A burning candle is placed between them 20 cm from the spherical mirror. (a) Where is the image of the candle in the plane mirror? Is it real or virtual? (b) Where is the image of the candle in the spherical mirror? Is it real or virtual? (c) The plane mirror forms an image of the image in the spherical mirror found in Part (b). Where is this image? Is it real or virtual?

38-7. An object 2 cm long is placed 32 cm in front of a converging lens whose focal length is 20 cm. Find (a) the position and (b) the size of the image. (c) Describe the image. (d) Draw a ray diagram.

38-8. An object 6 cm long is placed 60 cm in front of a converging lens whose focal length is 90 cm. Find (a) the position of the image, (b) the magnification, and (c) the size of the image. (d) Describe the image. (e) Draw a ray diagram.

38-9. A convenient approximate method for obtaining the focal length of a converging lens is to measure the image distance for a distant object. What percentage error would be made if the distant object used was the window of the laboratory, about 6 m from a lens of focal length 10 cm?

38-10. An object 5 cm long is placed 20 cm in front of a diverging lens whose focal length is -10 cm. Find (a) the position and (b) the size of the image. (c) Describe the image. (d) Draw a ray diagram.

38-11. An illuminated object and screen are 6 m apart. A converging lens is placed between them so that a real image 15 times the length of the object is formed on the screen. (a) Determine the distance of the lens from the object. (b) Determine the focal length of the lens.

38-12. A lens placed 40 cm from an object forms a real inverted image 16 cm from the lens. (a) What is the focal length of the lens? (b) Draw a ray diagram.

38-13. Two thin converging lenses, each of 10 cm focal length, are spaced 15 cm apart, their principal axes coinciding. An object 6 cm long is placed 20 cm in front of the first lens. (a) Determine the position of the image in the first lens. (b) Determine the position of the final image. (c) Determine the over-all magnification. (d) Is the image real or virtual? (e) Draw a ray diagram for the system.

38-14. An object is placed 40 cm from a thin converging lens of 8 cm focal

length. A second thin converging lens of 12 cm focal length is placed 20 cm behind the first lens. (a) Determine graphically and algebraically the position of the final image. (b) Determine the magnification produced by this lens combination.

38-15. An object is placed 16 cm from a thin converging lens of 32 cm focal length. A second thin converging lens of 6 cm focal length is placed 20 cm behind the first lens. (a) Determine graphically and algebraically the position of the final image. (b) Determine the magnification.

38-16. A converging lens forms an image on a screen 60 cm from it. A thin diverging lens is interposed between them at a distance of 40 cm from the converging lens. It is now found necessary to move the screen 10 cm away from the lens in order to produce a sharp image. Determine the focal length of the diverging lens.

38-17. A carbon arc is to be focused on a screen 10 m from the arc by means of a lens of frontal diameter 2 cm and of focal length 5 cm. (a) Find the position of the lens. (b) Using data furnished in Table 36-1 for a 1,500-watt carbon arc, find the illuminance of the screen.

38-18. Repeat Problem 38-17 for a lens of 20 cm focal length and frontal diameter 2 cm. Does the lens of long or short focal length provide the brighter image when the positions of the arc and screen are fixed?

39

Optical Instruments

39-1 The Camera

The photographic camera uses a converging lens to form a real, inverted image of an object. The image is focused on a film or plate which is coated with an emulsion containing silver bromide crystals. When a few incident quanta of light are absorbed in a grain of emulsion, the grain becomes activated and developable, and when the plate is developed, the bromine is removed from each activated grain, leaving a clump of silver behind. When the plate is "fixed," the remaining emulsion is removed from the plate, so that the image is made permanent. In most cameras the converging lens consists of several elements designed and arranged to reduce objectionable aberrations to a minimum. In general, the lens is designed for a particular arrangement of image and object, as in a camera, where the object is usually much farther from the lens than is the image. A camera lens designed to minimize aberrations for an infinitely distant object may not be well corrected for "close-ups."

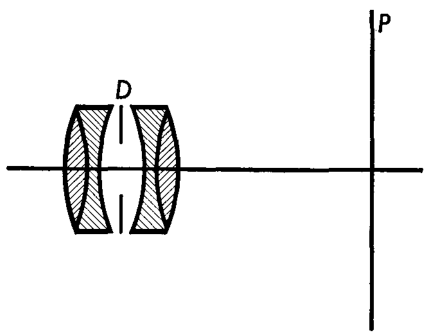


Fig. 39-1 A camera lens system.

A lens system used in some fairly good cameras is shown in Figure 39-1. Two achromats are spaced the proper distance apart, and an adjustable *iris* diaphragm is placed between them. A *shutter* is usually placed near the diaphragm to admit light for a preset time interval. The distance between the lens and film is adjusted when the camera is focused on the object to be photographed. In some box or "fixed-focus" cameras, there is no way to move the lens with respect to the film. The lens used in such a camera has a short focal length, and, as long as the object is at a distance greater

than about 6 ft from the lens, its image will be focused close enough to the principal focus so that the blurring is not severe. The designer of the box camera must be certain that all objects from about 6 ft to infinity are all pleasantly and equally blurred.

The number of developable grains of emulsion per unit area, which determines the blackness of the developed film, depends upon the number of light quanta which strike a unit area of the emulsion. This number is related to the product $E\,\Delta t$, where E is the illuminance, and Δt is the exposure time. Films are rated in arbitrary ways, as to their speed, but, in general, the higher the speed, the smaller the number of incident quanta per unit area required to produce a developable image. For convenience let us assume as a first approximation, that this relationship is a simple one, such as

$$S = \frac{\alpha}{E\,\Delta t},$$

where S is the speed of the film, and α is a constant of proportionality. From Equation (38-15) the illuminance of an image is given by $E = B\omega'$, where B is the brightness of the object, and ω' is the solid angle subtended by the lens at its image. If s' is the image distance, we have $\omega' = (\pi d^2/4)/(s')^2$, where d is the frontal diameter of the lens. Thus

$$\Delta t = \beta\,\frac{(s'/d)^2}{SB}, \tag{39-1}$$

where β is a new proportionality constant. Thus the exposure time Δt is decreased by using a film of higher speed S, by photographing a source of high brightness B, by decreasing the image distance s', or by using a lens of larger diameter d. Photographic exposure meters measure the brightness of the source and are provided with a computing scale which is correct for a distant object, in which case the image distance s' is equal to the focal length of the lens f. The number f/d is called the *f number* of a lens, generally given as $f/4.5$, or $f/2$. For other cases the indication of the exposure meter must be corrected for the proper image distance.

39-2 Projection Lantern

The projection lantern forms a real, inverted, and enlarged image on a screen which is at a great distance from the projection lens when the object is placed very close to the principal focus of the lens, as shown in Figure 39-2. Projection lanterns are widely used for the projection of motion pictures and lantern slides and for photographic enlarging. The system is focused by moving the lens with respect to the slide until a clear sharp

image is obtained. The slide must be uniformly and brightly illuminated. This is accomplished by using a very intense source of light, such as a carbon arc or a specially designed tungsten-filament projection lamp, and directing the light by means of a condenser lens so that every point of the source sends light to each point of the object and toward the projection lens. In Figure 39-2 a small section AB of the object is shown illuminated

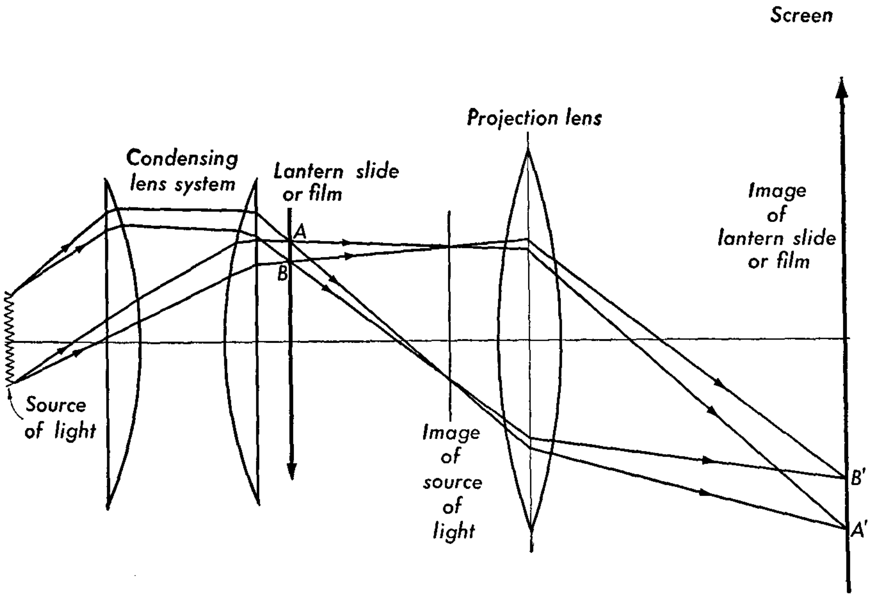


Fig. 39-2 A projection lantern.

by light from the two extreme points of the source. Rays of light from all intermediate points may be traced in a similar manner through the condensing lens system and through AB. These rays converge and form an image of the source of light at, or very close to, the projection lens. The latter should be large enough so that all of the light goes through it. The projection lens forms a real, inverted, and enlarged image $A'B'$ of the section AB. The image of every other section of the object may be traced in a similar manner. Focusing of the image on the screen is usually accomplished by moving the projection lens, keeping the object in a fixed position relative to the source of light.

The projection lantern can also be used for opaque projection, that is, for projecting pictures on opaque backings, by interposing mirrors between the condenser and the projection lens so that light is reflected properly from the opaque object.

39-3 The Eye

Many optical instruments are designed to be used in conjunction with the eye. The general structure of the eye is shown in Figure 39-3. Light enters the eye through the *cornea* C and passes through the *aqueous humor* A, of refractive index 1.35; the light then passes through the *pupil* P, the opening in the *iris* diaphragm I, and through the crystalline lens L. The lens is a lens-shaped transparent organ, suspended in the eyeball by a ring of tissue attached to the walls. The index of refraction of the lens varies from 1.388 for the outer layers to 1.411 for the inner layers. After passing through the lens, the light goes through a transparent jellylike substance called the *vitreous humor* V, of refractive index 1.336, until it strikes the *retina* R, which is the light-sensitive part of the eye. The retina is composed mainly of nerve tissue, which is connected to the brain by the optic nerve O. The image formed on the retina is real, inverted, and smaller than the object. The *optic axis* of the eye is a line passing through the vertex of the cornea and the center of the pupil.

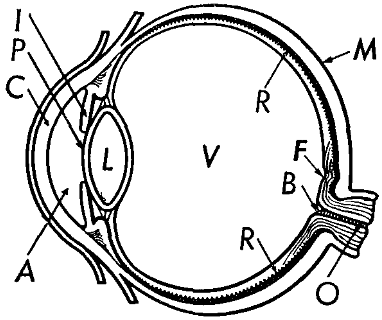


Fig. 39-3 Longitudinal section of a human eye.

The retina is a transparent membrane situated between the outer coating of the vitreous humor and the *choroid membrane* M. A section of the human retina is shown in Figure 39-4. It consists of eight fairly well defined layers; the outermost layer consists of *rods* and *cones;* the innermost layer consists of nerve fibers which lead from the rods and cones through the other layers to the optic nerve; the latter carries the impulses to the visual center in the brain. The retina is not uniformly sensitive throughout its area. At the place where the optic nerve enters the eye is the blind spot B; there is no vision when light falls on this spot. See Figure 39-3. A short distance away there is a small indentation called the *fovea* F, which is the most sensitive part of the retina. The fovea is approximately in the center of the retina and is about 0.5 mm in diameter. It contains over 30,000 cones, each with a separate nerve. There are no rods in the fovea. Rods appear in the other parts of the retina, and in the peripheral area practically only rods are present. That part of the image which falls on the fovea is seen most distinctly. Each rod or cone reacts to light as a unit. In the fovea the cones vary in size from about 0.0015 to 0.0054 mm in diameter, while the cones in the other parts of the eye are much larger. This sets a limit on the smallest details the eye can resolve. The resolving

power of the eye is about 1′ of arc. A pattern of dots or stars subtending less than 1′ of arc is seen as a blur. The field of distinct vision in the fovea is about 1°, and the rest of the eye is generally used as a means of centering

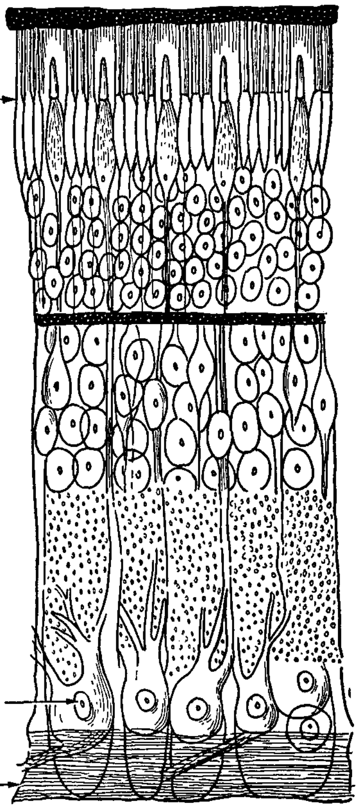


Fig. 39-4 Diagrammatic section of the human retina. Light enters from bottom of figure.

the desired image on the fovea. Thus, in reading, words are successively focused on the fovea.

The color sensitivity of the eye is associated with the cones. Incidentally, the relative response of the eye to different wavelengths is also dependent upon the intensity. Colored slides viewed in dim light sometime appear to have a different color balance than the same slides viewed in brilliant light. The rods are more sensitive than the cones and are responsible for night vision. In very dim light, colored objects are seen as shades of gray. Very faint objects can sometimes be seen by peripheral vision, where only rods are present, but they disappear when viewed directly by the cones in the fovea.

The eye lens is virtually uncorrected for chromatic aberration and suffers from spherical aberration. As the level of illumination is changed,

the sensitivity of the eye is changed, a process known as *adaptation.* The peripheral rods which are not sensitive to red light appear to control adaptation, so that a person wishing to become dark-adapted may wear red goggles in ordinary light, seeing by foveal vision, and become dark-adapted without having to wait long periods in dimly lit rooms.

Objects at various distances from the eye must form images on the retina if they are to be seen clearly. Since the image distance from the cornea to the retina is fixed, the only way this can be done is through a change in the effective focal length of the eye. This is called *accommodation* of the eye; it is accomplished by changing the shape of the crystalline lens through a change in the tension of the circular ligament supporting it.

In discussing the optics of the eye and the correction of certain defects with the use of lenses, it is common to talk about the *power* of the eye, or the power of the lens, instead of the focal length. The power of a lens is expressed in *diopters* and is defined as the reciprocal of the focal length expressed in meters. For example, if the focal length of a lens is 50 cm, that is, 0.5 m, its power is 2 diopters.

Each eye has a certain *range of accommodation.* This is the distance, measured along the optic axis, between the *near point* and the *far point* of the eye. The far point is that point on the optic axis which is sharply imaged on the retina when the eye is at its weakest power. The power of the relaxed normal eye is about 59 diopters. The near point is that point on the optic axis which is sharply imaged on the retina when the accommodation of the eye is most strongly exerted, or when the power of the eye is greatest. Eyes are classified according to the position of the far and near points, and to the range of accommodation. An object located from about 20 ft to infinity may be seen clearly without accommodation by a normal eye, for, as the object moves in from infinity, the image moves from the tips to the bases of the rods and cones.

The *emmetropic,* or normal, eye is one whose far point is at infinity and whose near point is close to the eye; this point is generally taken to be about 25 cm or 10 in.

The *myopic,* or nearsighted, eye has its far point at a finite distance from the eye; its near point is closer than in the normal eye. An eye is usually myopic because the eye is too long for its lens, or the lens is too strong for the eye. Myopia is usually corrected by the use of divergent lenses in spectacles, as shown in Figure 39-5.

The *hyperopic,* or *farsighted,* eye has a far point which is virtual; that is, its far point is situated behind the eye. The eye lens does not converge sufficiently to focus a nearby object on the retina. An eye is usually hyperopic because it is too short. Hyperopia is corrected by the use of spectacles of positive power, as illustrated in Figure 39-6.

The *presbyopic* eye is one which has lost some of its power of accommo-

dation because of the decrease in the elasticity of the lens tissue due to age. The amplitude of accommodation is very great during adolescence and is negligible beyond 60 years of age. The near point moves inconveniently far from the eye. The usual correction for presbyopia is a positive lens to be used for reading or viewing objects close to the eye.

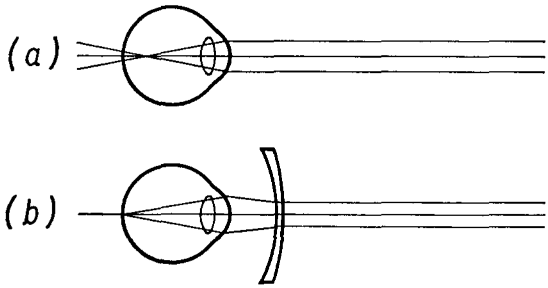


Fig. 39-5 (a) Myopia. (b) Myopia corrected by a lens of negative power.

A common defect of the eye, known as *astigmatism*, is due to the fact that the curvature of the cornea, or of the lens, is not the same in all directions about the optic axis. The cornea may be thought to have cylindrical curvature as well as spherical curvature. The astigmatic eye, viewing a collection of lines drawn out from the center of a circle, sees one group of radii sharply, while radii roughly perpendicular to these are seen less sharply. Thus, for example, the vertical lines may be clearly distinguished, while horizontal lines are less clear. In reading, an *e* may be perceived as a *c* by an astigmatic eye. Astigmatism is corrected by use of cylindrical lenses with axes properly oriented.

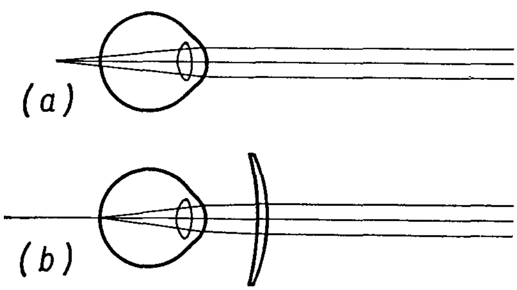


Fig. 39-6 (a) Hyperopia. (b) Hyperopia corrected by a positive lens.

39-4 The Magnifying Glass

When an object is to be examined minutely, it is usually brought as close as possible to the eye. The closer it is brought, the larger is the visual angle

α which it subtends at the eye, and the larger is its image on the retina, as shown in Figure 39-7. The image cannot be seen in sharp focus if the object is brought closer than the near point. This fact imposes a limit on the size of the retinal image and the smallness of detail which can be seen.

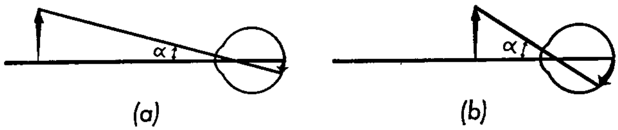


Fig. 39-7 The visual angle determines the apparent size of the object.

The visual angle can be increased and a magnified retinal image obtained with the aid of a converging lens used as a *magnifying glass*. To use the lens as a magnifier, the object is placed between the principal focus and the lens, as shown in Figure 39-8. The image is virtual and, for best

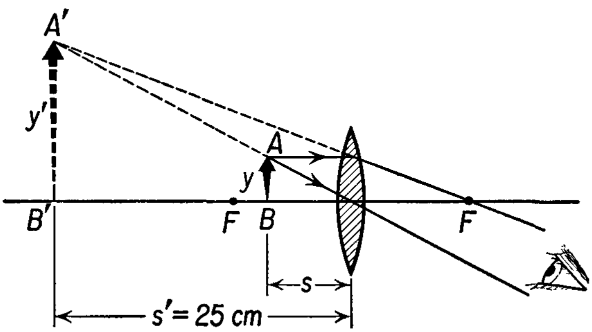


Fig. 39-8 Magnifying glass.

viewing, is formed at the near point at approximately 25 cm from the eye. If y' is the size of the image and y is the size of the object, the magnification m is given by

$$m = \frac{y'}{y} = -\frac{s'}{s},$$

where s' is the image distance and s is the object distance. From the lens equation

$$\frac{1}{s} = \frac{1}{f} - \frac{1}{s'},$$

so that

$$m = -\frac{s'}{f} + 1.$$

If we take the near point as 25 cm, we have $s' = -25$ cm and

$$m = \frac{25 \text{ cm}}{f} + 1. \tag{39-2}$$

A single lens is seldom used for magnifications greater than about 5 because of the amount of spherical and chromatic aberration produced. Two common types of eyepieces or *oculars* used in optical instruments are

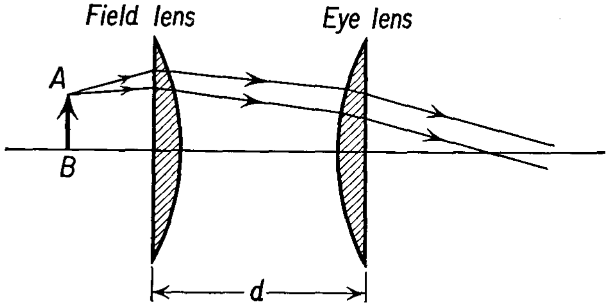


Fig. 39-9 Ramsden ocular consists of two plano-convex lenses of equal focal length separated by a distance equal to about two thirds of the focal length of either one.

the Ramsden ocular and the Huygens ocular. These are designed to reduce spherical and chromatic aberrations to a minimum while using lenses made of the same kind of glass. The Ramsden ocular, shown in Figure 39-9, consists of two converging lenses of equal focal lengths separated by two

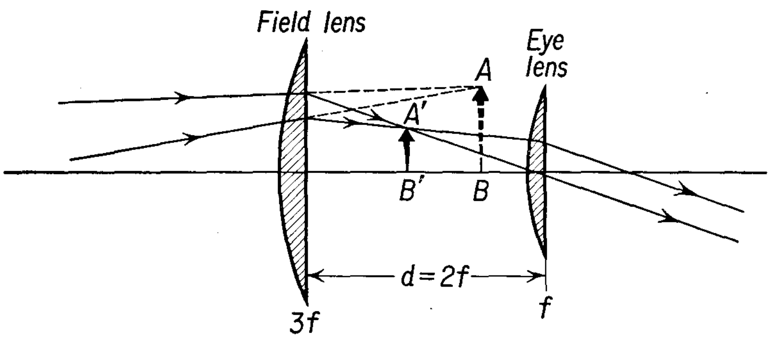


Fig. 39-10 Huygens ocular. Field lens has a focal length three times the eye lens. The distance between the lenses $d = 2f$, where f is the focal length of the eye lens.

thirds of the focal length of either. The lenses are planoconvex and are mounted so that the curved surfaces face each other. If f is the focal length of either lens, the effective focal length of the eyepiece is $3f/4$.

The Huygens ocular, shown in Figure 39-10, consists of two plano-convex lenses, the focal length of one being three times that of the other.

The two lenses are mounted with their curved surfaces facing in the same direction; the distance between the lenses is half the sum of their focal lengths. The lens with the shorter focal length is used nearer to the eye. If f is the focal length of the eye lens, the focal length of the combination is $3f/2$.

39-5 The Astronomical Telescope

In the use of an optical instrument, it is generally desirable that the image be in approximately the same position as the object, so that the focus of the eye can be shifted from the instrument to the position of the object without accommodation. Thus, in the case of the magnifier, the image is formed at

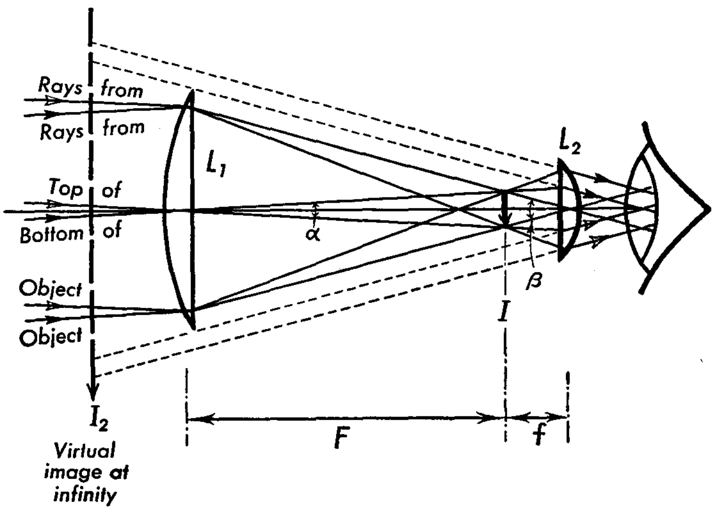


Fig. 39-11 Optical diagram of a simple astronomical telescope.

the near point, while in the case of the telescope, the image is formed at the far point. The astronomical telescope does not bring the image of the moon or of a planet closer to the eye but, rather, causes the image to subtend a larger angle. It is convenient to speak of the *angular magnification* M of the telescope as the ratio of the angle β subtended by the image to the angle α subtended by the object, thus

$$M = \beta/\alpha. \tag{39-3}$$

In its simplest form the astronomical telescope consists of two converging lenses—an objective lens L_1 and an eye lens L_2. The rays of light from one point on a distant object come into the telescope parallel to each other and inclined at a small angle α to the principle axis. These rays are focused at

a point in the focal plane of the objective. When the eye lens L_2 is placed at a distance equal to its focal length f from the principal focus of the objective, the rays from the image in the focal plane of the objective will leave the eye lens as parallel rays, as shown in Figure 39-11. In other words, parallel rays entering the objective lens leave the eye lens as parallel rays. The distance between the two lenses is $F + f$, where F is the focal length of the objective. From the figure the magnification of the telescope is

$$M = \frac{\beta}{\alpha} = \frac{F}{f}. \tag{39-4}$$

One very important function of the astronomical telescope is to collect more light than would be incident upon the retina with the naked eye. If all of the light incident upon the objective lens passes through the eye lens and is focused upon the retina, the luminous flux from a distant star is increased in the ratio of the area of the objective lens to the area of the pupil of the eye. This greatly enhances the visibility of faint celestial objects.

The image produced in the astronomical telescope is inverted, but this causes no difficulty in observing stellar objects.

39-6 The Terrestrial Telescope

Since the astronomical telescope produces inverted and reversed images, it is unsuitable for most terrestrial uses. The astronomical telescope can be modified to produce an erect image by inserting a converging lens between

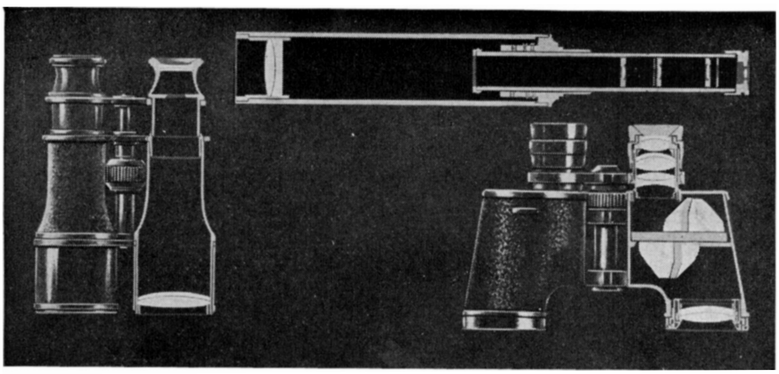


Fig. 39-12 Cut-away sections for comparison of the three types of terrestrial telescopes. (Courtesy of Bausch & Lomb Optical Company.)

the focal plane of the objective and the eye lens, as shown in Figure 39-12. Usually, terrestrial telescopes use an erecting system consisting of two con-

verging lenses with a diaphragm or stop between them to correct for spherical aberration. In the prism binocular the physical length of the telescope is shortened, and the image is erected by use of a pair of Porro prisms. Another method for producing an erect image is to make a Galilean telescope which uses a diverging lens for an eyepiece, as shown in Figure 39-13. The distance between the objective and eyepiece is $F - f$, where F is the focal length of the objective, and f is the focal length of the eye-

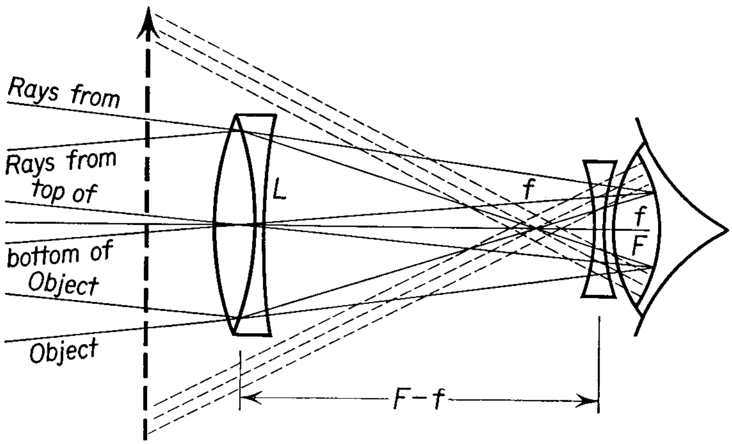


Fig. 39-13 Optical diagram of a Galilean telescope or opera glass.

piece. Parallel rays from the object are converged toward the focal plane of the objective and are deviated by the negative eye lens so that they emerge as parallel rays, forming a virtual image at infinity. Galilean telescopes are extensively used as opera glasses.

In a binocular used for daytime viewing, it is desirable that the brightness of the image through the binocular be approximately the same as the brightness of the object, so that there is no necessity for adaptation of the eye when viewing the image first through the binocular and then without its aid, as in following the flight of a bird. The average diameter of the pupil of the eye is about 5 mm. A binocular having an angular magnification of 8 and a front lens whose diameter is 40 mm (rated as an 8 x 40 binocular) gathers $(40/5)^2$ as much light as the eye but distributes this light over a retinal area 8^2 as great as the area illuminated by the unaided eye. Thus we see that an 8 x 40 binocular provides the retina with illumination comparable to the unaided eye. An 8 x 50 binocular yields an image of greater brightness, while an 8 x 30 binocular provides an image of lesser brightness than the unaided eye, provided that all the light entering the objective passes into the eye. Large-diameter objective lenses are gener-

ally used for night glasses or for viewing shaded objects rather than for general-purpose daytime observation.

39-7 The Compound Microscope

The compound microscope consists of two systems of converging lenses, shown in Figure 39-14 as single converging lenses. The objective lens L_1 has a very short focal length F, and the object is placed very close to but

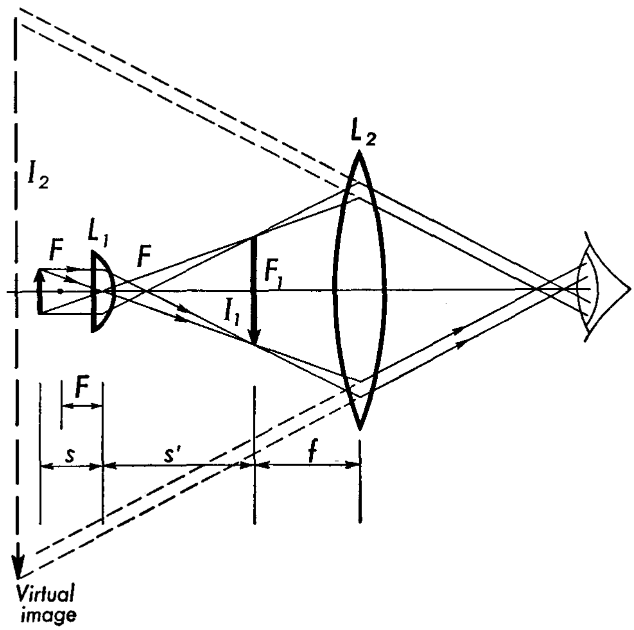


Fig. 39-14 Simplified optical diagram of a compound microscope.

just outside the principal focus of this lens. A real inverted image L_1 is formed at a distance s' from the objective lens. The eye lens I_2 is used for viewing this image. It is desirable for the image to be formed at the near point, located between the position of the specimen and the laboratory table, as shown in Figure 39-15. The magnification m_1 produced by the objective is

$$m_1 = \frac{s'}{s}.$$

To a good approximation the object distance is very nearly equal to the focal length of the objective lens F. Thus we may write

$$m_1 = \frac{s'}{F}.$$

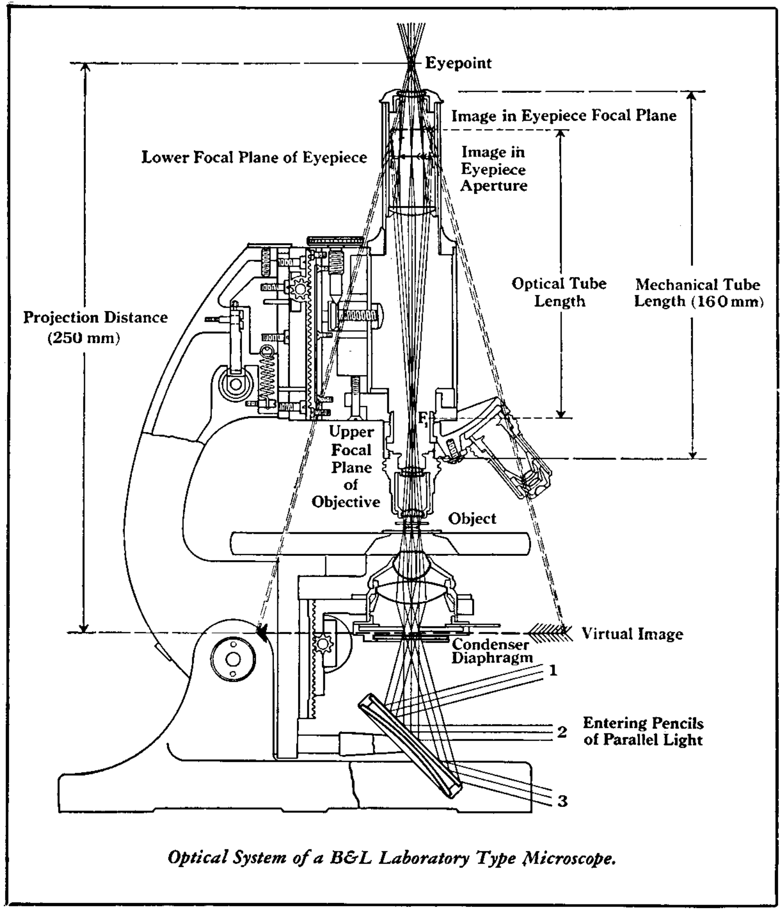


Fig. 39-15 Complete optical system of a compound microscope. (Courtesy of Bausch & Lomb Optical Company.)

Similarly, the magnification m_2 produced by the eye lens may be approximated as

$$m_2 = \frac{25 \text{ cm}}{f},$$

neglecting the 1 in Equation (39-2), where f is the focal length in centimeters, of the eye lens. The total magnification m of the compound microscope is

$$m = m_1 m_2 = \frac{s'}{F} \times \frac{25 \text{ cm}}{f}. \tag{39-5}$$

Magnifications of several hundred diameters are common with com-

pound microscopes, so that the area of the image is tens of thousands of times the area of the object. The object must be strongly illuminated to provide sufficient illumination for the image to be seen. This is usually accomplished by focusing light onto the object from a bright source by means of a mirror and a condensing lens, as shown in Figure 39-15.

39-8 The Prism Spectroscope

A prism spectroscope is used for determining the composition of the light incident upon it from a source. The light enters a narrow slit S, placed at the principal focus of a converging lens, and emerges as a parallel beam, as shown in Figure 39-16. The light is said to be *collimated* by this lens. The slit and lens are mounted at the ends of a lighttight tube called a

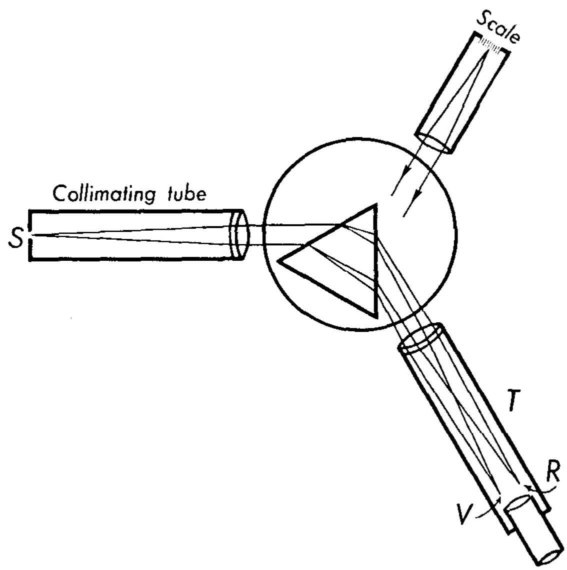


Fig. 39-16 A prism spectroscope.

collimating tube. The purpose of the collimator is to avoid astigmatism in the final beam. The collimated light is dispersed by a prism made of some suitable transparent material, such as glass, quartz, or rock salt. Rays of any small wavelength interval are deviated through nearly the same angle and emerge from the prism in a parallel beam. The telescope T can be rotated so that its axis is parallel to any one beam, and that beam is converged to the principal focus of the telescope objective. Each converged beam is an image of the slit formed by monochromatic light. Since the spectrum of a monochromatic source appears as a line, we speak of

discrete spectra as line spectra. In contrast, an incandescent body emits a continuous spectrum. The images of the slit are viewed with the aid of the telescope eyepiece. A scale in a side tube is sometimes brought into the field of view by reflecting it from one face of the prism, for purposes of measurement or calibration. A prism spectroscope can only be used for the measurement of wavelength when it has been calibrated with known spectral lines. Wavelengths cannot be measured directly with the spectroscope, for the spectroscope provides no direct means for comparing the wavelength of light with a standard of length. Such primary measurements of wavelength are made by means of interference and diffraction effects, to be discussed in the next chapter.

Problems

39-1. A lantern slide 3 in. wide is to be projected on a screen 6 ft wide at a distance of 25 ft from the projection lens. The picture should fill the whole screen. What should be the focal length of the projection lens?

39-2. (a) Determine the magnification produced by a converging lens used as a magnifier if its focal length is 4 in. (b) If this lens is used by a person whose near point is 15 cm, what is the greatest magnification he can obtain?

39-3. A person has a near point of 100 cm. What spectacle lens should he wear so that he can read newsprint at a distance of 25 cm from his eye?

39-4. The far point of a myopic eye is 50 cm. What is the focal length of the lens which should be placed in front of this eye so that it can see infinitely distant objects?

39-5. A farsighted person is prescribed lenses of $+2.00$ diopters. What is his near point?

39-6. What is the power of a converging lens which produces a magnification of $5\times$ when used as a simple magnifier?

39-7. (a) What is the focal length of a spectacle lens of -1.5 diopters? (b) What is the far point of the eye for which this spectacle lens is prescribed?

39-8. A small laboratory telescope has an objective whose focal length is 18 cm and an eye lens whose focal length is 3 cm. (a) What is the angular magnification of the telescope? (b) How far apart are the lenses? (c) If this telescope is focused on the moon, whose angular diameter is approximately $\frac{1}{2}^{\circ}$, what will be the angular diameter of the image?

39-9. A camera lens whose focal length is 3 in. has its position relative to the film adjustable for object distances from 3.5 ft to infinity. Determine the maximum displacement of the lens for these extreme positions.

39-10. A compound microscope has an objective of 4 mm focal length which forms an image 15 cm from it. The eye lens has a focal length of 2.5 cm and forms an image 25 cm from the eye lens. Determine the linear magnification of the microscope.

39-11. The distance between the objective lens and the eye lens of an astronomical telescope, when adjusted for parallel light, is 80 cm. Determine

the focal length of the lenses when the measured value of the angular magnification is 12.

39-12. A box camera of the fixed-focus type has a lens whose focal length is 4 in. and whose diameter is 0.5 in. Determine the positions of the images formed when the objects are (a) 6 ft, (b) 15 ft, and (c) 25 ft from the lens. (d) If the lens is set at the proper image distance for an object 10 ft from the lens, how large will the illuminated area be from a point source 6 ft from the lens?

39-13. The field lens of a Huygens ocular has a focal length of 9 cm, and the eye lens has a focal length of 3 cm. The lenses are spaced 6 cm apart. Determine the effective focal length of this ocular. Trace two parallel rays through this ocular, assuming (a) that they enter through the field lens and (b) that they enter through the eye lens. Determine the focal point in each case.

39-14. The lenses of a Ramsden ocular have focal lengths of 9 cm each and are placed 6 cm apart. (a) Determine the effective focal length of this ocular. Trace two parallel rays through this ocular and determine the position of the focal point.

40

Light as a Wave Motion

40-1 Wave Versus Particle

Let us recount some of the characteristics of the motion of particles and the propagation of waves, with a view toward analyzing the behavior of light. In accordance with Newton's first law, a particle moves in a straight-line path in the absence of external forces. Thus we might infer, as Newton suggested, that light is composed of particles, and that, in a continuous medium, there is no deflecting force on the light particles. At the interface between two media, light may be propagated in a straight line parallel to the interface. Thus even at an interface there is no force on the particles of light unless the light passes through the interface, and in that event the force acting must be perpendicular to the interface. To account for the fact that light is refracted toward the normal on passing from a rare to a dense medium, as from air to water as shown in Figure 40-1, it is necessary to assume that the force is directed from the rare to the dense medium. The normal component of the velocity is increased, while the tangential component remains the same, so that the refracted ray is more nearly directed toward the normal. From this we see that a particle theory of light implies that the velocity of light in the dense medium is greater than the velocity of light in the rare medium.

As we have already seen, a wave theory accounts for refraction by requiring that the velocity in the dense medium is less than the velocity in the rare medium. This particular contest between the wave and particle theories was settled by Focault in 1850 (almost 175 years after the controversy between the points of view of Newton and Huygens was initiated) by measurement of the velocity of light in water and in air. Focault showed that the velocity of light in water was less than its velocity in air, in quantitative agreement with the wave-theory explanation of the index of refraction of water.

A second aspect of the differences between particles and waves lies in their respective principles of *superposition.* When a shotgun is fired at a

target, a certain number of pellets penetrate each square centimeter of target area. When two identical guns are fired at the same target, we infer that an increased number of pellets penetrates each square centimeter of target, and that this number is approximately twice the number of pellets per unit area obtained with one gun; that is, we expect particles to obey an *arithmetic superposition principle.* At a point in a medium where the paths of two waves intersect, the medium is simultaneously displaced by the two waves, so that the resultant displacement is the vector sum of the

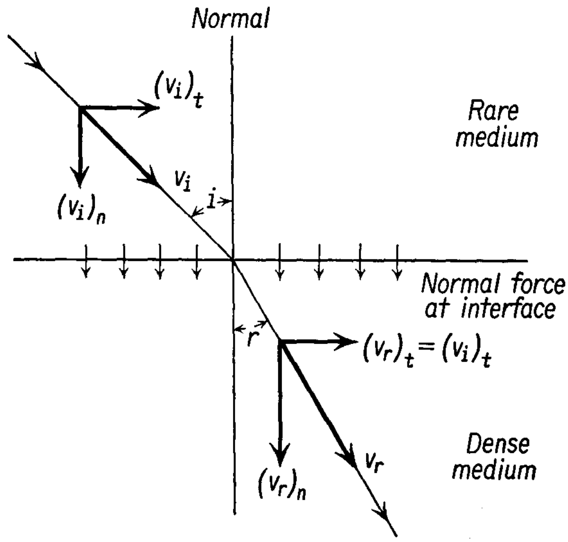


Fig. 40-1 If light were composed of a beam of particles, the velocity v_r of the refracted ray in an optically dense medium would be greater than the velocity v_i of the incident ray in the optically rare medium.

individual displacements. The magnitude of the instantaneous displacement of the medium is obtained from a *vector superposition principle.* We have already seen in Chapters 20 and 21 that the superposition principle was capable of explaining both *beats* and *standing waves.* One vibrating source at the end of a long string generated a wave in the string which caused every particle of the string to vibrate. Two identical vibrating sources at opposite ends of the string did not yield a wave of twice the original displacement at every point of the string, but, instead, nodes were produced at intervals of a half wavelength, at which there was no vibration at all.

The difference between the superposition principles appropriate to particles and to waves results in our expectation that waves must display such phenomena as beats, standing waves, and interference and diffraction

effects, under appropriate conditions, while particles do not exhibit such phenomena. In the succeeding sections we shall examine some of the interference and diffraction effects exhibited by light. These lead us to the conclusion that light is a wave motion.

40-2 Interference of Light from Two Sources

Let us suppose that two vibrating needles are immersed in a ripple tank, and that these vibrate in phase with each other and are driven at the same frequency. Each of the needles is a source of waves which spread out along the surface of the water of the tank as a series of ripples, or Huygens' wavelets. Let us represent the crest of each ripple by a solid line and the trough of each ripple by a dashed line, as in Figure 40-2. Each circular ripple spreading out from a particular source S_1 represents a disturbance which was initiated by the source at some particular time, while a circle of smaller radius represents a disturbance initiated by the source at a later time, when the phase angle of the source vibration was a different value. For example, the phase angle of any dashed circle is 180° greater than the phase angle of the next greater solid circle, concentric with it, and so on. At a point on the water surface where a crest from source S_1 intersects a crest from source S_2, the amplitude of the resulting disturbance will be the sum of the two separate disturbances. Where the crest from S_1 intersects the trough from S_2, the resulting disturbance will be of zero amplitude. Another way of saying the same thing is that any point P on the water surface where the disturbance from S_1 is always in phase with the disturbance from S_2 will vibrate with twice the amplitude due to either source alone, while at any point where the disturbance from S_1 is out of phase with the disturbance from S_2 by 180°, the disturbance will have zero amplitude. This is the two-dimensional analogue of the production of standing waves in a string, and if the two sources are the proper distance apart, that is, an odd number of half wavelengths, a series of nodes and loops will appear along the

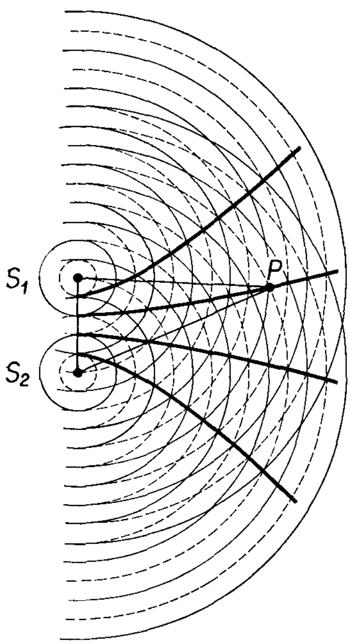


Fig. 40-2 Ripple system, showing lines of nodes.

line S_1S_2, just as in the case of the string. Not only will there be nodes where the phase difference is 180°, or π radians, but there must also be nodes appearing wherever the phase difference between the two disturbances is π, 3π, 5π, . . . , $(2m - 1)\pi$, where m is any integer.

We may find the loci of points at which a node will occur by a simple geometric consideration. If the distance S_1P is an integral number of wavelengths, the disturbance at P due to S_1 oscillates in phase with the disturbance at S_1, for when S_1 is at crest, the disturbance at P due to S_1 is at crest. If the point P is $\frac{1}{2}$, $\frac{3}{2}$, $\frac{5}{2}$, . . . , $(2m - 1)/2$ wavelengths (called a half-integral number of wavelengths) from S_2, the disturbance at P due to S_2 is 180° out of phase with the disturbance at S_2 and is therefore 180° out of phase with the disturbance at S_1, for the two sources are in phase. The point P must then be a nodal point. In fact, if the path difference $S_2P - S_1P$ is an odd half-integral number of wavelengths, the point P is a nodal point. From analytic geometry we know that the locus of points from two foci such that the sum of the distances from the foci to the locus is constant, is an ellipse, while if the difference of these two distances is constant, the locus is a *hyperbola*. Thus we must expect to find a family of hyperbolas at which there is no resultant disturbance; each hyperbola represents a definite path difference $S_2P - S_1P$ such as $3\lambda/2$, $7\lambda/2$, and so on. A photograph of a ripple tank under these conditions is shown in Figure 40-3(a). Note that the nodal lines are not straight but are appropriately curved in this small section of the interference pattern. A very similar effect can be produced by vibrating a single source close to a reflecting wall. The wall acts as a mirror which generates a virtual source on the opposite side of the barrier, and nodes are obtained between the waves from the real source and the virtual source, as shown in Figure 40-3(b).

We cannot examine the behavior of light waves in transit the way we can observe water waves, for the light waves are invisible until they interact with a screen. Since light waves are emitted by independent atoms and molecules which are not in phase with each other, we cannot generate interference effects by setting two sources of light side by side. We may simulate two sources by making use of Huygens' principle, for if a wave is incident upon a slit, each point of the slit acts as a source of wavelets, as was shown in Figure 20-11(a) in connection with our discussion of the diffraction of a wave. In the event that two adjacent slits are placed in a barrier, these slits act as sources, and an interference pattern is produced, as shown in Figure 40-3(c). The light from each atom of a source spreads out as a spherical wave. If a double slit is placed in the vicinity of the source so that the source is equidistant from each slit, the light from each atom arriving at the slits is in phase. When the light from the pair of slits intercepts a screen, we must expect to find adjacent bright and

dark regions on the screen, if light is a wave motion. The dark regions correspond to the positions of nodes, while the bright regions are the positions of antinodes.

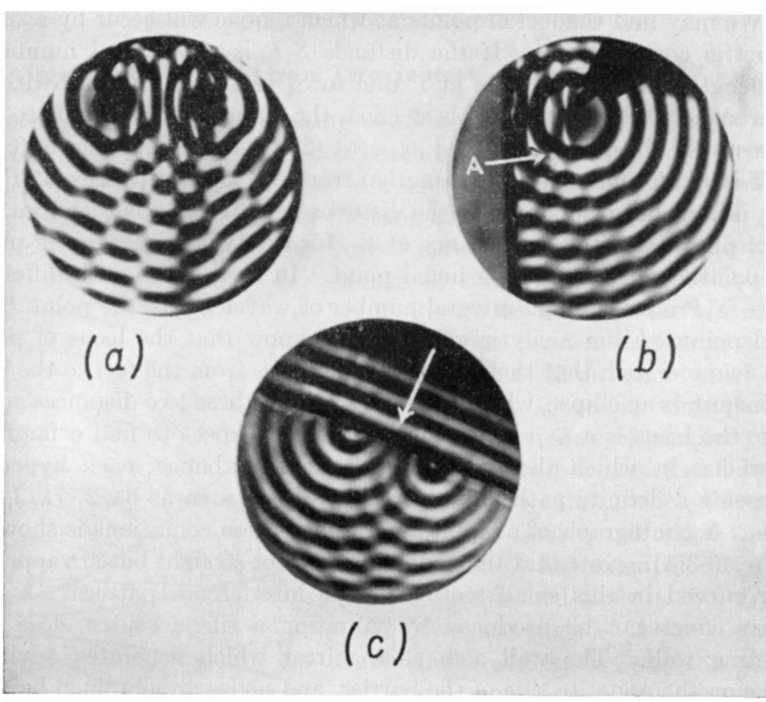


Fig. 40-3 (a) Nodal lines produced by interference of ripples from two sources. (b) Nodal lines produced by interference of ripples from a point source and its virtual image. (c) Nodal lines produced by interference of ripples from two slits. (Photographs courtesy of The Ealing Corporation.)

In Figure 40-4 monochromatic light from a source is limited by a slit S equidistant from two narrow slits S_1 and S_2. These two slits act as new sources which emit light of the same frequency and phase. A point P on the screen will be bright when the light reaching it from the slits S_1 and S_2 is in phase; that is, the light path from S_2 to P must exceed the light path from S_1 to P by an integral number of wavelengths, or, if m is an integer,

$$S_2P - S_1P = m\lambda.$$

This path difference may be found on the figure by swinging an arc of radius S_1P to intersect the line S_2P at A. The path difference between the two rays is S_2A. In the case of small angles we may approximate the arc by the chord S_1A, and we may assume that the chord is perpendicular

to the path S_2P at A. From the figure we find

$$\sin\theta = \frac{S_2A}{S_1S_2} = \frac{m\lambda}{d},$$

where d is the distance between the slits, and θ is the angle describing the position of P with respect to the center of the slits. Similarly,

$$\tan\theta = \frac{OP}{CO} = \frac{x}{L},$$

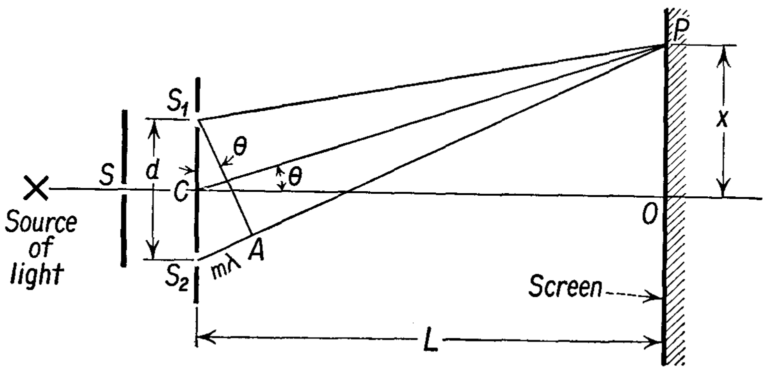


Fig. 40-4

where x is the distance from the central image to P, and L is the distance CO from the slit to the screen. For small angles $\sin\theta = \tan\theta$, so that

$$\frac{m\lambda}{d} = \frac{x}{L},$$

or,

bright fringes:
$$x = m\frac{\lambda L}{d}. \tag{40-1}$$

We may locate the position of the dark fringes on the screen by observing that the path difference for P to be a node is given by $S_2A = (2m - 1)(\lambda/2)$, an odd number of half wavelengths. Thus

dark fringes:
$$x = \frac{(2m-1)}{2}\frac{\lambda L}{d}. \tag{40-2}$$

To get an idea of the order of magnitude of the quantities involved, suppose that a sodium lamp emitting monochromatic yellow light illuminates a pair of slits separated by a distance of 0.05 cm, and we find that the first bright line is displaced a distance of 0.24 cm from the central image (for which $m = 0$) when the screen is 200 cm from the slits. We find, from

Equation (40-1), that

$$\lambda = \frac{xd}{L} = \frac{0.05 \text{ cm} \times 0.24 \text{ cm}}{200 \text{ cm}} = 6 \times 10^{-5} \text{ cm}.$$

Such an experiment makes it possible to measure the wavelength of light with a ruler. A two-slit interference pattern produced with light from a mercury arc is shown in Figure 40-5.

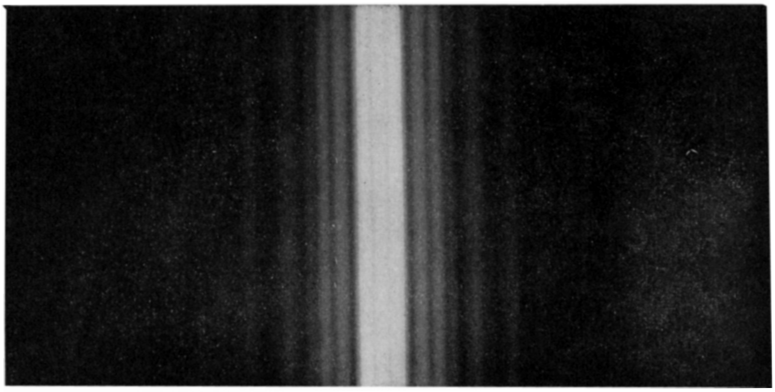


Fig. 40-5 Photograph of the interference pattern produced by the passage of light from a high-pressure mercury arc through two slits which are close together. No lens was used; the light was allowed to fall directly on a photographic plate. (Courtesy of Central Scientific Co.)

If the slits are illuminated with white light, each color will produce its own set of interference bands, and these will overlap. The central image is produced by light of all wavelengths which are in phase so that the central image is white and bright. Series of colored interference bands appear on either side of the central image. These interference bands may be easily seen by viewing light from a distant source, such as a street lamp, through two pinholes punched in a piece of cardboard. It is interesting to note that if we limit our conception of light to the postulates of geometric optics, we must infer that light has passed through the opaque center of the slits in the two-slit experiment, as in Figure 40-4.

40-3 Interference from Thin Films

Let us consider the effect of a thin parallel plate of transparent material upon a light beam normally incident upon the plate, as shown in Figure 40-6. The beam is partially reflected and partially transmitted by each of the two interfaces. The reflected light is made up of two parts, some reflected from the first surface, and some reflected from the second surface

and transmitted back through the first surface. If these interfere *constructively*, that is, if the two parts are in phase or in step, the reflected beam will be bright, while if the two beams are out of phase by $180° + m(2\pi)$, or are out of step by an odd number of half wavelengths, the two beams

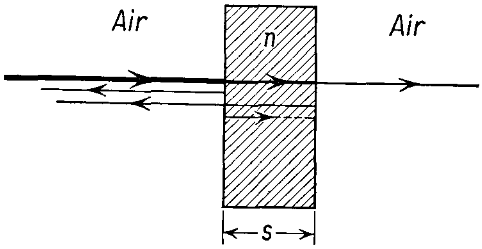


Fig. 40-6

will interfere *destructively*, and the reflected light will be of zero intensity. The phase difference between the two beams will be due in part to the path difference between the two beams and in part to phase changes which take place on reflection.

Let us suppose that two pieces of string of different masses per unit length are joined together, and that the two are under tension. In accordance with Equation (20-10), the velocity of propagation of a wave in a string is given by $v = (S/m)^{1/2}$, where S is the tension, which is the same in both sections of the string, and m is the mass per unit length. The velocity of propagation of a wave is different in the two strings, and if a wave is initiated in one of them, it is partially transmitted and partially reflected at the interface, or at the point where the strings are joined. Experimentally we find that the wave is transmitted without a change of phase, no matter in which string the wave starts. When the wave approaches the interface from the denser string, in which the wave velocity is smaller, the wave is reflected without a change of phase. However, when the wave approaches the interface from the lighter string, in which the wave velocity is greater, the reflected wave undergoes a change of 180° of phase with respect to the incident wave, as shown in Figure 40-7. These results are consistent with our previous discussion in Section 20-4, where we saw that a wave reflected from a barrier, which we may conceive of as a very heavy string, changes its phase on reflection by 180°. At normal or nearly normal incidence, these phase changes also occur optically. Thus when light is transmitted from one medium to another, there is no phase change; when light coming from a rare medium is reflected from an interface with a denser medium, it experiences a phase change of 180°; when light coming from a dense medium is reflected from an interface with a rarer medium, it experiences no phase change on reflection, the phase of the reflected wave being the same as the phase of the incident wave.

When a string or any other medium is set into vibration at a certain frequency f, the particles of the medium respond at this frequency. Thus when a wave progresses from one medium to another, the frequency remains the same. Since the speed of the wave changes as it goes from one

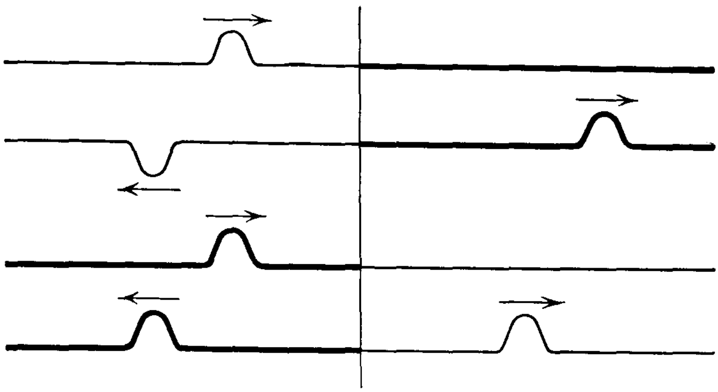


Fig. 40-7

medium to another, the wavelength must change. If λ_1 is the wavelength in a medium in which the velocity is v_1, and λ_2 is the wavelength in the medium in which the velocity is v_2, then

$$\frac{v_1}{v_2} = \frac{f\lambda_1}{f\lambda_2} = \frac{\lambda_1}{\lambda_2}.$$

From Snell's law the index of refraction is

$$n = \frac{v_1}{v_2}; \tag{37-4}$$

therefore

$$n = \frac{\lambda_1}{\lambda_2}. \tag{40-3}$$

In studying the interference effects in thin films, the effect of the distance traversed in the medium upon the difference in phase between the ray reflected at the first surface and the ray reflected at the second surface is most easily accounted for by expressing that distance in wavelengths. But here we must be careful to use the wavelength of the light in the medium in which the light is moving rather than the wavelength in vacuum. If s is the thickness of the film, the light traverses a distance $2s$ in its passage into and away from the second reflecting interface. The difference in phase between light reflected from the second interface and light newly incident upon the first interface, because of the path

difference, is $2s\frac{2\pi}{\lambda'}$, where λ' is the wavelength in the film. In addition, if the film is immersed in a medium of different refractive index, either the wave reflected at the first interface or the wave reflected at the second interface is shifted in phase by reflection, by 180° or π. Thus the total phase shift $\Delta\phi$ between the two beams is

$$\Delta\phi = 2s\frac{2\pi}{\lambda'} + \pi.$$

If the film is an air wedge between two glass plates, the wavelength in the air film is the same as that measured by a spectrometer, and we may drop the prime from λ', yielding

air film:
$$\Delta\phi = 2s\frac{2\pi}{\lambda} + \pi. \tag{40-4}$$

If the film is a wedge of material of refractive index n, the wavelength λ' may be obtained from Equation (40-3) as

$$\lambda' = \frac{\lambda}{n},$$

and we have

dielectric film:
$$\Delta\phi = 2ns\frac{2\pi}{\lambda} + \pi. \tag{40-5}$$

When the phase difference between the two reflected beams is an integer times 2π, the waves reinforce each other, but when the phase difference is an odd integer times π, the waves interfere destructively, and no light is reflected from the film.

The interference produced by thin films has been used to reduce the loss of light by reflection which takes place at the surfaces of lenses in optical instruments. A thin film of lithium fluoride or calcium fluoride of thickness about one quarter the wavelength of sodium light is deposited on the lens. The refractive index of this material is intermediate between that of the air and the glass of the lens. On the way into the lens, the light is reflected from a denser medium, so that a phase change of π is produced at the air-fluoride interface, and the same phase change is produced at the fluoride-glass interface. Thus the total phase difference between the two reflected rays is entirely due to the path difference. This is made equal to half a wavelength to make the rays interfere destructively. White light is made of many wavelengths, so that reflection is not eliminated at all wavelengths. The deposited film is generally appropriate for yellow light, so that coated lenses look purple (white minus yellow) by reflected light.

Interference bands can easily be produced by making a wedge-shaped

air film between two plane pieces of glass, as shown in Figure 40-8(a). When this wedge is illuminated with monochromatic light, a series of parallel bright and dark lines will be observed if the two glass surfaces A and B are perfectly plane. If one glass surface is known to be plane, as in an *optical*

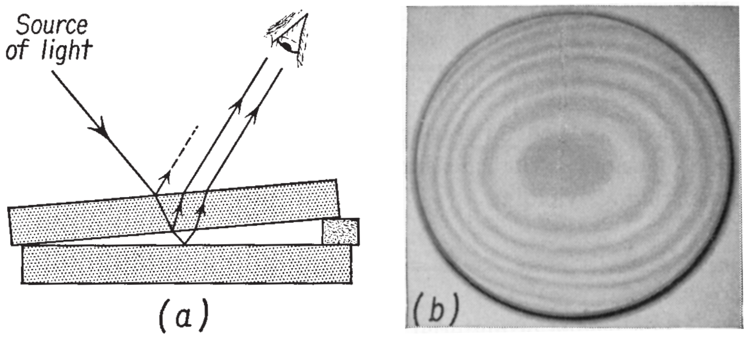


Fig. 40-8 (a) Thin air film between glass plates. (b) "Contours" or interference pattern produced in testing the surface of a telescope objective; the oval rings indicate that the surface is not quite spherical. (Photograph courtesy of Bausch & Lomb Optical Co.)

flat, the flatness of the second surface may be determined by observing the interference pattern produced when a wedge-shaped film of air is set up between them. The interference pattern then takes on the appearance of a contour map, as shown in Figure 40-8(b) with the dark lines representing the loci of equal thickness of the air film.

Newton observed and studied the interference produced by a thin film of air between the convex surface of a lens and a plane piece of glass,

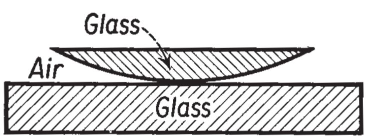


Fig. 40-9 Apparatus for producing Newton's rings.

as shown in Figure 40-9. Because of the circular symmetry of the arrangement, a series of bright and dark rings may be seen when the system is illuminated with monochromatic light. These are called *Newton's rings.* Thomas Young first explained these rings on the basis of the wave theory of light and explained why the central spot should be dark on the basis of the change in phase on reflection. He reproduced Newton's experiment by using an oil film between a crown-glass lens and a flint-glass plate. The oil had an index of refraction intermediate between the crown glass and the flint glass. In this case the central spot was white and the other bright and dark regions were shifted in a corresponding manner.

When a thin, wedge-shaped film or a Newton's ring apparatus is illuminated with white light, colored fringes are formed as a result of the destructive interference of first one wavelength, then another. The pattern of Newton's rings formed with monochromatic light is shown in Figure 40-10.

The colors observed in soap bubbles, oil slicks, butterfly wings, and even some minerals are associated with the interference effects from thin films rather than with pigments.

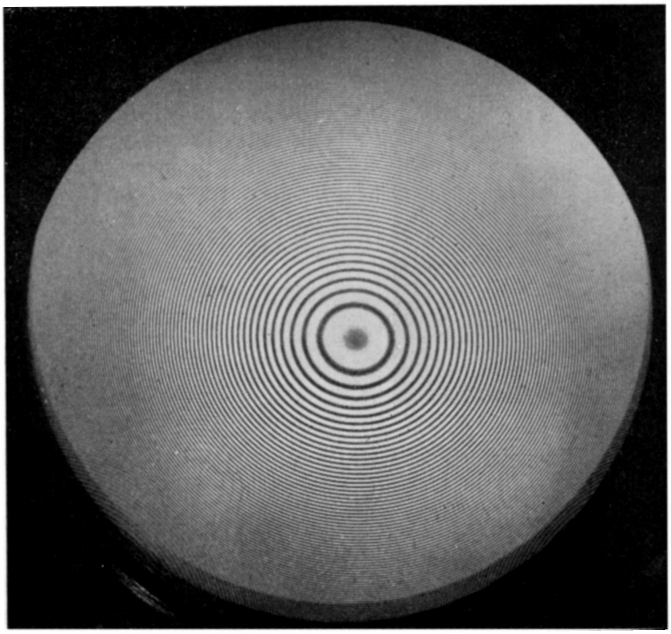


Fig. 40-10 Newton's rings; pattern produced with sodium light. (Courtesy of Bausch & Lomb Optical Company.)

40-4 Diffraction of Light

The terms "interference" and "diffraction" are used rather loosely. In general, "interference" is used to describe wave effects involving a limited number of rays, as in the two-slit interference pattern, while "diffraction" is used when a broad wave front is limited by a barrier or an aperture and an infinite number of elements of the wave front must be considered. We have already noted that, in the case of sound, the waves bend around corners and also spread out in passing through a narrow aperture; that is, sound waves are *diffracted*. Since interference experiments show definitely that light is propagated as a wave motion, we must expect to find diffraction

effects associated with light. As in the case of the two-slit interference pattern, we shall find that the angular aperture in which diffraction effects are observed is associated with the parameter λ/d, where d is the linear dimension of the obstacle. In our daily experience we observe diffraction effects with sound, for the dimensions involved are of the order of the wavelength of sound. Diffraction effects are not commonly observed with

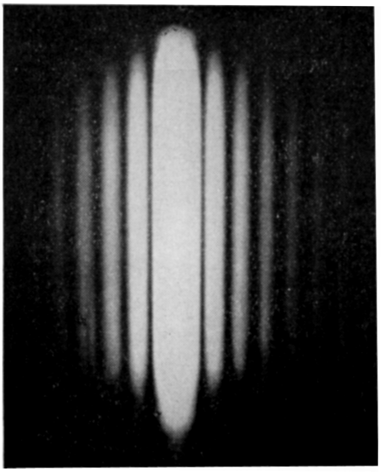


Fig. 40-11 Photograph of a diffraction pattern produced by the passage of light through a narrow slit. (Reproduced by permission from *College Physics*, 2nd ed., by Sears and Zemansky, 1952; Addison-Wesley Publishing Company, Inc., Reading, Mass.)

light because the wavelength of light is much smaller than the dimensions of common objects.

When a beam of light from a distant source passes through a narrow slit and falls upon a screen at some distance from it, the pattern on the screen consists of a bright image of the slit and a series of bright and dark fringes on either side of the central bright fringe, as shown in Figure 40-11. Only a small portion of the incident wave front passes through the narrow slit to produce this diffraction pattern. The appearance of the bright and dark regions on the screen can be explained by assuming that each point of this section of the wave front acts as a source of light. The slit is divided into imaginary elements. The amplitude and phase of the disturbance generated at the screen by each of these elements is computed, and the resultant disturbance is determined by the superposition principle. Since both amplitude and phase must be taken into account, the resultant is computed by vector methods, similar to those used in finding the voltage in an a-c circuit.

Imagine AB of Figure 40-12 to be the edges of the slit, greatly magnified, and the wave front approaching it to be a plane monochromatic wave, so that each point of the wave front incident upon the slit is vibrating in the same phase. Point C is the center of the slit, and CO is a perpendicular

line from the slit to the screen. We may locate the dark fringes by dividing the slit into an even number of elements such that the light from one element reaching the screen just cancels the light from an adjacent element.

Let us divide the slit into two equal parts, AC and CB, and consider the conditions under which a point D on the screen is dark. If the point D is so located that the light from a small element i near A is 180° out of phase with the light from a small element i' near C, these two elements

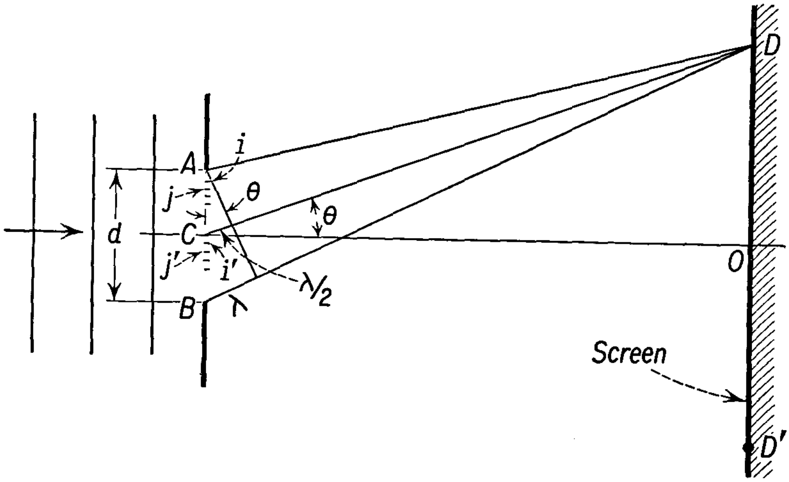


Fig. 40-12

together will contribute no illumination to the screen. In the same way the pair of elements j and j' will interfere destructively at D, and so on. This will be the case when the path differences $CD - AD = \lambda/2, \ldots,$ $BD - CD = \lambda/2$, for each pair of elements across the slit. Thus the path difference between the light reaching D from the bottom of the slit and the light reaching D from the top of the slit is equal to a wavelength. We have

$$BD - AD = \lambda,$$

and, using the same small angle approximations we used in our analysis of the two-slit interference effect, we find

$$\sin\theta = \frac{\lambda}{d} \tag{40-6}$$

for the location of a dark band, or fringe. Clearly, a dark fringe also appears at the symmetrical point D'.

To find the location of the second dark band, we divide the slit into four equal parts and repeat the above argument for the top two parts and

for the bottom two parts separately. Thus we observe that the second dark fringe will occur when the path difference between the top and bottom of the slit is 2λ. In general, the mth dark fringe is found at the screen when

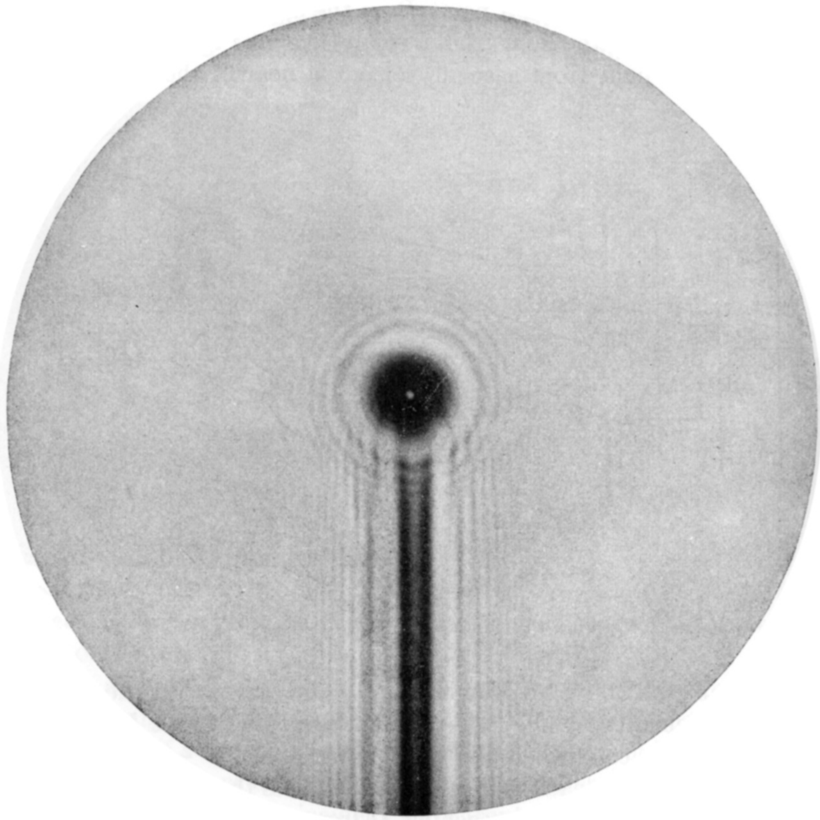


Fig. 40-13 Photograph of the shadow of a ball bearing supported on a rod. Note the diffraction pattern around the rod and ball bearing and the bright spot in the center of the shadow of the ball bearing. (Reproduced by permission from *College Physics*, 2nd ed., by Sears and Zemansky, 1952; Addison-Wesley Publishing Company, Inc., Reading, Mass.)

the path difference between light reaching the screen from the top and bottom of the slit is $m\lambda$, or, from Equation (40-6), we have

mth dark fringe: $$\sin\theta = m\frac{\lambda}{d}\cdot \tag{40-7}$$

Thus the diffraction pattern of a single slit consists of a central maximum with alternating dark and bright fringes on either side of the central region.

Thus we note that the parameter λ/d is characteristic of the width of the diffraction pattern.

The diffraction pattern produced by plane waves incident upon a circular aperture consists of a bright central disk surrounded by fainter circular rings. If d is the diameter of the aperture, it can be shown that the angle θ subtended by the radius of the first dark ring is given by

$$\theta = 1.22\frac{\lambda}{d}\cdot \tag{40-8}$$

When plane waves of light are limited by an obstacle, such as a disk or ball bearing, rather than by an aperture, waves diffracted from the edge of the obstacle reach a point in the center of the shadow in phase with each other, for the center of a circular shadow is equidistant from all points on the rim. A bright spot will be found at the center of a circular shadow, as though light passed directly through the obstacle, as shown in Figure 40-13.

40-5 Diffraction and Resolving Power

Diffraction phenomena can also be observed when light passes through large apertures, such as the lenses of microscopes and telescopes. The effect of such phenomena is to limit the *resolving power* of the instrument; that is, the ability of the instrument to show increasingly greater detail at higher magnifications. If light from a point source is focused by a converging lens, the image will not be a sharp point even if the lens has been perfectly corrected for all aberrations. The very best image which can be formed will be determined by the diameter of the lens opening, according to Equation (40-8), as shown in Figure 40-14.

The size of the diffraction pattern will be determined by the wavelength of the light used, the diameter of the lens, and the focal length of the lens used. Thus the best astronomical telescope forms the smallest image of a star. If we consider two points sending light through an optical system, the image of each point will be a diffraction pattern. If the points are close together, these patterns may overlap, so that it will not be possible to distinguish them as two separate points. The images are said to be resolved if the dark ring of one pattern passes through the center of the other pattern, or if the two central disks are separated a distance equal to the radius of one of them. If two points cannot be resolved by an instrument, merely increasing the magnification serves no useful purpose, for one simply obtains a larger, fuzzier image. These considerations provide us with a fundamental limit on the magnification of a microscope or a telescope. In practice, the smallest separation of two point sources which can be resolved by a microscope with visible light is about 1.8×10^{-5} cm, and the highest magnifications used are generally less than 2,000 $\times$. To obtain higher

resolving power, it is necessary to use shorter wavelengths. The 200-in. reflector at Mt. Palomar can distinguish between two stars which are separated by 1.3×10^{-7} radian or 2.6×10^{-2} sec of arc. This may be compared to the eye whose pupil has a diameter of about 3 mm, so that the theoretical resolving power is such that the eye might resolve objects separated by 47 sec of arc. Actually, the average person can distinguish objects separated by about 1 min of arc.

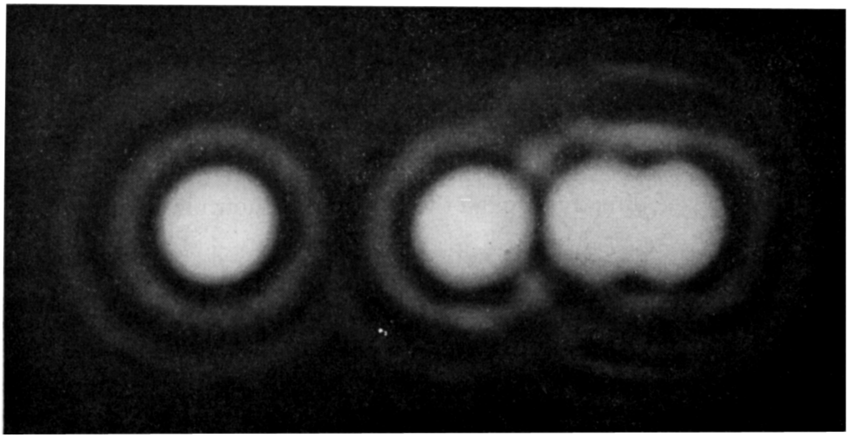


Fig. 40-14 Photograph of the diffraction patterns of light produced by a lens. Four point sources of light were used. The two patterns on the right can just be resolved as due to two sources. (Reproduced by permission from *College Physics*, 2nd ed., by Sears and Zemansky, 1952; Addison-Wesley Publishing Company, Inc., Reading, Mass.)

Identical considerations apply to the design of radar reflectors and the reflectors used in radiotelescopes which are now being used to study sources of radio noise and line emission of microwaves from the sun and other cosmic sources. An antenna having a reflector of 60 ft diameter must be used with 21-cm radiation to resolve the sources of radiation to an angle of about 1°. This particular wavelength is emitted by neutral hydrogen atoms in the hydrogen clouds of the galaxy. The largest radiotelescope now in operation in the United States is the 60-ft paraboloid of Harvard University, shown in Figure 40-15, while reflectors three and four times this diameter are under construction.

40-6 The Diffraction Grating

The *diffraction grating* is widely used for the measurement of the wavelength of light and for spectrum analysis. Diffraction gratings are used as reflection gratings or as transmission gratings. A reflection grating consists of a

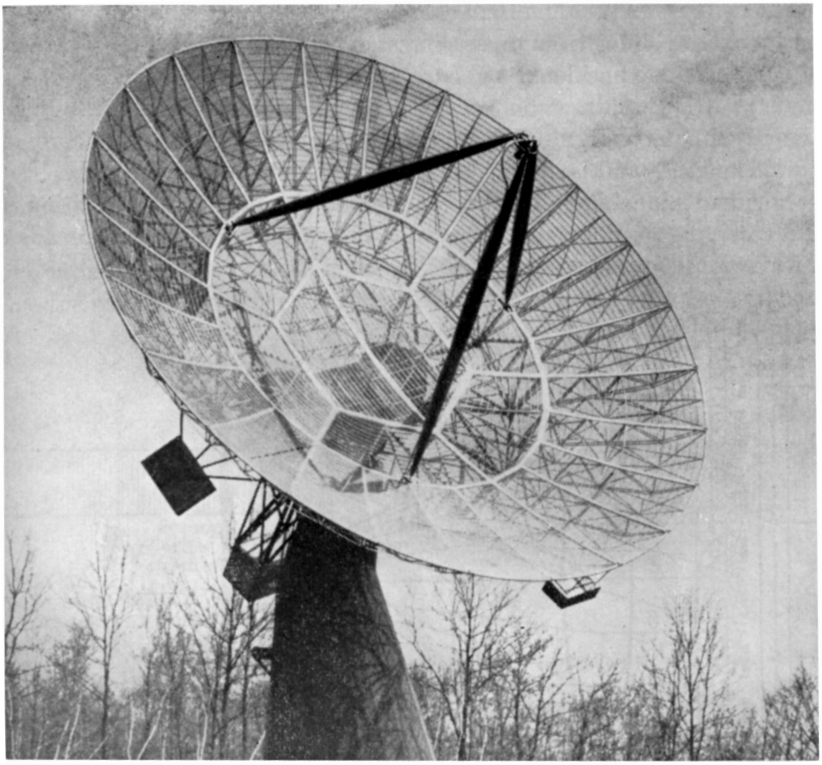


Fig. 40-15 The George R. Agassiz 60-foot Radio Telescope of Harvard College Observatory at the G. R. Agassiz Station, Harvard, Mass. (Photo by Robert D. Cox, Sky Publishing Corp; reproduced by permission of Harvard College Observatory.)

series of parallel rulings or scratches made on a polished reflecting surface. The number of rulings varies from about 400 per centimeter in some gratings to about 6,000 per centimeter in other gratings. A transmission grating has a series of parallel rulings made on a flat glass surface. The light is transmitted through the spacings between the scratches. Good gratings are difficult to prepare, and, for ordinary purposes, replicas are used. These replicas can be made by pouring a solution of collodion in ether over a ruled grating. After the ether has evaporated, the collodion layer is stripped off and is cemented to one side of an optical flat. The collodion retains the impression and acts as a fairly good diffraction grating.

To understand the action of a grating, let us consider a set of plane parallel waves incident on a transmission grating, as shown in Figure 40-16. The parallel light is generally produced by placing the light source in front of the slit of a collimating telescope. The spaces between the rulings can be considered as a series of equally spaced narrow slits, a few of which are

shown in the figure. The light which passes through the grating can be considered as coming from these slits, and, according to Huygens' principle, the slits can be considered as sources of waves. These waves will be of circular section, as drawn in a plane perpendicular to the rulings. For the sake of simplicity, let us assume that the incident light is monochromatic of wavelength λ and is directed normal to the plane of the grating. Since the incident plane wave front is parallel to the plane of the grating, the light emerging from the slits at any one time is all in phase and spreads out in wave fronts of circular section from each slit as center.

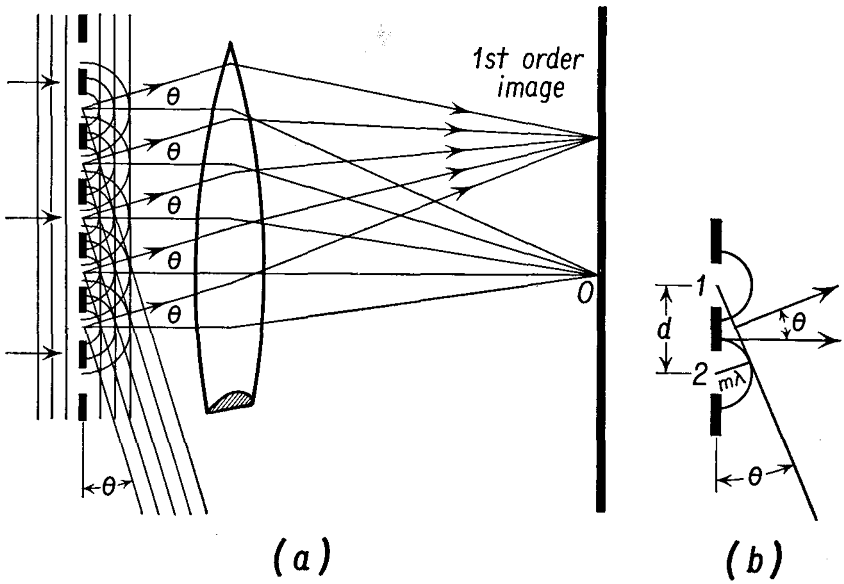


Fig. 40-16 (a) Action of a diffraction grating on a parallel beam of monochromatic light. The relative dimensions of the grating and the lens are drawn out of correct proportion in order to illustrate the effect of the lens on the diffracted beam. (b) The relationship between the grating space d, the wavelength λ, and the angle of diffraction θ.

When parallel light composed of plane wave fronts is incident upon a lens, it is brought to a focus at a point in the focal plane of the lens. If the incident parallel light is everywhere in phase across a plane perpendicular to the direction of propagation of the light, it will be in phase at the focus, and a bright spot will be produced. Let us consider the phase relationships between light from adjacent slits incident upon the lens in various directions making an angle θ with the normal to the grating.

Light leaving the grating in the direction of the normal is in phase in a plane perpendicular to its direction of motion. This light will be brought to a focus by a converging lens and will produce what is known as the

central image at O, in the focal plane of the lens and on the principal axis of the lens. Light leaving the grating at an angle θ with the normal to the grating will interfere constructively, provided that the path traversed by light from adjacent slits differs in length by an integral multiple of λ. From Figure 40-16(b) it can be seen that constructive interference will occur when

$$\sin\theta = \frac{m\lambda}{d}, \tag{40-9}$$

where m is an integer, and d is the distance between adjacent slits. The image for which $m = 1$ is called the first-order image, and so on. Thus if violet light and red light are incident on the grating, the first-order image

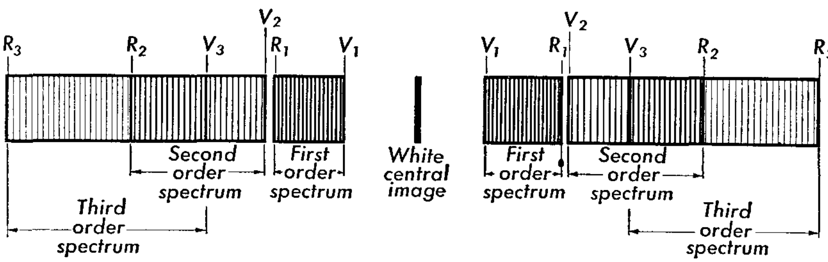


Fig. 40-17 Relative positions of the first three orders of spectra produced by a diffraction grating on either side of the central white image. Notice that the second and third orders overlap considerably.

of violet light will be deviated through a smaller angle than that of red light, since the wavelength of violet is shorter than that of red light. If white light is incident on a diffraction grating, a series of continuous spectra will be obtained on each side of the central image. The central image itself will be white, since all of the wavelengths from the source are focused in it. In any one order, the spectrum produced by a diffraction grating has the colors in the reverse order from that produced by a prism. In a diffraction-grating spectrum the violet is deviated least, while in a prism spectrum the red is deviated least.

The diffraction grating is often used as the dispersing element in a spectrometer, in place of a prism, producing images of several orders, as shown in Figure 40-17. Because the wavelength of visible light ranges from about 3,800 A to about 7,500 A, a factor of slightly less than 2, there is a break between the first-order spectrum and the second-order spectrum, but the third-order spectrum overlaps the second-order spectrum.

The prism spectroscope has one important advantage over the diffraction-grating spectroscope in that all the energy which passes through the prism is concentrated in a single spectrum. In a diffraction-grating spectroscope the energy from the source of light is spread over several

orders, and a large fraction of this energy is concentrated in the zero order, or central image. On the other hand, the diffraction grating provides a direct means of measurement of wavelength from the measurement of an angle and the spacing between the rulings of the grating. The dispersion produced by the grating can be calculated from Equation (40-9), while the dispersion of a prism does not follow a simple law. For most optical glass the dispersion is much greater in the violet region than in the red region of the spectrum.

Problems

40-1. Yellow sodium light whose wavelength is 5,893 A comes from a single source and passes through two slits 1 mm apart. The interference pattern is observed on a screen 175 cm away. (a) How far apart are two adjacent bright bands? (b) Taking the distance between the first minima on either side of the central maximum as the "shadow" of the solid region between the two slits, how wide is this shadow?

40-2. Light from a mercury arc is passed through a green filter and then falls upon two narrow slits 0.06 cm apart. The interference pattern is formed on a screen 250 cm away. The distance between two adjacent green lines is found to be 2.27 mm. Determine the wavelength of the light.

40-3. Five per cent of the incident light striking a glass-air surface is reflected back. What percentage of the incident light is transmitted after passage through an optical system containing eight surfaces?

40-4. Calculate the thickness, in centimeters, of a nonreflecting film of refractive index 1.40 to be used for coating a glass plate. Assume that sodium light is incident upon it.

40-5. Prove that if an object of thickness T is placed at one edge of a glass plate, and a second plate is placed atop the first so as to make an air wedge, then the number N of dark lines produced in light of wavelength λ is $N = 2T/\lambda$.

40-6. Calculate the number of dark lines that will be produced when green light of wavelength 5,461 A is incident normally upon a wedge-shaped air film produced by inserting a piece of steel 0.02 cm thick between two glass plates, at one end.

40-7. Two glass plates 10 cm long are in contact at one end and are separated by a thin sheet of paper at the other end, forming a wedge-shaped air film. Red light of wavelength 6,600 A is incident normally upon the glass. Experiment shows that there are 17 dark lines per centimeter. (a) Calculate the thickness of the paper. (b) How many lines per centimeter would be produced by green light of wavelength 5,400 A?

40-8. In a Newton's ring apparatus a planoconvex lens is placed upon a flat glass plate, convex side down. The diameter of the first dark ring is observed to be 0.1 cm. What is the radius of curvature of the lens? Monochromatic light for which $\lambda = 6{,}000$ A is used in normal illumination. [HINT: If r is the radius of the Newton ring, s the thickness of the air at this position, and R the

radius of curvature of the lens, then $r/s = (2R - s)/r$, and, to a first approximation, $r^2 = 2Rs$.]

40-9. Light from a distant source is incident normally upon a single slit. The wavelength of the incident light is 5,893 A. It is found that the width of the central maximum on a screen located 1 m from the slit is 0.1 cm. What is the width of the slit?

40-10. What must be the diameter of the reflector of a radar antenna in the form of a parabolic mirror if the antenna is to separate two airplanes 1° apart? The wavelength to be used is 3 cm.

40-11. A diffraction grating has 6,000 lines per centimeter. White light is incident on the slit of a diffraction-grating spectrometer so that the collimated beam falls normally on this grating. At what angle will the blue light of 4,400 A wavelength be found (a) in the first order and (b) in the second order?

40-12. In the grating of Problem 40-11, at what angle will the red light of 7,200 A wavelength be found (a) in the first order and (b) in the second order? (c) What is the highest order spectrum in which this red light will be found?

40-13. Blue light of 4,500 A is used to determine the number of lines on a grating. When this grating is used with a spectrometer, the second-order image is found at an angle of 30° from the central image. Determine the number of lines per centimeter on the grating.

40-14. The yellow line in the spectrum of sodium, sometimes called the D line, consists of two lines very close together when viewed with a spectroscope of moderate resolving power. The wavelengths of these lines are $D_1 = 5{,}896$ A and $D_2 = 5{,}890$ A. Determine the angular separation of these lines when viewed with a diffraction grating having 6,000 lines/cm and viewed (a) in the first order and (b) in the second order.

41

Polarized Light

41-1 Polarization. Transverse Waves

The phenomena of interference and diffraction show that light is propagated as a wave motion, but they do not show whether light is a *longitudinal* wave or a *transverse* wave. The fact that the velocity of light is the same as the velocity of radio waves and the radiation of visible light from accelerated electrons, as in a betatron, indicates that light is an *electromagnetic* wave. We recall from Section 20-9 that a wave can be shown to be transverse if a device can be found which will prevent passage of the wave in one orientation and will allow the wave to be transmitted when in a second orientation at right angles to the first, as in the case of a slit and a transverse wave on a string. Since a longitudinal wave will pass through a slit, however that slit is oriented, longitudinal waves may be distinguished from transverse waves by our inability to demonstrate the property of *polarization.* Such materials as *Polaroid* enable us to demonstrate that light waves are transverse waves. A beam of light in which all the vibrations are in one direction is said to be *linearly polarized,* or *plane polarized.*

On the basis of the electromagnetic theory of light, a linearly polarized monochromatic beam consists of a varying electric field accompanied by a similarly varying magnetic field traveling with the velocity of light, as shown in Figure 41-1. If we take the direction of motion of the beam as the x direction, the vectors representing the electric and magnetic fields at any point along the beam will be in a plane perpendicular to the x axis. The vectors $\mathbf{E}$ and $\mathbf{H}$ are perpendicular to each other and to the direction of propagation in air or vacuum. The direction of propagation of the electromagnetic wave is given by the direction of the vector $\mathbf{E} \times \mathbf{H}$. *We shall arbitrarily take the direction of the electric field intensity* $\mathbf{E}$ *as the direction of vibration of linearly polarized light.*

Unless special conditions prevail, the light emitted by a source is unpolarized. The light is emitted in short polarized bursts by independent atoms and molecules in random phase and random polarization, and the net effect is that of an unpolarized beam. We may represent the vibrations

in an unpolarized beam as taking place in all directions at right angles to the direction of propagation of the beam, as shown in Figure 41-2. In a radio or television signal the elecromagnetic wave is generally polarized,

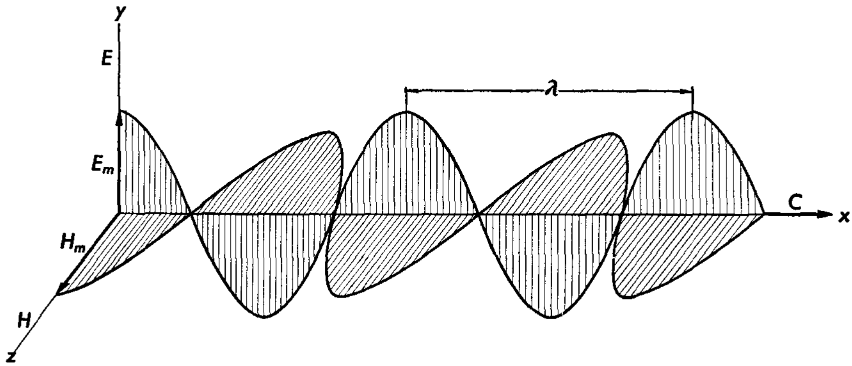
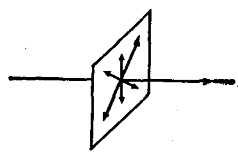


Fig. 41-1 The electric and magnetic field intensities of a linearly polarized beam of light moving in the x direction. The direction of vibration of the beam is taken as the direction of the electric field, in the above case, the y direction.

with the electric vector directed in the plane formed by the direction of propagation and the wire or rod from which the antenna is made. The polarization of a television signal may be demonstrated by rotating an antenna about an axis parallel to the direction of propagation. The signal

Fig. 41-2 The vibrations in a transverse wave are in a plane at right angles to the direction of propagation.



received is a maximum when the receiving antenna is horizontal, that is, parallel to the transmitting antenna.

The human eye is insensitive to the state of polarization of a light beam, so that experiments on the polarization of light must be conducted with the aid of some polarizing substance or device. It has been suggested that the eyes of insects may be sensitive to the polarization of light, and that this may be the mechanism by which bees find their way, as we shall see in the next section.

41-2 Polarization by Scattering

Light can be polarized by scattering from small particles or by molecules of a substance. For example, the blue light of the sky, produced by the scattering of sunlight by air molecules, is partially polarized. When we

look at the sky through a sheet of Polaroid, the intensity of the transmitted light is a minimum when the axis of the Polaroid is at right angles to the direction of vibration of the light.

If unpolarized light is directed vertically down a tube of water containing some fine particles in suspension, the light that is scattered in a horizontal direction will be found to be polarized, as shown in Figure 41-3.

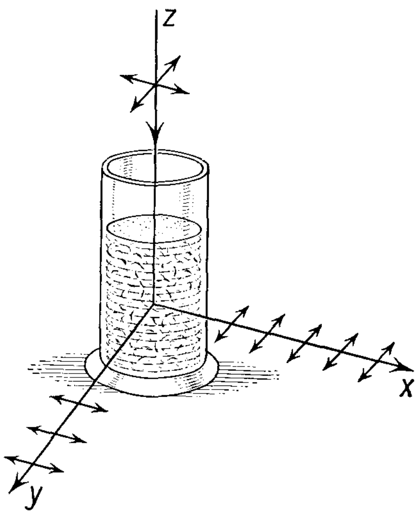


Fig. 41-3 Light traveling downward, z-direction, is scattered by small particles in the liquid. Light scattered in the x-direction has its vibrations in the y-direction. Light scattered in any horizontal direction is linearly polarized.

The direction of vibration of the incident light is in a horizontal plane. The electrons of the substance are caused to vibrate in the horizontal plane by the varying electric field of the incident light. These electrons reradiate the energy absorbed from the incident light, so that the direction of polarization of the scattered light must be horizontal. Since light is a transverse wave motion, the light scattered in the x direction, for example, can have no vibrations in the x direction; hence the only vibrations present when the scattered light is examined along the x axis will be those in the y direction. Light which is scattered in the forward direction or in the backward direction with respect to the incident beam need not be polarized. Only that light scattered at right angles to the original beam is completely polarized, as shown in the figure. Note that the direction of the incident beam can be determined as lying in a plane formed by the direction of the scattered radiation and the normal to the direction of vibration. Thus, by means of a sheet of Polaroid, the direction of the sun can be determined even when the sky is overcast. First one must find the direction in which the Polaroid has the greatest effect on the skylight. The normal to the direction of vibration of the scattered light is found by rotating the Polaroid to minimum intensity. The axis of the Polaroid is then directed along the direction

of the sun. In photography, polarizing filters are sometimes used to yield unusual effects, for by rotating the filter the intensity of the skylight reaching the film can be varied, and one may photograph white clouds against a dark sky to provide good cloud contrast.

The polarization of x-rays by scattering was used by Barkla in 1911 to show that x-rays are transverse waves. As we have seen, a beam of electromagnetic waves scattered at 90° has its direction of vibration normal

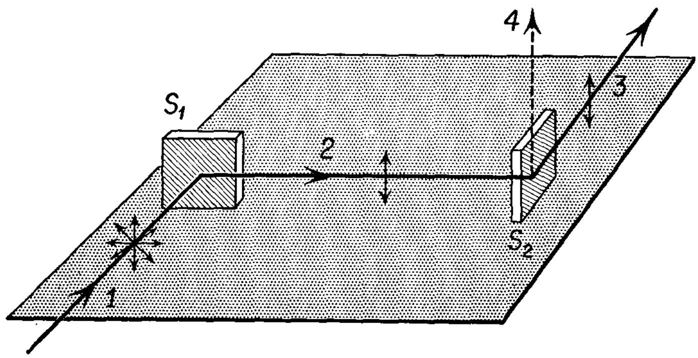


Fig. 41-4 Scattering of x-rays.

to the plane formed by the incident and the scattered beams. If this scattered beam is incident upon another block of scattering material, as in Figure 41-4, the waves scattered by the second block must have their direction of vibration in the same direction as the first scattered beam. Hence if a beam of x-rays is scattered first from S_1 and then from S_2, the x-rays scattered from the S_2 must have maximum intensity in direction 3 and zero intensity in direction 4. Barkla's experiment showed that this was indeed the case and thus demonstrated the transverse character of x-rays.

41-3 Polarization by Reflection

Light can be polarized by reflection from a plate of glass by the proper choice of the angle of incidence. For example, if the index of refraction of the glass is 1.54, the angle of incidence should be 57°, as shown in Figure 41-5(a). In general, the angle between the reflected ray and the refracted ray is 90° when the reflected light is completely polarized. At this angle of incidence, the reflected light has its direction of vibration parallel to the glass surface. If the reflected ray strikes another glass plate parallel to the first, the light will be reflected from the second plate in the usual manner. Now if the second glass plate is rotated through 90° about the ray incident on it as an axis, it will be found that no light is reflected from it; instead,

all the light is transmitted through it, as shown in Figure 41-5(b). When the second plate is turned through another 90° about the same axis, the light will again be reflected, as in Figure 41-5(c).

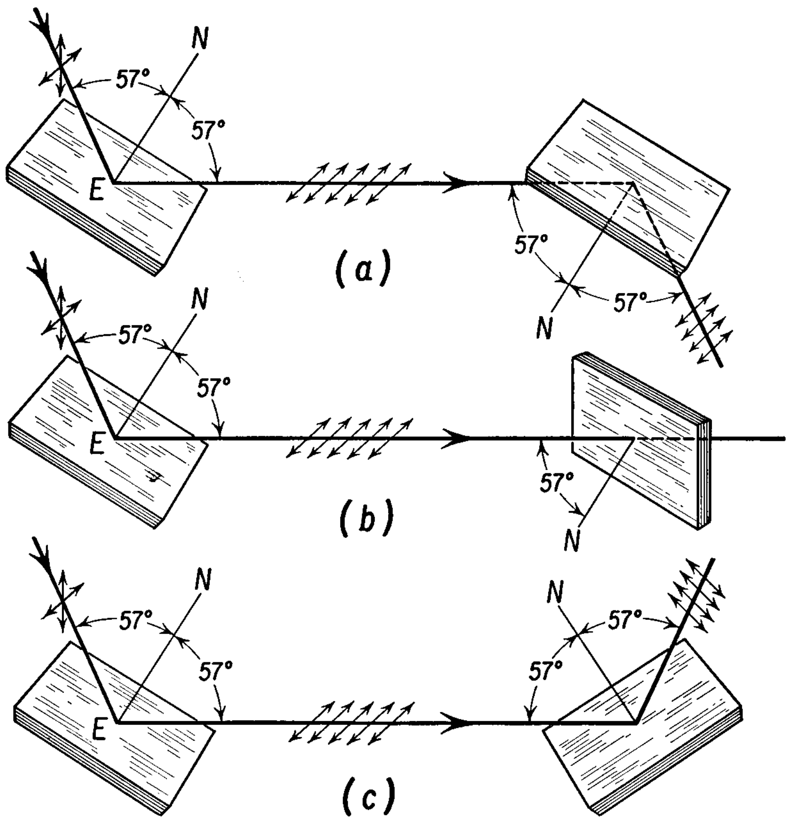


Fig. 41-5 Polarization of light by reflection from a glass plate.

This peculiar behavior of the beam reflected from the first plate can be explained by assuming that only those vibrations which are parallel to the surface of the first plate are reflected from it, when the angle of incidence is such that the angle between the reflected and refracted rays is 90°. Let us call this particular angle of incidence i_p, the *polarizing angle*. The unpolarized incident beam has light of all possible directions of vibration in a plane at right angles to the direction of motion. We may imagine these vibrations to be resolved into two components—one vibrating parallel to the surface of the glass plate, and the other at right angles to this direction. When the angle of incidence is at the polarizing angle, the reflected beam contains only vibrations which are parallel to the glass surface.

When this reflected beam strikes the second glass plate at an angle of incidence equal to the polarizing angle, it will be reflected at maximum intensity when the surface of the plate is parallel to the direction of the vibrations. As the second plate is turned about the ray incident upon it as an axis, the intensity of the light reflected from it decreases and becomes zero after a rotation of 90°, then starts increasing again as the glass plate is turned beyond this position, and becomes of maximum intensity after a rotation of 180°. Thus, as the second glass plate rotates about the ray incident upon it at the polarizing angle, there will be two positions of maximum intensity of the reflected beam and two positions of zero intensity.

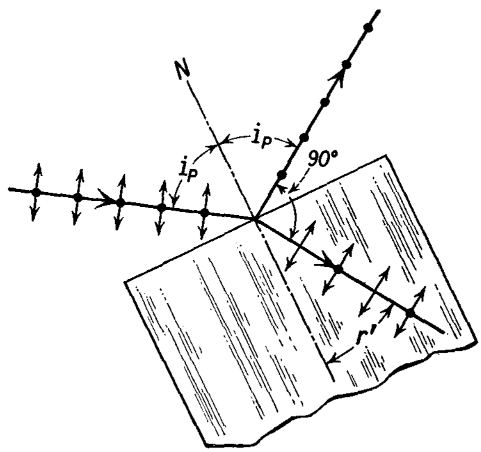


Fig. 41-6 The incident ray is unpolarized. The vibrations are shown resolved into two components; the component parallel to the glass surface is represented by dots. The reflected ray is completely polarized; the refracted ray is only partially polarized.

This combination of two glass surfaces arranged so that the light strikes the first surface at the polarizing angle, and the second glass surface mounted so that it can rotate about the reflected beam as an axis, is one form of *polariscope*. The first glass plate, which reflects polarized light, is called the *polarizer*, and the second glass plate, which is used to analyze the light, is called the *analyzer*.

The correct polarizing angle i_p for a particular glass can be found by applying Snell's law to the beam which strikes it. Assuming that the glass is in air, we have

$$\frac{\sin i_p}{\sin r'} = n,$$

where r' is the angle of refraction, as in Figure 41-6. Since the angle of reflection is equal to the angle of incidence, and the reflected ray is perpendicular to the refracted ray, $i_p + r' = 90°$; that is, the angles i_p and r' are complementary. In this case

$$\sin r' = \cos i_p;$$

hence Snell's law becomes

$$\frac{\sin i_p}{\cos i_p} = n = \tan i_p. \tag{41-1}$$

Thus the tangent of the polarizing angle is equal to the index of refraction of the glass. If $n = 1.54$, the polarizing angle is 57°.

Only about 8 per cent of the incident light is reflected from a glass surface at the polarizing angle. By using a bundle of thin glass plates, say seven or eight, the intensity of the reflected beam may be increased to about 40 per cent of the incident beam. This light is all linearly polarized, with the direction of vibration parallel to the glass surfaces. A small percentage of the light will be absorbed by the glass plates, and the remainder will be transmitted. If the number of reflecting surfaces is large, the transmitted beam will consist principally of light whose direction of vibration is perpendicular to the reflecting surface.

41-4 Polarization by Crystal Absorption

Some crystals, such as tourmaline, possess the property of absorbing those vibrations which are perpendicular to the axis of the crystal and of transmitting the vibrations which are parallel to this axis. If a beam of light which is unpolarized is sent through a thin plate of tourmaline, the trans-

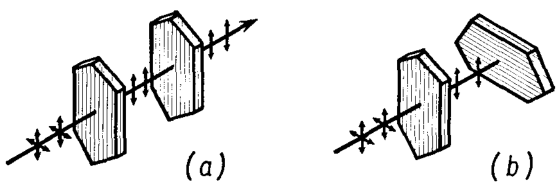


Fig. 41-7 Polarization by absorption in a crystal. (a) Linearly polarized light from the polarizing crystal is transmitted by the analyzer. (b) The linearly polarized light is absorbed by the analyzer when it is turned through 90°.

mitted beam will be linearly polarized parallel to the crystal axis. If this linearly polarized beam is now allowed to fall on a second tourmaline plate set so that its axis is at right angles to the axis of the first crystal, no light will be transmitted by the second crystal, as shown in Figure 41-7. But, if the second tourmaline crystal is rotated through 90° about the incident beam as an axis, thus making the axes of the two crystals parallel, the linearly polarized beam will be transmitted through the second crystal. At intermediate positions the intensity of the transmitted beam will be smaller than that transmitted when the axes are parallel. This phenomenon is known as *dichroism*.

One variety of Polaroid contains crystals of a synthetic material, known as herapathite, oriented in a nitrocellulose film. More recently, specially treated polyvinyl films have been shown to be dichroic and are marketed as varieties of Polaroid.

We have already seen that the energy per unit volume in the electric field in vacuum is $\epsilon_0 E^2/2$, while the energy density of the magnetic field in vacuum is $\mu_0 H^2/2$. In an electromagnetic wave in vacuum, the electric and magnetic fields vary with time, and we may compute the average energy in the field by taking the rms values of these quantities, rather than the instantaneous values of the electric and magnetic field intensities, as in the case of the rms values of alternating current. The flow of energy in the wave depends upon the square of the amplitude of the vibration rather than upon the amplitude itself. We call the energy reaching a unit area of surface per unit time the *intensity* of the wave. Thus the intensity of a light wave depends upon the square of the amplitude of vibration.

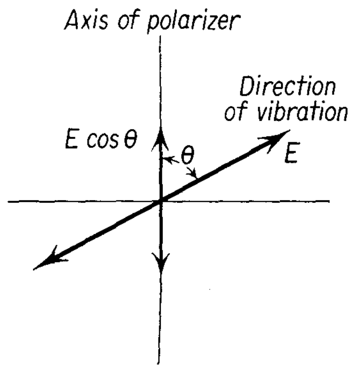


Fig. 41-8

When plane polarized light passes through a polarizer whose axis makes an angle θ with the direction of vibration, as shown in Figure 41-8, we may imagine the vibrations of the light wave to have components parallel and perpendicular to the axis of the polarizer. If the amplitude of vibration of the polarized light is E, only the component parallel to the axis of the polarizer passes through it. This component has amplitude $E \cos \theta$. The intensity I of a beam of light is related to its amplitude of vibration through the equation

$$I = \alpha E^2,$$

where α is some constant of proportionality. If the incident beam is of intensity I_0, the transmitted beam will be of intensity I given by

$$I = I_0 \cos^2 \theta. \tag{41-2}$$

This relationship is known as the *law of Malus*. If unpolarized light is incident upon a perfect polarizer, only half the incident light intensity is transmitted as a plane polarized beam, for the light intensity may be resolved into two equal components one of which is vibrating parallel to the axis of the polarizer and the other is vibrating perpendicular to the axis of the polarizer. When linearly polarized light is transmitted by a perfect analyzer, the intensity of the transmitted light varies from zero to

the incident intensity, depending upon the angle θ between the direction of vibration of the light and the axis of the analyzer, in accordance with Equation (41-2).

41-5 Polarization by Double Refraction

When a narrow beam of unpolarized light is sent through certain transparent crystals, such as calcite and quartz, the refracted beam is split into two parts which travel through the crystal and emerge as two separate beams, as shown in Figure 41-9. When an object is viewed through such a crystal, two separate images are seen. If the emergent beams are analyzed

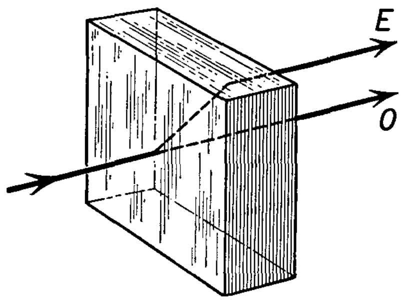


Fig. 41-9 Double refraction by a crystal.

with a Polaroid film, it will be found that each beam is polarized, but that the directions of their vibrations are at right angles to each other. One of these beams obeys the ordinary laws of refraction and is called the *ordinary ray,* designated as O in Figure 41-9. The other ray E is known as the *extraordinary ray,* for it does not always lie in the plane of incidence; its speed, and hence its index of refraction, depends upon its direction of propagation through the crystal.

In general, if an unpolarized beam of light is sent through a crystal at an arbitrary angle, two linearly polarized beams emerge. However, it is possible to find one direction in the crystal such that both the ordinary ray and the extraordinary ray will travel with the same speed. This particular direction in the crystal is known as the *optic axis.* If the arrangement of atoms in the calcite crystal is examined in a plane perpendicular to the optic axis, the atoms are observed to be symmetrically distributed. In no other direction is this true. The electrical properties of the crystal, such as its permittivity, and the optical properties are found to vary in different directions in the crystal. In all crystals except those displaying cubic symmetry, double refraction is observed. In some crystals more complicated than calcite, there is not one direction in which the ordinary ray and the extraordinary ray travel with the same speed but, rather, there are two such directions. Crystals with one optic axis are called *uniaxial*

crystals, while the latter types are called *biaxial crystals*. We will limit our discussion to the case of uniaxial crystals, which may be understood by a special type of wave surface suggested by Huygens.

In isotropic media the device of spherical wave fronts was used to describe the propagation of light. Let us assume that the ordinary ray is propagated by such spherical wave fronts, but that the extraordinary ray is propagated by ellipsoidal wave fronts. A point source of light within such a crystal is thought to emit two wave fronts, corresponding to different directions of vibration, as shown in Figure 41-10. In the direction of the optic axis, the two wave fronts are tangent.

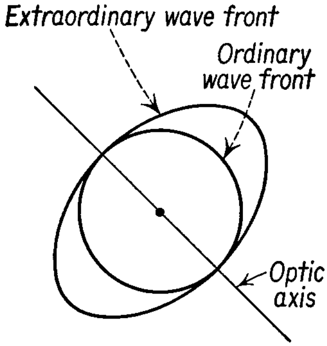


Fig. 41-10

Following the Huygens construction, let us take the case of a plane wave AB incident upon the plane surface S of a uniaxial crystal, in which the optic axis lies in the plane of incidence, as shown in Figure 41-11. By the time the portion of the incident wave front at B reaches the surface at C, the O and E wavelets at A have reached the indicated size. Drawing the tangent to the O wavelet from C, we find the refracted ordinary wave front to be CD, and the refracted ray to be AD. In a similar way the E wave front is CF, and the refracted E ray is AF. The vibrations of the O ray are normal to the plane of the figure, while the vibrations of the E ray lie in the plane of the figure. Both vibrations are tangent to their respective wavelets. In general, it is found that *the* O *vibrations are normal to a plane containing the refracted* O *ray and the optic axis, called the principal plane of the* O *ray; the* E *vibrations lie in a plane containing the* E *ray and the optic axis, called the principal plane of the* E *ray.* Note that the vibrations of the E ray are not perpendicular to the direction of propagation of the E ray. The incident light is doubly refracted.

In the above case the refracted E ray lies in the plane of incidence. When the optic axis does not lie in the plane of incidence, the tangent plane to the E wave front may be skewed so that the refracted E ray also does not lie in the plane of incidence. The E ray is truly extraordinary. The behavior of a crystal under the influence of the varying electric field of the light ray is related to the behavior of a long bar of rectangular cross section subjected to a periodic driving force. When the force is normal to one face of the rod, waves are generated in the rod which are propagated with one velocity, appropriate to the stiffness of the rod in the direction of the applied force. Thus there are two velocities of wave propagation down the rod, for the stiffness of the rod is different in the two normal

directions. If the applied force is not normal to one face of the rod, the rod responds as though it were subjected to two independent normal forces—the components of the applied force normal to the faces of the rod. Two separate wave motions are generated which travel down the rod with

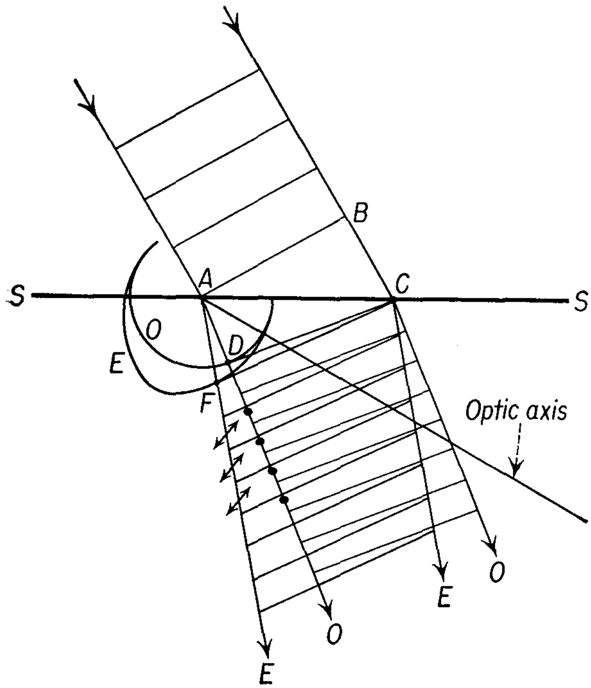
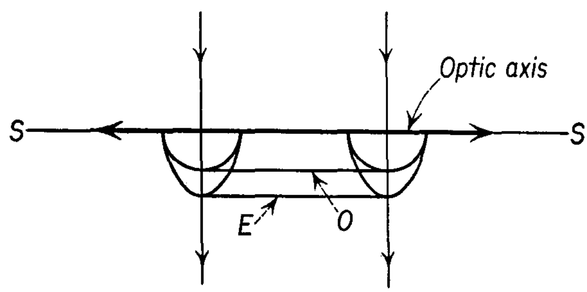


Fig. 41-11

Fig. 41-12



different speeds; each speed is characteristic of a particular direction of vibration. The displacement of the rod at any point is the vector sum of the two displacements due to the two wave motions, at right angles to each other.

The Huygens ellipsoid which describes the propagation of the extraor-

dinary wave requires that we have knowledge of the velocity of propagation of the E ray in every direction. Once this is known, the behavior of the E ray and the O ray, under arbitrary conditions, can be determined. Thus, if a plane wave of unpolarized light is incident normally upon a section cut from a uniaxial crystal so that the optic axis is parallel to the crystal face, the O ray and the E ray are propagated along the normal to the crystal face with different speeds but in the same direction, as shown in Figure 41-12.

41-6 The Nicol Prism

An ingenious method of producing linearly polarized light was devised by W. Nicol in 1828 and is known as a *Nicol prism*. A natural crystal of calcite is shown in Figure 41-13. In such a crystal the direction of the optic

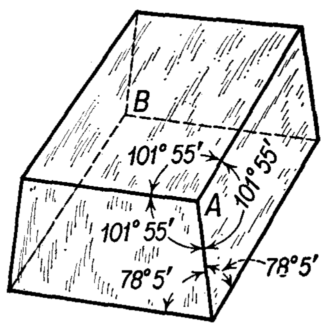


Fig. 41-13 Sketch of calcite crystal.

axis is parallel to a line making equal angles with the three intersecting angles at A or B. Using a fairly long, clear crystal, the end faces are cut so as to obtain a piece whose cross section is a parallelogram with an

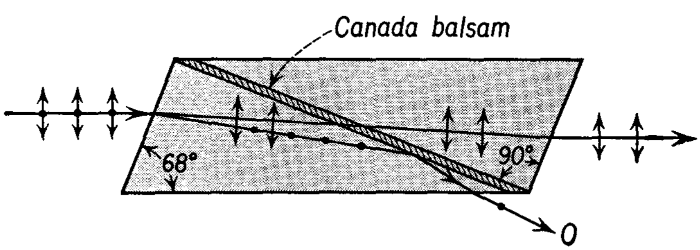


Fig. 41-14 Nicol prism.

included angle of 68°, with the sides of the section parallel to the principal axis of the crystal. This piece is cut in two along the diagonal, as shown in Figure 41-14, in cross section. The two parts of the piece are then

cemented together with a cement widely used in optical work, known as Canada balsam. As shown in Figure 41-14, a plane wave incident upon the crystal is doubly refracted. The index of refraction of the ordinary ray in calcite is 1.66, and in Canada balsam it is 1.55. The angle of incidence of the ordinary ray when it strikes the Canada balsam is greater than the critical angle, so that this ray is totally reflected. The index of refraction of the extraordinary ray depends upon its direction. For the possible directions in the Nicol prism, the index of refraction of the E ray is about 1.49; hence it will be transmitted through the Canada balsam and emerge from the prism.

Nicol prisms are used in optical devices as analyzers, polarizers, or both. The Nicol prism suffers the disadvantage, compared to sheets of Polaroid, that it is limited in aperture, but it has the advantage that it produces a completely polarized beam of light with all the vibrations in one direction.

41-7 Elliptically Polarized Light

In our study of simple harmonic motion in Chapter 12, we have seen that the projection of the motion of a particle in uniform circular motion onto two mutually perpendicular axes yields two simple harmonic motions of equal amplitude and frequency which are 90° out of phase with each other. One of these projected motions was taken along the x axis and the other along the y axis in Chapter 12. Thus we may think of these two simple harmonic motions as the x and y components of a uniform circular motion.

The general case of motion of a particle subjected to two perpendicular harmonic motions is quite complex. The path followed by the particle is known as a *Lissajous* figure. In studying the behavior of polarized light, we shall be interested in the special case where the two motions are of the same frequency f, and therefore of the same angular frequency ω, although there may be an arbitrary phase difference ϕ between the two simple harmonic motions. Let us examine the path of a particle subjected to two simple harmonic motions, of the same frequency but of different amplitudes, when these are directed along two perpendicular directions, say the x and y axes. We have

$$x = A \sin \omega t,$$

and

$$y = B \sin (\omega t + \phi).$$

The simplest case is one in which $\phi = 0$, that is, the two motions are *in phase*. Here we find that

$$x = \frac{A}{B} y;$$

that is, the path of the vibrating particle is a straight line.

The second case of interest to us is that in which ϕ is equal to 90°, so that $y = B \cos \omega t$; for this case

$$\frac{x^2}{A^2} + \frac{y^2}{B^2} = 1.$$

This is the equation of an ellipse with semiaxes A and B; hence a particle subjected to two simple harmonic motions of different amplitudes which are 90° out of phase experiences an elliptical motion. If the two motions are of the same amplitude, the path is circular.

From Chapter 35 we know that light can be emitted by an accelerated charged particle. When the charged particle oscillates with a frequency f, the emitted light is of the same frequency and is linearly polarized, with its direction of polarization perpendicular to the direction of propagation and lying in the plane formed by the acceleration vector and the direction of propagation. If the charged particle moves in an ellipse, the light emitted is elliptically polarized; that is, it may be resolved into two linearly polarized components which are of different amplitudes but of the same frequency, and which are 90° out of phase with each other. The two components are along the axes of the ellipse.

We can generate circularly polarized light and elliptically polarized light from plane polarized light with the aid of a thin section of a doubly refracting crystal known as a *quarter-wave plate*. Such a plate is also useful in identifying elliptically polarized light. We shall find it convenient to discuss these plates by means of the concept of the *optical path length*, which is defined as the distance the light would travel in vacuum in the time it takes to traverse a given path in a medium. The time to traverse a distance s in a medium is s/v, where v is the velocity of light in the medium. If n is the index of refraction of the medium, the optical path length s' is given by the equation

$$s' = \frac{s}{v} c = ns. \tag{41-3}$$

If a plane wave is incident normally upon a slab of medium of thickness s, the phase difference between light just incident upon the front face of the slab and the light just leaving the rear face of the slab is given by $\Delta\phi$, where

$$\Delta\phi = \frac{2\pi}{\lambda} ns, \tag{41-4}$$

and λ is the wavelength of the light in vacuum. Note that there is no phase change on transmission, and that with this consideration Equation (41-4) is in agreement with Equation (40-5).

When light is incident normally upon a slab of a uniaxial crystal with parallel faces, cut so that the optic axis lies in the plane of the crystal face,

the ordinary and extraordinary rays are refracted normally, as shown in Figure 41-12, so that there is no angular separation between the two rays. Since the refractive indices n_O and n_E are different in this direction, we find that the phase changes of the two rays are different. Plates which are cut of such thickness that the difference in phase between the emergent O and E rays is π are called *half-wave plates*, while if the difference in phase between the two emergent rays is $\pi/2$, the plate is called a *quarter-wave plate.*

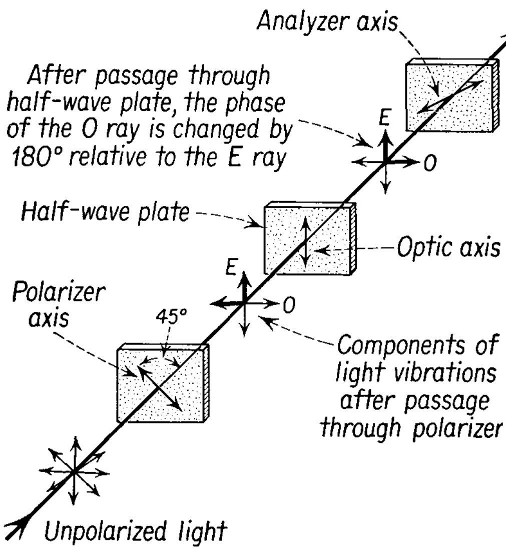


Fig. 41-15 A half-wave plate rotates the plane of polarization of linearly polarized light.

We have seen that linearly polarized light may be resolved into two components whose directions of vibration are mutually perpendicular. Referring to Figure 41-15, suppose that a beam of monochromatic light first goes through a polarizer and then strikes a half-wave plate with the direction of its optic axis at an angle of 45° with the direction of vibration of the incident beam. We can consider the light on its passage through the half-wave plate as resolved into two components, the E ray with its direction of vibration parallel to the optic axis, and the O ray with its direction of vibration perpendicular to the optic axis. Although these two vibrations are in phase when the light enters the crystal, the effect of traversing the half-wave plate is to introduce a phase difference of π between them, as shown in the figure. The sum of these two vibrations is again a linear vibration but at right angles to the direction of the vibrations of the incident beam. This can be verified by rotating an analyzer

until extinction of the beam is produced. It must be remarked that a half-wave plate is such only for a particular wavelength.

If a quarter-wave plate is substituted for the half-wave plate of Figure 41-15, then the phase difference between the E and the O vibrations of the emergent beam will be $\pi/2$; that is, when the E beam has maximum amplitude, the O beam has zero amplitude, and when the E beam has zero amplitude, the O beam has maximum amplitude. The emergent beam is *circularly polarized.* If this beam is now examined with the analyzer, it will be transmitted at all angles with equal intensity, similar to the behavior of an unpolarized beam. To detect the fact that the emergent beam is circularly polarized, it is necessary to introduce another quarter-wave plate in the path; the effect will be to convert the circularly polarized beam into a linearly polarized beam.

If the thickness of the crystal is such as to produce a phase difference of $\pi/4$, for any particular color, it will convert a linearly polarized beam into an *elliptically polarized* beam. When examined with an analyzer, the intensity of one direction of vibration will be different from that in a direction at right angles to it.

41-8 Interference with Polarized Light

We have already considered the phenomenon of interference using ordinary or unpolarized light. From the preceding discussion and from the vector superposition principle, it must be clear that destructive interference can only take place with waves which are linearly polarized in the same di-

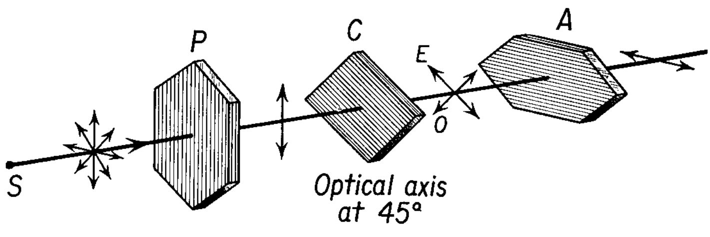


Fig. 41-16 A doubly refracting crystal C placed between the polarizer P and the analyzer A when P and A are in crossed positions.

rection. It was unnecessary to make this restriction in the case of ordinary interference phenomena, for the requirement of coherent radiation led us to consider the interference as composed of the sum of the effects produced by the light emitted from individual atoms. In the present discussion we wish to consider special interference effects which are produced with linearly polarized light.

Consider, for example, a polariscope arranged in the *crossed* position, as shown in Figure 41-7(b), so that no light passes through the analyzer. If a piece of doubly refracting material, such as a quartz crystal C, with its optic axis in the plane of the crystal face and oriented at 45° with the axis of the polarizer P, is placed between the polarizer and the analyzer A, colored light passes through the analyzer (see Figure 41-16). When white light is incident on the polarizer, linearly polarized white light is

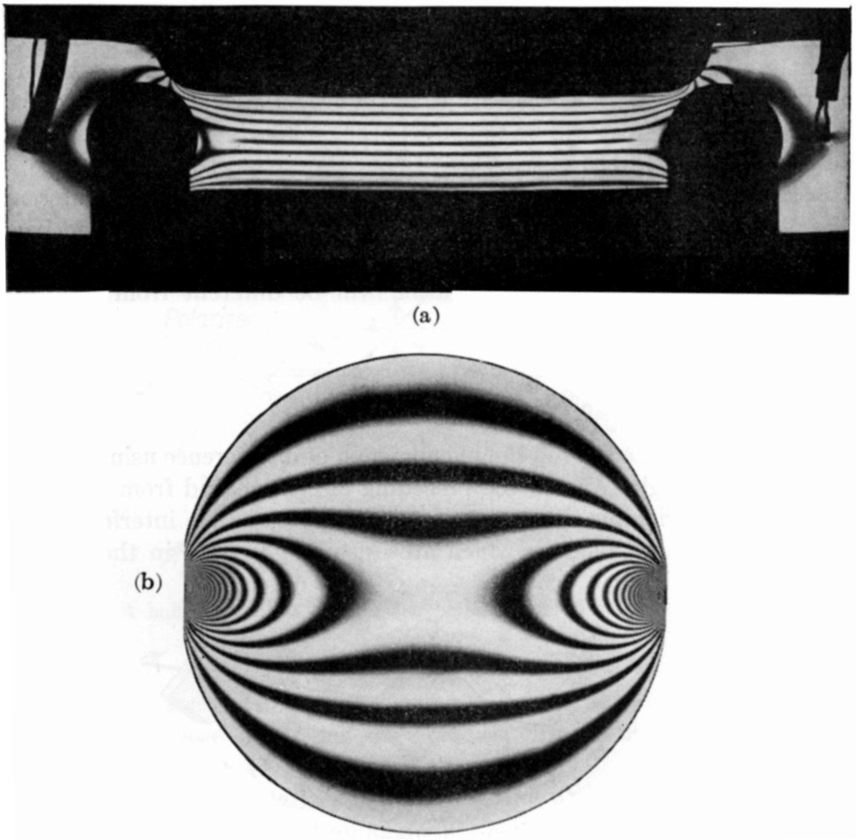


Fig. 41-17 Interference pattern produced by passing polarized light through a piece of transparent plastic in a crossed polariscope. (a) Strain pattern when a bar is subjected to a bending moment. (b) Strain pattern in a cylindrical disk subject to a diametral compression. (Photographs by George Gerard, New York University.)

incident upon the crystal. The beam is divided into two components, the extraordinary ray vibrating parallel to the optic axis, and the ordinary ray vibrating perpendicular to it. Since they travel with different velocities, there will be a difference in phase when they emerge. But when the two vibrations pass through the analyzer, only the horizontal components

of these vibrations will be transmitted, and interference will occur between these two transmitted rays. If the path difference due to passage through the crystal is a half wavelength, or an odd number of half wavelengths, these two rays will annul each other. When this is the case for the yellow light, the light transmitted through the analyzer will be "minus yellow," or purple. Hence specimens of clear, colorless, doubly refracting material will appear colored when viewed through a polariscope. These effects may be observed with pieces of mica or Scotch tape, and colored decorative patterns may be obtained from an overlaid mosaic of transparent substances.

Many transparent materials, such as glass and clear Bakelite, which are not noticeably doubly refracting under ordinary conditions, become so when subjected to stresses. When such a substance is placed between the polarizer and analyzer of a polariscope, a colored pattern may be observed. The pattern of colors can be related to the strains in the substance produced by the stresses. In modern engineering practice, models of structures are built of transparent Bakelite, and the stresses induced in elements of the structure due to applied loads may be determined with the aid of polarized light. This type of study of the stress in a structure is called *photoelasticity*. Strain patterns in loaded specimens illuminated with monochromatic light are shown in Figure 41-17. When glassware is cooled too rapidly, internal stresses often develop owing to nonuniform contraction. These internal stresses may be studied by placing the object in a polariscope.

41-9 Further Effects and Applications

(a) *Optical Activity.* Many solids and liquids *rotate* the direction of polarization of linearly polarized light. If a tube of sugar solution is placed between a polarizer and an analyzer, it is found that the sugar solution rotates the direction of polarization by an amount which depends upon the length of the tube, the sugar concentration, and the type of sugar used. In some cases the rotation is clockwise, or right-handed, and is called *dextrorotatory,* while in other cases the rotation is left-handed, or *levorotatory.* Such a polariscope may be used to measure the concentration of the solution of a known sugar; it is called a *saccharimeter.* This type of optical activity is associated with a corkscrewlike structure of the molecules of the substance.

(b) *Kerr Effect.* Many substances acquire the property of *birefringence* and become doubly refracting when placed in a sufficiently strong electric field. The effect is due to anisotropic alignment of the molecules of the substance when placed in the electric field. This effect is especially marked in carbon disulphide and nitrobenzene. A cell of nitrobenzene of proper thickness can be made into a temporary half-wave plate, under

the influence of an electric field, and can act as a very fast optical shutter when placed between crossed sheets of Polaroid. The Kerr effect is used as the optical shutter in the *geodimeter*, mentioned in Section 36-8, and in other devices to measure the velocity of light.

(c) *Cotton-Mouton Effect.* Some substances placed in a strong transverse magnetic field become doubly refractive, in this magnetic analogue of the Kerr effect. The Cotton-Mouton effect is associated with the alignment of the molecules in a magnetic field. Both the Kerr effect and the Cotton-Mouton effect vary with temperature, as we might expect, for the tendency of the molecules to become disaligned depends on their random motion, which increases with temperature.

(d) *Glare Reduction.* Light incident upon a diffusely reflecting surface at large angles is sometimes reflected at an angle of reflection approximately equal to the angle of incidence. This results in *glare.* When the reflected light is analyzed with a Polaroid, it is found that the component with its vibrations parallel to the surface is much more intense than the component whose vibrations are perpendicular to the surface. Glare may be reduced considerably by using polarizing spectacles oriented to absorb the horizontal vibrations.

Problems

41-1. The index of refraction of one type of flint glass is 1.65. Determine the polarizing angle when the glass is (a) in air and (b) immersed in water of index 1.33.

41-2. The indices of refraction of yellow light of 6,000 A in a doubly refracting medium are 1.710 for the ordinary ray and 1.740 for the extraordinary ray. (a) Determine the velocity of each of these waves in this medium transverse to the optic axis. (b) Determine the minimum thickness of this material necessary to produce a path difference of half a wavelength for these two rays. (c) If white light is incident on a polariscope in the crossed position, what will be observed when this doubly refracting material is placed between the polarizer and analyzer? What will be the intensity of the light passing through the analyzer compared to the light incident upon the doubly refracting plate?

41-3. Derive a formula relating the critical angle θ_c of a transparent substance to its polarizing angle θ_p.

41-4. A beam of unpolarized light is incident upon a polarizer. (a) If the incident beam has a luminous flux of 5 lumens, what will be the luminous flux passing through the polarizer? (b) The light is then incident upon an analyzer whose axis is at an angle of 53° with the axis of the polarizer. What will be the luminous flux which passes through the analyzer?

41-5. A beam of unpolarized light traveling in the $+x$ direction is scattered in all directions by a cloud of fine particles. (a) Is the light scattered at an angle of 37° with the $+x$ direction completely polarized? (b) If this scattered light is

examined by a Polaroid, what are the relative intensities of the maximum and minimum light transmitted by the Polaroid?

41-6. The index of refraction for the ordinary ray n_O of quartz is 1.544, while the index for the extraordinary ray n_E is 1.553 for sodium light. (a) Is the ordinary Huygens wavelet inside or outside the extraordinary wavelet? (b) A ray of polarized light is incident normally upon a slab of quartz in which the optic axis is in the plane of incidence and makes an angle of 30° with the normal. What is the angle of refraction of the ordinary ray? (c) What is the angle of refraction of the extraordinary ray? Use a Huygens construction.

41-7. A slab of quartz has its optic axis in the plane of incidence. (a) What is the thickness of a quarter-wave plate for sodium light? (b) What is the thickness of a half-wave plate? (c) Will the phase of the ordinary ray be advanced or retarded with respect to the extraordinary ray? Use the data of Problem 41-6.

41-8. Circularly polarized light is incident upon a polariscope. Compared to the incident beam, (a) what is the intensity of the light passing through the polarizer? (b) When the analyzer is in the crossed position, what is the intensity of the light passing through the analyzer? (c) When the axis of the analyzer makes an angle of 53° with that of the polarizer, what is the intensity of the beam passing through the analyzer?

41-9. Elliptically polarized light in which the amplitude of the x vibration is three times the amplitude of the y vibration is incident upon a polarizer. Compared to the incident beam, (a) what is the maximum intensity transmitted through the polarizer? (b) What is the minimum intensity transmitted through the polarizer? (c) What is the intensity transmitted through the polarizer when it is rotated to a position midway between the maximum and minimum position?

41-10. Elliptically polarized light, in which the amplitude of the x vibration is twice the amplitude of the y vibration, is incident upon a quarter-wave plate, and the light passing through the plate is then examined by an analyzer. Compared to the incident beam, (a) what is the maximum intensity of the light transmitted by the analyzer and (b) what is the minimum intensity of the light transmitted by the analyzer?

41-11. Three perfect polarizers are placed in the path of a beam of unpolarized light so that the axis of the second polarizer makes an angle of 30° with the first, while the axis of the third polarizer makes an angle of 90° with the first and an angle of 60° with the second. Find the percentage of the incident light transmitted by the combination.

41-12. A television antenna is found to receive a maximum signal when its axis lies along the direction of propagation and it is oriented horizontally. What fraction of the maximum signal is received when the plane of the antenna is rotated about the direction of propagation so that it lies at an angle of 37° with the horizontal?

Part Six

ATOMICS AND NUCLEONICS

42

Optical Spectra and Atomic Structure

42-1 Types of Optical Spectra

When light passes through a prism spectroscope or a diffraction-grating spectroscope, an optical spectrum is obtained in which the intensity of the radiation may be analyzed as a function of wavelength. The spectrum may be observed visually in the limited wavelength region to which the eye is sensitive; it may be focused on a photographic plate or upon a thermocouple or thermopile. Our knowledge of the structure of atoms and molecules is largely dependent upon the analyses of optical spectra, for these spectra are characteristic of the emitting atoms or molecules. Even before the spectra of atoms were properly understood in terms of the atomic structure, it was possible to determine the chemical composition of an unknown substance by study of its spectrum.

The spectra obtained from radiating bodies are called *emission spectra* and are classified as *continuous spectra, band spectra,* and *line spectra,* according to their appearance in a spectroscope. Continuous spectra are emitted by solids, liquids, and dense opaque gases at high temperatures. The spectrum of the sun, or of a black body, is a continuous spectrum and, as we have seen in Section 18-6, the shape of the black-body spectrum is characteristic of its temperature. Gases at low pressures emit *band* or *line* spectra. Line spectra have their origin in the energy changes which take place in the atoms of a gas, while band spectra are associated with similar changes in the molecules.

To early investigators the line spectra of atoms and the continuous spectra of black bodies posed seemingly insurmountable problems, for all explanations of the wavelengths of the spectral lines, or of the shape of the continuous spectra, were inadequate. For example, by analogy with

the properties of organ pipes and vibrating bodies, a black body was assumed to be capable of sustaining electromagnetic vibrations in much the same manner as an organ pipe is capable of generating a fundamental and its overtones, under appropriate excitation, or as a room is capable of sustaining standing waves. Each harmonic was assumed to be a mode of vibration and to have energy kT, as in the case of the specific heats of gases, but the calculated spectrum was in violent disagreement with the observed spectrum. To explain the shape of this spectrum Max Planck proposed, in 1900, that the energy of a given mode of vibration depended upon its frequency in accordance with the relationship

$$\boxed{\mathcal{E} = hf,} \tag{42-1}$$

where $\mathcal{E}$ is the energy, f is the frequency of vibration, and h is a constant now known as *Planck's constant* whose measured value is

$$h = 6.624 \times 10^{-27} \text{ erg sec.}$$

Planck's conception, in contradiction to the well-established principle of *equipartition of energy*, successfully predicted the shape of the black-body radiation spectrum. Planck's formula was subsequently interpreted by Einstein as implying that radiation was emitted or absorbed in discrete units called *quanta*, each of which had an energy given by Equation (42-1). This interpretation helped to explain many phenomena observed in the interaction of radiation and matter such as the photoelectric effect and the emission and absorption of radiation. According to the quantum theory of light, the energy in a beam of monochromatic light of frequency f is determined by the number of quanta hf in the beam rather than by the square of the amplitude of vibration.

From our discussion of black-body radiation we have seen that the efficiency of a body as a radiator of energy is equivalent to the efficiency of the body as an absorber of radiation. Thus we should expect that, when white light irradiated a gas, it would absorb precisely those wavelengths which it was capable of emitting. A spectrum produced in this manner is called an *absorption spectrum.* Contrary to classical expectations it was discovered that only a few of the lines in the emission spectrum of a gas were to be found in its absorption spectrum.

42-2 Spectrum of Hydrogen

Hydrogen, the simplest of the elements, has been investigated most extensively both theoretically and experimentally. The knowledge gained from this study has acted as a guide to the study of more complex elements.

As shown in Figure 42-1, the radiation emitted by hydrogen is a line spectrum, whose wavelengths were shown by Balmer, in 1855, to be related by a formula which may be stated as

$$\frac{1}{\lambda} = R\left(\frac{1}{2^2} - \frac{1}{n^2}\right) \tag{42-2}$$

for light in the visible spectrum. In this equation λ is the wavelength, R is a constant known as the *Rydberg constant* whose value for hydrogen is $R = 109{,}677.76\ \text{cm}^{-1}$, and n is an integer greater than 2. By substituting

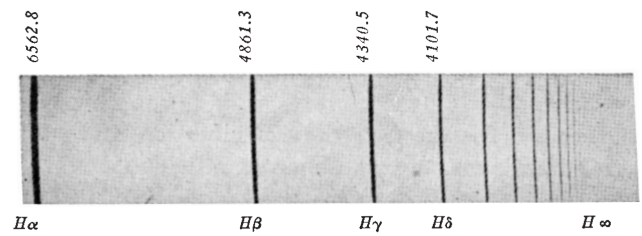


Fig. 42-1 Photograph of the emission spectrum of hydrogen showing the Balmer series lines in the visible and the near ultraviolet regions. The numbers represent the wavelengths of the corresponding lines in angstrom units. (Reprinted by permission from *Atomic Spectra and Atomic Structure*, by G. Hertzberg, Dover Publications.)

for n in Equation (42-2) the successive values 3, 4, 5, . . . , we obtain the wavelengths of the lines in the visible region and also in the near ultraviolet region. These lines form a series known as the *Balmer series* of hydrogen.

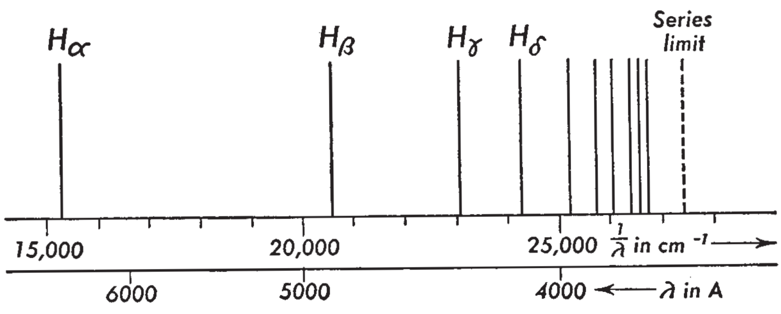


Fig. 42-2 Graph of the positions of the lines of the Balmer series. The upper scale is the reciprocal of the wavelength in centimeters, while the lower scale is the wavelength in angstrom units.

The relative positions of these lines are shown in Figure 42-2. As n approaches infinity, the wavelength difference between adjacent lines gets smaller; the lines crowd together and approach a limit known as the *series limit*.

In addition to the Balmer series in the visible spectrum, hydrogen emits radiation in the ultraviolet region, in a series known as the *Lyman series*, which is described by a formula like Equation (42-2), except that the first term in the parentheses is replaced by $1/1^2$, and the symbol n is replaced by integers from 2 on; there are lines in the infrared which can be described by a series formula in which the first term in the parentheses is replaced by $1/3^2$, and so on. The other series of hydrogen are shown in Figure 42-3.

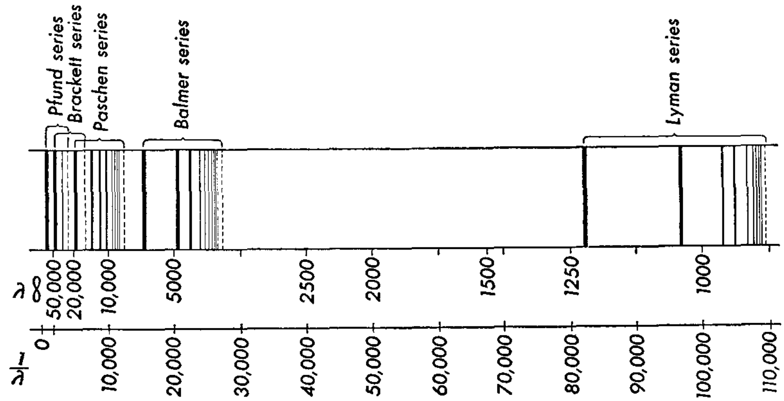


Fig. 42-3 Relative positions of the lines of the different spectral series of hydrogen. The upper scale is in angstrom units, while the lower scale is the reciprocal of the wavelength in centimeters.

None of this great regularity in the spectrum of hydrogen could be explained by classical ideas. In absorption at ordinary temperatures, hydrogen exhibits only the Lyman series in the ultraviolet, while several of the absorption lines in the solar spectrum, known as the *Fraunhofer lines*, lie in the Balmer series. The first quantitatively correct derivation of the Balmer formula was given by Bohr, in 1913, in his model of the hydrogen atom. This theory has played such an important role in the development of atomic physics that, even though it has been modified and extended by later developments, it is worth while presenting the original simplified theory.

42-3 Bohr's Theory of the Hydrogen Atom

In his theory of the hydrogen atom, Bohr chose as a model a positively charged central nucleus, consisting of a proton, about which a negatively charged electron revolved in a circular orbit. In accordance with classical theory, a revolving electron experiences a centripetal acceleration and should radiate electromagnetic waves, so that the electron should lose energy and slowly spiral into the nucleus. Bohr supposed that this would not happen in the atom, and that only certain orbits were permitted the

electron according to special quantum conditions. An allowable orbit was called a *stationary state,* for the electron could remain in a particular allowed orbit without the emission of radiation or without a change in the energy of the system.

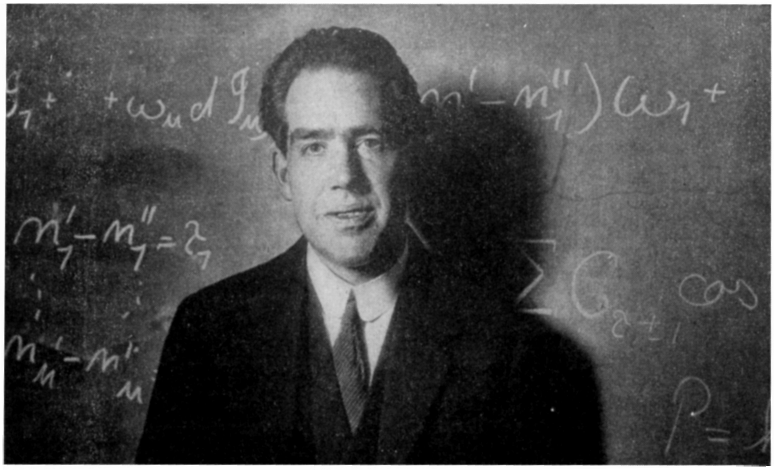


Fig. 42-4 Niels Bohr. He made outstanding contributions to the modern theory of the nuclear structure of the atom and formulated the theory of nuclear processes. (Courtesy of American Institute of Physics.)

If e represents the magnitude of the charge of the electron or the proton, the electrical force between the electron and proton is given by Coulomb's law, expressed in Gaussian (electrostatic) units by

$$F = \frac{e^2}{r^2}, \tag{42-3}$$

in which r is the radius of the circular orbit. The nucleus, having a mass of 1,836 times that of the electron, is assumed to remain stationary at the center of the circle, as though its mass were infinite. The force on the electron owing to the attraction of the two charges is directed toward the center and supplies the centripetal force necessary to keep the electron in its circular motion. Thus we have

$$\frac{mv^2}{r} = \frac{e^2}{r^2}, \tag{42-4}$$

in which m is the mass of the electron, and v is its speed.

The energy of this system is partly potential energy and partly kinetic energy. The potential energy of the negatively charged electron in the field of the nucleus is given by the expression $-e^2/r$, while the kinetic

energy is simply equal to $\frac{1}{2}mv^2$. From Equation (42-4) we see that

$$mv^2 = \frac{e^2}{r},$$

so that the total energy $\mathcal{E}$ of the electron in the atom is given by

$$\mathcal{E} = \tfrac{1}{2}mv^2 - \frac{e^2}{r} = -\frac{1}{2}\frac{e^2}{r}. \tag{42-5}$$

The minus sign for the total energy indicates that the electron is bound to the nucleus, and that positive work is required to remove the electron to infinity where it has zero energy when at rest.

Nothing thus far in the theory indicates that the radiation should be emitted in the form of sharp spectral lines. Bohr introduced two postulates to account for the emission of sharp spectral lines. The first of these postulates may be stated as follows: *Only those orbits are permissible for which the angular momentum of the electron is an integral multiple of* h/2π. Thus we have

$$\boxed{mvr = n\frac{h}{2\pi},} \tag{42-6}$$

where n is an integer called a *quantum number*. The orbits which satisfy this condition are called *stationary* orbits. The second of Bohr's postulates states that *whenever the energy of the atom is decreased from its initial value* $\mathcal{E}_i$ *to some final value* $\mathcal{E}_f$, *the atom emits radiation of frequency* f *in the form of quanta or photons of energy* hf *such that the energy of the system of atom plus quantum is conserved, or*

$$\boxed{\mathcal{E}_i - \mathcal{E}_f = hf.} \tag{42-7}$$

The radii of the permissible orbits can be obtained by eliminating v from Equations (42-4) and (42-6) and solving for r, yielding

$$r = n^2\frac{h^2}{4\pi^2me^2}. \tag{42-8}$$

By substituting known values of h, m, and e, and setting $n = 1$, we find that the radius of the first orbit is

$$r_1 = 0.529 \times 10^{-8}\text{ cm} = 0.529\text{ A}.$$

The energy of any orbit characterized by any quantum number n can

be determined by eliminating r from Equations (42-5) and (42-8), yielding

$$\mathcal{E} = -\frac{2\pi^2 me^4}{n^2h^2}. \tag{42-9}$$

Applying Bohr's second postulate, we find that the frequency f of the radiation emitted when an electron goes from its initial orbit of quantum number n_i to another orbit n_f is

$$f = \frac{\mathcal{E}_i - \mathcal{E}_f}{h} = \frac{2\pi^2 me^4}{h^3}\left(\frac{1}{n_f^2} - \frac{1}{n_i^2}\right). \tag{42-10}$$

Here is a result that can be tested by comparison with experiment. Equation (42-10) may be re-expressed in terms of the wavelength of the emitted line by making use of the equation $\lambda f = c$, where c is the speed of light. We find that

$$\frac{1}{\lambda} = \frac{2\pi^2 me^4}{ch^3}\left(\frac{1}{n_f^2} - \frac{1}{n_i^2}\right). \tag{42-11}$$

This equation can be checked directly against experimental results. On comparison with Equation (42-2), which represents the experimental values of the wavelengths of the Balmer series, we find that our theoretical expression predicts a value for the Rydberg constant R of

$$R = \frac{2\pi^2 me^4}{ch^3} = 109{,}740 \text{ cm}^{-1},$$

the latter value calculated from known values of the constants in the equation. This value is in remarkably good agreement with the experimental value of R quoted earlier. The Bohr theory thus provided the first successful and satisfying explanation of the emission of sharp spectral lines by hydrogen, not only qualitatively but quantitatively as well. The model not only gave correct values of the wavelengths of the Balmer series, and the other spectral series of hydrogen, but explained the reason for the integers appearing in the wavelength formulas.

The hydrogen atom was pictured as a series of imaginary concentric rings surrounding the nucleus, as shown in Figure 42-5, representing the allowed orbits of the electron. Each ring was characterized by a quantum number n which described the radius of the orbit, the angular momentum of the electron in the orbit, and the energy of the electron in the orbit. At any instant of time a neutral hydrogen atom had an electron in one of these orbits. In emission, an electron could make a spontaneous transition from any of these orbits to any other orbit of lower energy, that is, closer to the nucleus, and in this transition light would be emitted of appropriate energy, or wavelength. Thus if the electron jumped from orbit $n = 4$ to

orbit $n = 2$, the wavelength of the radiation corresponded to the blue line H_β ($\lambda = 4861.3$ A) of the Balmer series. In a gaseous discharge tube, collisions between neutral atoms and charged ions caused the atoms to become ionized; that is, the atoms were given sufficient energy in the collision to overcome the binding energy of the electron to the proton. The

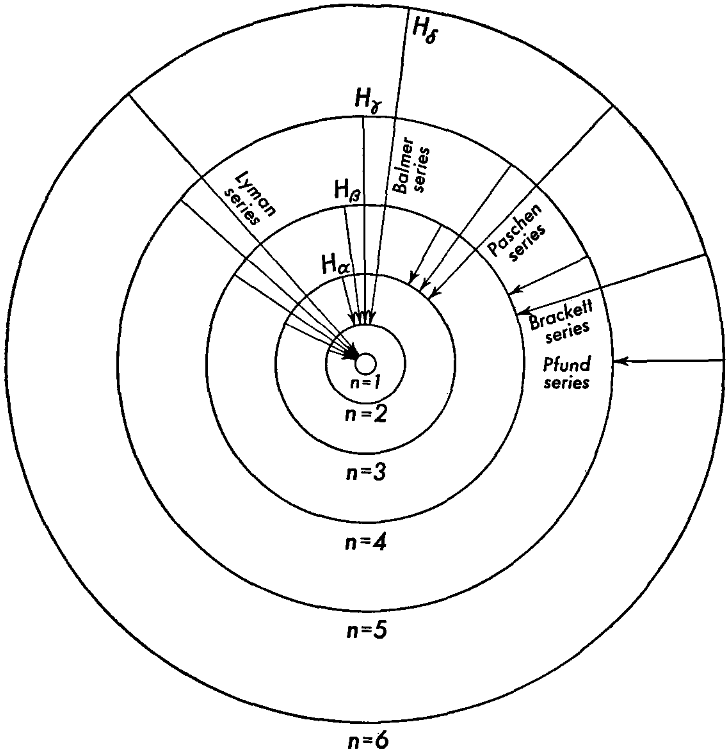


Fig. 42-5 Possible quantum jumps between stationary orbits giving rise to the different spectral series of hydrogen.

electron then had positive energy of any amount. Since any system generally seeks its configuration of lowest potential energy, an unbound electron would tend to drop into one of the unfilled orbits of neutral hydrogen, with the emission of a quantum of radiation. On passing from the outer orbits to the orbit for which $n = 1$, the electron might fall into intermediate orbits for a short time emitting an appropriate quantum of radiation in each transition. Since any small quantity of hydrogen contains an enormous number of atoms, all of the lines of the hydrogen spectrum are emitted simultaneously from a gas discharge tube.

42-4 The Hydrogen Energy-Level Diagram

The results discussed above can be represented in a simple and elegant manner by means of an energy-level diagram, as shown in Figure 42-6. In this diagram the energies of the various orbits corresponding to the different quantum numbers n are plotted on a vertical scale. Two different sets of numbers are shown in the diagram. The numbers on the right are the reciprocals of wavelengths in centimeters, and, when multiplied by the constants ch, will yield the energy $\mathcal{E}$ in ergs. Differences between these numbers for any two energy levels give the reciprocal of the wavelength, called the *wave number* (the number of waves per centimeter), of the radiation emitted when an electron goes from an upper level to a lower one. The numbers on the left give the energy in electron volts. Recalling that an electron volt is the energy acquired by an electron in falling through a potential difference of 1 volt, we see that the energy, expressed in electron volts, may be converted to energy expressed in joules by multiplying by the charge of the electron, in coulombs. On the left of the diagram the lowest level, called the *ground state*, is taken as the zero of energy. Thus we see that 13.60 ev of energy must be delivered to a neutral hydrogen atom, in its ground state, to ionize the atom. This is called the *ionization potential*, for if hydrogen is placed in a gaseous discharge tube which is provided with electrodes, there can be no significant conduction of electricity through the gas until the electrode potential difference is higher than the ionization potential. Beyond the level designated as $n = \infty$, the electron is outside the atom and can have any value of energy; this is the shaded region of the diagram. The numbers on the vertical lines between any two levels represent the wavelength in Angstrom units of the radiation emitted when the energy of the atom changes from one level to another. All lines ending at the same energy level represent spectral lines of the same series. The thickness of the lines in the diagram represents the relative intensities of the spectral lines, although it must be pointed out that one of the failings of the Bohr theory is that it makes no attempt to predict the line intensities. This prediction has been accomplished with the aid of more advanced methods called *wave mechanics* or *quantum mechanics*.

The advantage of an energy-level diagram is that it is independent of any particular model of the atom. Such a diagram may be constructed from experimental data using only the wavelengths of the spectral lines of an atom. The construction is based upon Bohr's postulate that the frequency of spectral lines is determined by Equation (42-7), regardless of the particular mechanical model chosen to represent the atom. Thus all spectra, whether atomic, molecular, or nuclear, are analyzed in terms of energy-level diagrams, based upon Bohr's analysis of hydrogen.

At temperatures where the mean translational energy of $\frac{3}{2}kT$, where k is the Boltzmann constant and T is the absolute temperature, is consider-

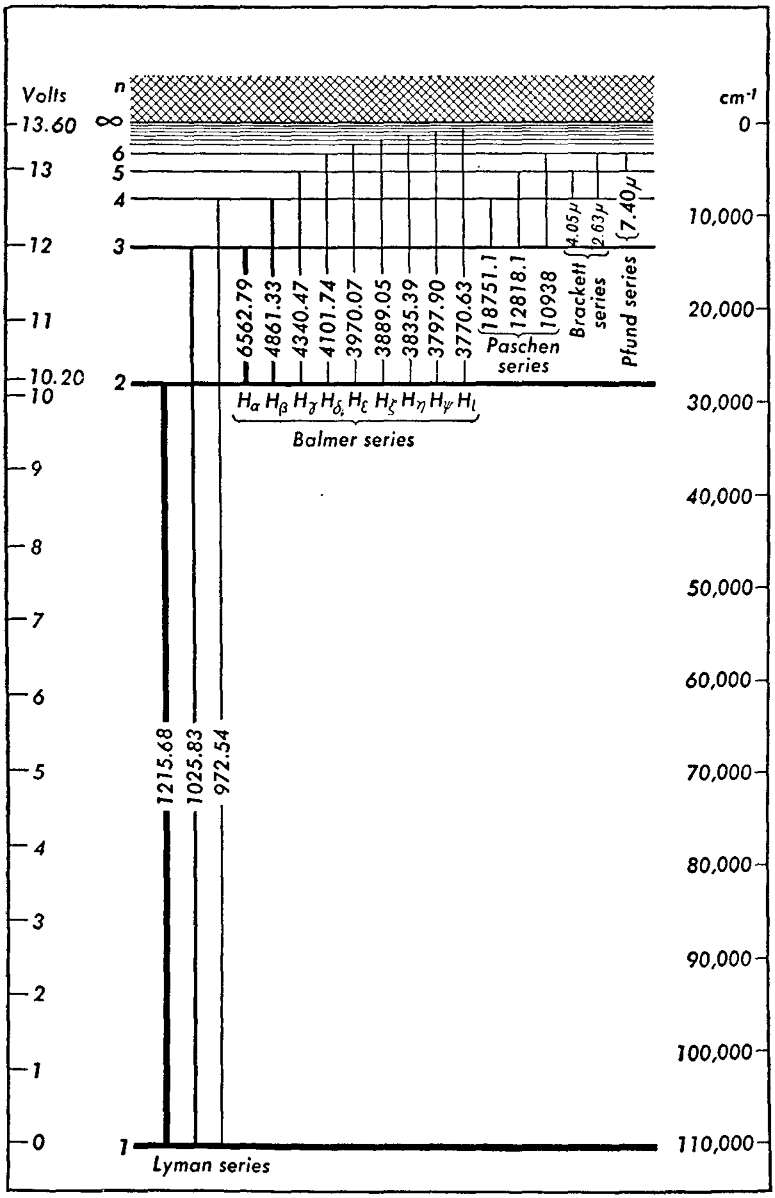


Fig. 42-6 The energy-level diagram for hydrogen.

ably less than the difference in energy of 10.20 ev between the ground state and the first excited state, we would expect to find the vast majority of hydrogen atoms in the ground state, for there is insufficient energy available to raise atoms from the $n = 1$ state to the $n = 2$ state. Thus

only lines corresponding to transitions from the $n = 1$ state to higher levels should occur in absorption, at such temperatures. Only when the gas is at sufficiently high temperatures should we expect to find absorption lines of the Balmer series, for only at high temperatures will there be a significant population of hydrogen atoms whose electrons are in the $n = 2$ level. Thus the question of why not all of the spectral lines appear in an absorption spectrum is resolved by the Bohr theory.

If the temperature is sufficiently high, the kinetic energy of translation exceeds 13.6 ev, and the gas is completely ionized. In this event we would expect that the protons and electrons act as separate particles which are not bound to each other and which are free to be much closer together than would be permitted by the Bohr quantum conditions. We would expect to see no spectral lines from a gas at sufficiently high temperatures. Thus the sun and the stars emit continuous spectra rather than line spectra. The line spectra are formed in absorption when the continuous spectrum passes through the cooler gases in the outer atmospheres of the sun and the stars. Furthermore, the stars and the sun consist of gases at high temperatures, but these gases are of much greater density than the solid matter on earth. On earth we must expect that the closest distance of approach of two hydrogen atoms is approximately the diameter of the first Bohr orbit. At the temperature of a star, the electron is unbound, and there is no such restriction, so that the distance of approach of two protons, or of a proton and an electron, is limited only by their kinetic energies, and stellar densities may seem enormous when compared to the densities of solid substances on earth.

When hydrogen is illuminated with light whose quantum energy is greater than 13.6 ev, each quantum has sufficient energy to liberate an electron from the atom. This is the *photoelectric effect* in a gas. Light of longer wavelength cannot cause ionization of hydrogen gas, no matter how intense the beam or how long the period of illumination. This prediction from the energy-level diagram is in agreement with experimental results but is quite different from our expectations based on classical theory. From a purely classical viewpoint, the energy flux in a light beam is related to its vibration amplitude, so that light of any frequency should be able to liberate photoelectrons if its intensity is sufficiently great. Here too, the quantum viewpoint is in agreement with experiment, while the classical analysis is not.

42-5 Atomic Structure

Attempts to extend the Bohr theory quantitatively to more complex atoms met with failure. The best that could be accomplished was to use the Bohr model of hydrogen as a qualitative guide or, at best, in a semi-

quantitative way. For example, the Z electrons of a neutral atom of atomic number Z were assumed to be moving in circular orbits about the nucleus. Sommerfeld attempted to generalize Bohr's theory to permit the atom to have elliptic orbits, and while the Sommerfeld theory has been superseded, certain details of the model he used remain useful in implementing our understanding of atomic structure. Sommerfeld supposed that the electrons could move in elliptic orbits and found that the orbit could be described by a principal quantum number n and an azimuthal quantum number which we will call l, corresponding to modern notation. The energy of the electron was found to depend upon n exactly as in Bohr's calculation as given by Equation (42-9), but the angular momentum depended upon l and was given by $(lh)/(2\pi)$. For each principal quantum number n, the value of l could be any integer between zero and $n - 1$, corresponding to orbits of different eccentricity, as shown in Figure 42-7. Thus where $n = 3$, there are three possible orbits having angular momenta corresponding to $l = 0$, $l = 1$, and $l = 2$. In the elliptical orbit the principal quantum number designates the length of the semimajor axis, as it designates the orbital radius in the Bohr circle. Thus the elliptic orbits fall into shells for which $n = 1$ (called the K shell), $n = 2$ (called the L shell), $n = 3$ (called the M shell), and so on. The orbits within a shell are called S orbits if $l = 0$, P orbits if $l = 1$, D orbits if $l = 2$, and F orbits if $l = 3$. Thus we speak of an electron occupying a D orbit in the M shell, or of a $3D$ electron, when its orbit has a principal quantum number of 3 and an azimuthal quantum number of 2.

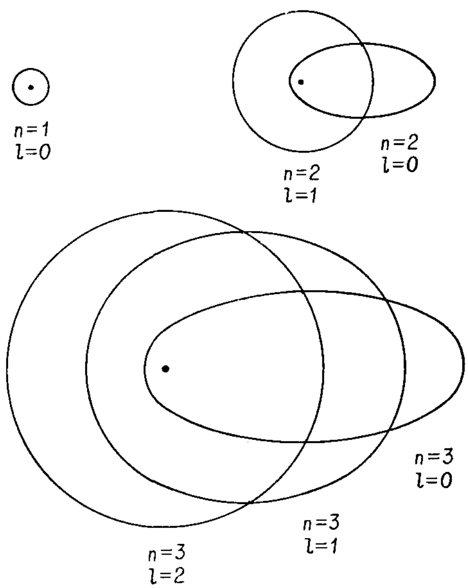


Fig. 42-7 Possible electronic orbits for a given total quantum number n.

An electron moving in a closed orbit is analogous to a loop of wire carrying current, in that the electron generates a magnetic field and also experiences a force when in a magnetic field which tends to orient the plane of the orbit perpendicular to the external field, as in the case of a galvanometer coil. Experiment has shown that not all orientations are possible for an atomic orbit but that the orbit may take up only certain permitted

orientations, according to a principle known as *space quantization.* The orbit can only take up those orientations in which the projection of the azimuthal quantum number representing the orbital angular momentum vector onto the direction of the magnetic field is itself an integer, called the magnetic quantum number m. Furthermore, the projection of l can be in the direction of the magnetic field or opposed to it. Thus we find that the quantum number m may have $2l + 1$ values ranging from $+l$ through 0 to $-l$. For an orbit for which $l = 2$, the quantum number m may take on the values 2, 1, 0, -1, -2. In addition to this perplexing number of parameters which we might call orbital quantum numbers, we already have mentioned that the electron has an intrinsic angular momentum called *spin,* designated by the letter s. Thus each electron in an atom is designated by four quantum numbers n, l, m, and s. The number n designates the shell in which the orbit is located, the number l designates the eccentricity of the orbit and its angular momentum, the number m designates the orientation of the plane of the orbit with respect to a magnetic field, and the number s designates the spin of the electron, which may be oriented parallel to or opposed to the direction of the magnetic field.

Together with an extremely valuable guiding principle introduced by Pauli in 1925, these quantum numbers serve as a vitally important guide to the problem of atomic structure. We would normally expect a system to fall into its configuration of lowest potential energy. Thus an atom with many electrons would have all of its electrons in the K shell, and all atoms would be essentially indistinguishable from each other, except perhaps through their different masses. But we know experimentally that this is not the case, and that atoms display characteristic valences. Furthermore, we know that the chemical properties of different atoms repeat in cycles in the periodic table.

To solve the problem of how electrons are arranged in the periodic table, Pauli proposed a principle, now known as the *Pauli exclusion principle,* which states that *no two electrons in an atom can exist in the same state.* Now the state of any electron can be completely specified by the four quantum numbers n, l, m, and s, so that the Pauli principle states that the group of values assigned to these four numbers must be different for all electrons in the atom. Subject to this condition, the electrons in an atom must occupy the states of lowest energy.

Thus an atom for which $Z = 1$ (hydrogen) must have its electron in the $n = 1$ shell. In this shell the value of l is 0, so that $m = 0$. In an atom for which $Z = 2$ (helium), there are two electrons which must fall into the K shell but which must have their spins oppositely oriented. No more than two electrons can occupy the K shell. An atom for which $Z = 3$ must locate its third electron in the L shell for which $n = 2$. This single electron

in the outermost shell is responsible for the fact that the chemical valence of lithium is $+1$. The fact that the K shell of helium is filled is responsible for the fact that helium is a noble gas. Thus we would expect that each time a shell is filled we would find an inert gas, and that each time a new shell is begun we would find a substance of valence $+1$. To fill the K shell requires 2 electrons; to fill the L shell requires 8 electrons, so that the next element of unit valency is sodium for which $Z = 11$.

While there are many complications in the periodic table which require far more elaboration than it is possible to give here, the basic structure and pattern of the table and of the structure of atoms are determined by the energy-level diagram of hydrogen and by our theory of the hydrogen atom. A table of the distribution of electrons in atoms is given as Table 5 in Appendix A, while the periodic table of the elements is given as Table 6 in Appendix A. From Table 5, Appendix A, we may note that the shells fill in regular order until we reach potassium, for which $Z = 19$. At this point we note that the S level in the N shell begins to be occupied before the D level in the M shell. In the case of iron, cobalt, and nickel, the $4S$ levels are occupied while the $3D$ levels are being filled. The spins of the $3D$ electrons may reorient themselves without altering the chemical structure of a molecule, or the crystal structure of a solid, for these depend upon the outermost electrons. It is this distortion in the level structure which is responsible for ferromagnetism. Another interesting aberration occurs in the case of the *rare earths*, beginning with element 57. Here the outer shells are nearly identically filled, but deep inside the atom, two shells removed from the valence electrons, the electrons populate the $4F$ subshell, for which $n = 4$, and $l = 3$. This results in nearly identical properties for the rare-earth elements and makes purely chemical separation of these elements extremely difficult.

42-6 Changes in Our Fundamental Concepts

Around the beginning of the present century, many changes began to take form in our fundamental concepts of physics. Light, which had been successfully explained as a wave motion, was found to have particlelike characteristics. The electron, which originally was an extremely small charged particle, was found to have spin and a magnetic moment. The concept of *electron spin* was introduced by Uhlenbeck and Goudsmit in 1925 to explain a fine splitting of atomic energy levels, called the *fine structure*. This concept was later found to be of great value in explaining the magnetic properties of matter. The Pauli exclusion principle was a special rule for the organization of atoms and electrons which had no classical analogue and which proved invaluable in explaining the electrical properties of metals as well as the periodic table. Another concept which

has played an important role in modern physics is one of the complete indistinguishability of fundamental particles; that is no two electrons can be distinguished from each other. The formal recognition that it is impossible to label individual electrons has had important consequences in theoretical physics, as in the theory of valency.

In 1925 another new concept was introduced by Louis de Broglie to the effect that every particle of mass m moving with speed v has associated with it a wave of wavelength λ given by

$$\lambda = \frac{h}{mv}, \tag{42-12}$$

where h is Planck's constant. Thus an electron, as well as any other particle, should have a wave associated with it. The existence of waves associated with electrons was verified in a series of remarkable experiments in 1927 and 1928 by Davisson and Germer, and by G. P. Thomson, and has led to an entirely new field of physics known as *electron optics.* These new ideas concerning the spin of the electron and the waves associated with electrons in motion were incorporated into a new theory known as *wave mechanics* or *quantum mechanics,* mainly through the work of Heisenberg, Schroedinger, Dirac, and Pauli, beginning about 1926, in which the wave properties of electrons and the quantum character of atomic energy levels was built into the theory, as the inertial character of mass is built into Newton's second law of motion. As a result of these developments, remarkably rapid progress was made in the next few years in describing atoms, molecules, and nuclei. We shall consider these concepts in greater detail in the succeeding chapters.

Problems

42-1. The red line of the Balmer series of hydrogen has a wavelength of 6,563 A. Determine the energy of a quantum of this radiation (a) in ergs and (b) in electron volts.

42-2. The first line of the Lyman series of hydrogen has a wavelength of 1,216 A and is emitted in a transition from the second to the first Bohr orbits. Calculate the energy, in electron volts, of this transition.

42-3. Ultraviolet light of wavelength 800 A is incident upon a tube containing hydrogen. Determine the minimum kinetic energy with which an electron will be ejected from hydrogen.

42-4. At what absolute temperature would the mean kinetic energy of molecules of hydrogen in a gas tube equal the energy required to raise hydrogen from the ground state to the first excited state ($n = 2$)?

42-5. The light emitted by sodium atoms in a transition from the first excited state to the ground state is of wavelength 5,896 A. At what temperature

is the mean kinetic energy of sodium atoms equal to the energy required to raise sodium to its first excited state?

42-6. A neutron has a mass of 1.67×10^{-24} gm. What is the de Broglie wavelength of a neutron whose velocity may be determined by assuming the neutron to have a kinetic energy corresponding to a temperature of 20°C? [HINT: Assume $\mathcal{E}_k = kT$.]

42-7. Derive an equation for the energy with which an electron is bound to a nucleus whose charge is Ze.

42-8. From the result of Problem 42-7, find the energy, in electron volts, with which an electron is bound to sodium ($Z = 11$) in the state for which $n = 1$.

42-9. (a) Find the de Broglie wavelength of an electron in the nth Bohr orbit of hydrogen. (b) What is the relationship between this wavelength and the circumference of the nth orbit? (c) Is there any relationship between the Bohr condition on the quantization of angular momentum and the result you obtained in part (b).

42-10. The earth is 149×10^6 km from the sun and has a mass of 5.98×10^{24} kg. What is the de Broglie wavelength of the earth?

42-11. A ball which weighs 5 oz is pitched with a velocity of 100 ft/sec. What is the de Broglie wavelength of the ball? Would you expect such a ball to display predominantly wavelike or particlelike behavior?

42-12. An electron has a mass of 9.1×10^{-28} gm and a charge of 4.8×10^{-10} stcoul. What is the de Broglie wavelength of an electron which has been accelerated through a potential difference of 100 volts? Would you expect this electron to display wavelike or particlelike behavior in the presence of obstacles 1 mm in diameter?

42-13. The negative μ-meson has a mass of 206 times the mass of the electron and a charge equal to the charge of the electron. Such a meson can form a mesic atom by falling into an appropriate Bohr orbit. Find the radius of the first Bohr orbit of the meson in silver.

42-14. According to the *correspondence principle* of Bohr, we must expect phenomena to look alike from the classical and quantum viewpoints in the limit of large quantum numbers, for we know that the classical picture of matter is reasonably accurate for macroscopic phenomena. Classically, we would expect that the frequency of light emitted from the hydrogen atom would equal the frequency of rotation of the electron in its orbit. Show that, for very large quantum numbers, the frequency of rotation of the electron and the frequency of the radiation emitted when the quantum number changes by unity approach the same value $f = 2cR/n^3$.

43

X-Rays

43-1 Discovery of X-Rays

The study of the electric discharge through gases led directly to the discovery of x-rays by W. C. Roentgen in 1895. While operating a gas-discharge tube, Roentgen observed that a platinum-barium cyanide screen at some distance from the tube fluoresced. He shielded the tube so that no visible radiation could reach the screen, but the fluorescence could still be observed. On interposing various materials between the tube and the screen, he found that the intensity of the fluorescence could be diminished, but that it was not completely obliterated. He interpreted these observations as being due to radiation coming from the walls of the tube which penetrated the absorbing screens and caused the screen to fluoresce, and he called the new radiation *x-rays* to indicate their unknown character. The x-rays were produced when the cathode rays struck the glass walls of the electric-discharge tube.

Ever since their discovery, x-rays have played an important part in the investigations of atomic physics and have added immeasurably to our knowledge of the structure of the atom. Almost immediately after their discovery, they were used by physicians as aids in diagnosis, and later in therapy. Industry turned to the use of x-rays for the study of the internal structure of materials and for the examination of castings to determine the presence of internal voids.

X-rays are produced whenever a stream of electrons strikes some substance and are rapidly accelerated. The electrons may come from the cathode of a gas-discharge tube, or, more commonly, from a heated filament. In the Coolidge-type x-ray tube, shown in Figure 43-1, a high potential difference, ranging from a few thousand volts to about a million volts, is applied between the cathode and anode. The electrons are accelerated and strike the target, that is, the anode, causing it to emit x-rays. For many purposes it is desirable to limit the region from which x-rays are

emitted, called the focal spot, and this may be done by surrounding the filament with a metallic cup which shapes the lines of force in such a way as to focus the electrons into a small region on the target. It is difficult to maintain potential differences greater than a million volts between two electrodes separated by the glass walls of a vacuum tube. The *betatron,*

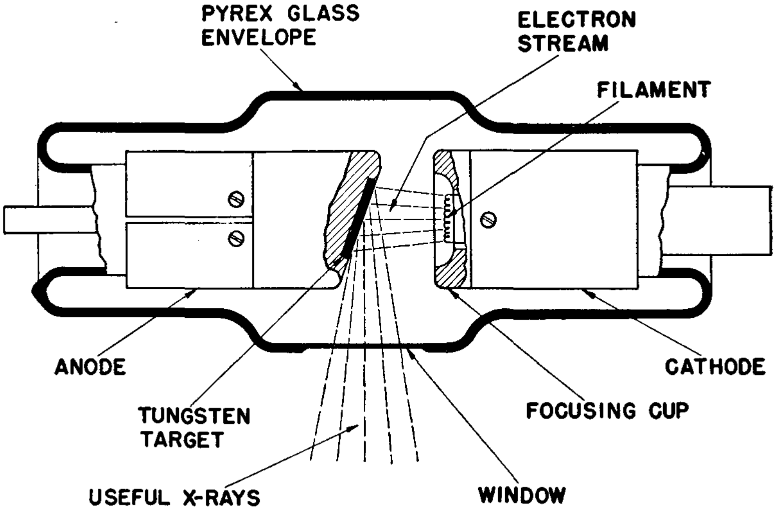


Fig. 43-1 A modern Coolidge-type x-ray tube. (Courtesy of General Electric X-ray Department.)

developed by D. W. Kerst in 1941, utilizes Faraday's law of electromagnetic induction as the means of accelerating electrons. The betatron is pictured in Figure 43-2 and is shown schematically in Figure 43-3.

In Figure 43-3 we see that electrons from a heated filament F are accelerated by a small difference in potential applied to an injector grid G inside a doughnut-shaped vacuum tube. The electrons are made to travel in a circular path by a magnetic field perpendicular to the plane of the paper, at the position of the orbit. The electrons are constrained to move within a ring of small cross section by the shape of the magnetic field. Electrons moving within the vacuum tube may be thought of as though they were moving in a circular loop of wire or in the coil of a transformer. The magnetic flux is caused to change rapidly through the central portion of the doughnut. This changing flux induces an emf such that each time the electron completes a circular turn, its energy has increased by the product of the electron's charge by the emf, as in the case of the transformer. As the energy of the electron is increased, it is necessary to adjust the magnetic field intensity at the orbit to keep the electrons rotating in a circle

of fixed radius. In many betatrons the changing magnetic field is produced by supplying 60-cycle alternating current to the flux coils of the electromagnet. Electrons are injected into the tube for a very short time at the beginning of a cycle and continue traveling around the circular orbit until the magnetic flux reaches its maximum value, in 1/240 sec. Each electron makes hundreds of thousands of revolutions during this quarter cycle,

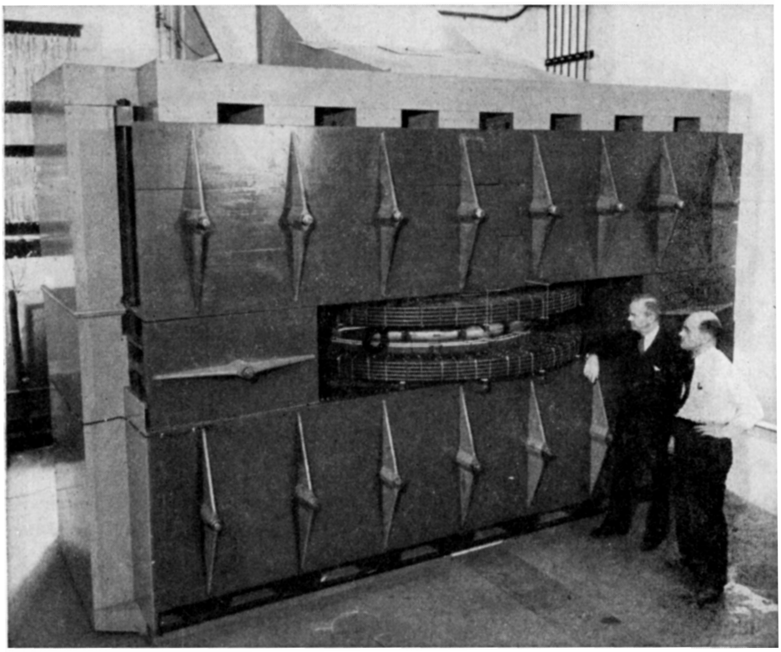


Fig. 43-2 The 100,000,000 electron-volt betatron. The betatron tube is in the center between the poles of the electromagnet. (Courtesy of General Electric Company.)

gaining energy in each revolution. When the electron has gained its maximum energy, current is sent through an additional set of coils, expanding the radius of the electron orbits so that the electrons strike a target, generating x-rays.

Betatrons are now being operated at energies up to 300 million electron volts (abbreviated Mev) and modified forms of betatrons are operating at higher energies. Since the electrons are injected at the beginning of a cycle, the x-ray output occurs in short pulses, at the peak of each cycle, when the electrons have reached maximum energy. The electrons are moving with speeds very close to the velocity of light, and at these high speeds the emitted x-rays are practically confined to a narrow cone of aperture ranging from 2° to 15° in the forward direction.

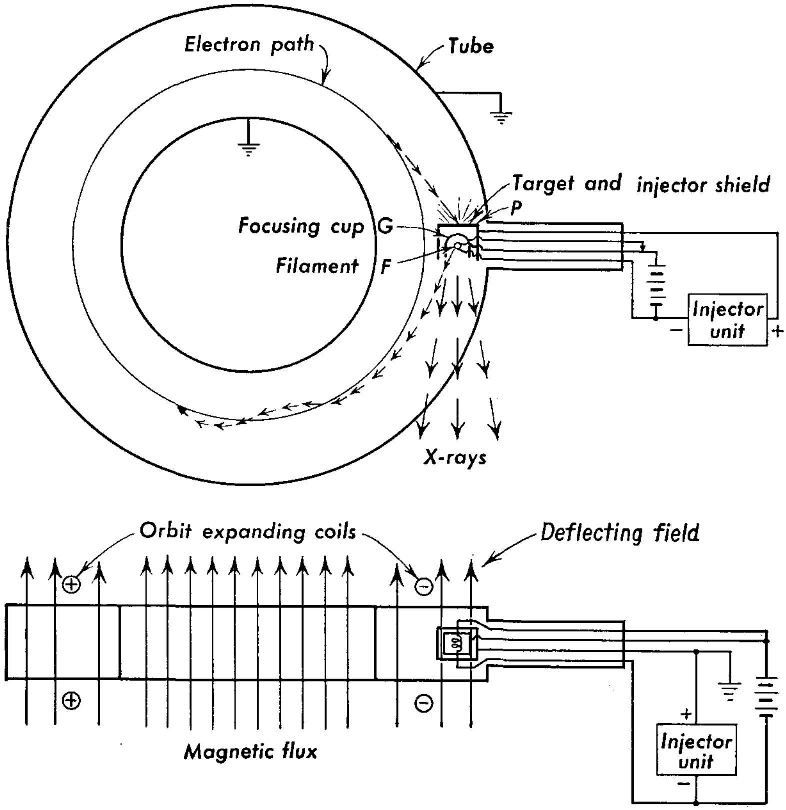


Fig. 43-3 Path of an electron in a betatron tube.

43-2 Diffraction and Interference of X-Rays

Early attempts were made to determine the character of x-rays, but the results were quite puzzling. The rays were not deflected by an electric or magnetic field and so did not consist of charged particles. Although it is known today that x-rays are refracted very slightly by matter, early attempts to show refraction failed. The rays passed through electrical conductors, contrary to our expectations for the behavior of an electromagnetic wave. Early attempts to produce interference effects also failed. In 1912 von Laue argued that the reason that interference effects had not been demonstrated was the very short wavelength of x-rays, and suggested that x-rays might have a wavelength of the order of the spacing of the atoms in the lattice of a crystalline solid. With his collaborators he was able to produce diffraction effects by passing a beam of x-rays through a thin crystal.

If a narrow pencil of x-rays is sent through a small thin crystal, such as a piece of rock salt, and received on a photographic plate a short distance away, as shown in Figure 43-4, the photograph will show a series of small spots arranged in a definite pattern. As shown in Figure 43-5, this pattern,

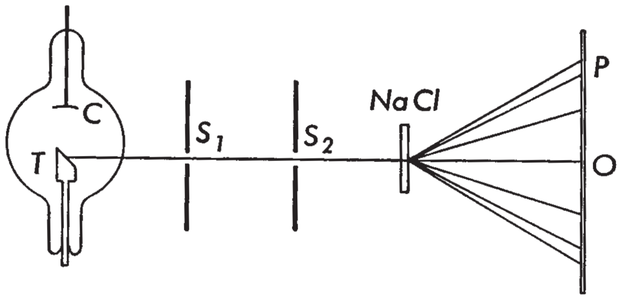


Fig. 43-4 Arrangement of apparatus for producing a Laue diffraction pattern using a rock salt crystal. S_1 and S_2 are pinholes and P is the photographic plate.

as obtained with rock salt, has a fourfold symmetry characteristic of a cubic crystal. The pattern may be interpreted by thinking of each ion of the crystal as a scattering center for the incident beam of x-rays. The crystal thus acts as a three-dimensional diffraction grating. The pattern

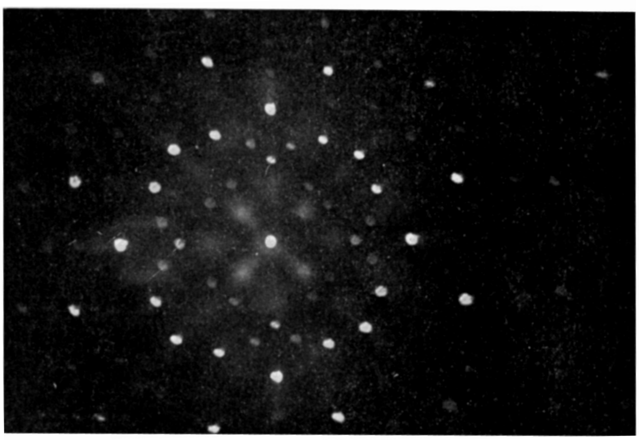


Fig. 43-5 Photograph of Laue diffraction pattern of rock salt. (From photograph by J. G. Dash.)

obtained on the photographic plate is called a Laue pattern. From the distribution and intensities of the points on the photographic plate, the arrangement of the ions in the crystal can be deduced.

A slightly different arrangement of the x-ray beam and crystal, used by Bragg, gives a simpler pattern which may be interpreted more easily.

This is shown in Figure 43-6, in which x-rays coming from the target T of the x-ray tube pass through two narrow slits, and are then incident upon the face of a crystal which is mounted on a spectrometer table. The crystal scatters the x-rays in all directions, but the photographic plate is set so as to receive only that part which comes from the face of the crystal.

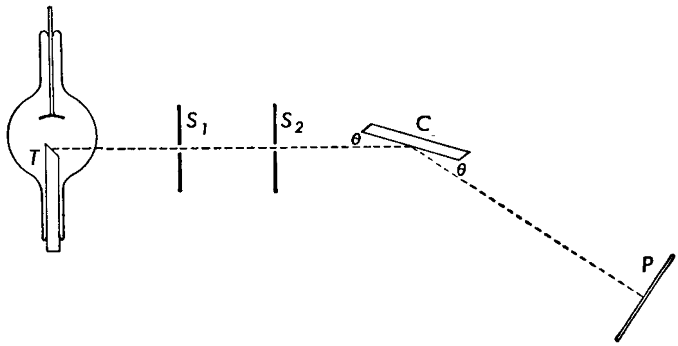


Fig. 43-6 The single crystal x-ray spectrometer with photographic plate. S_1 and S_2 are narrow slits, C is the crystal, and P is the photographic plate.

The angle θ between the incident beam and the face of the crystal is changed slowly by rotating the crystal. In general, the photograph will show a series of sharp lines against a continuous background. If an ionization chamber is used as a detector, in place of a photographic plate, it is found that the intensity of the x-rays is a maximum when the scattered beam makes an angle θ with the face of the crystal equal to the angle that the incident beam makes with it. For this reason the beam is sometimes said to be "reflected" from the crystal.

A simple explanation of the effect of the crystal in this case may be given with the aid of Figure 43-7, in which the ions of the crystal are arranged in layers parallel to the surface of the crystal. The distance d between atomic layers is shown greatly enlarged. The x-ray beam incident upon the crystal at an angle θ with its face penetrates the crystal and is scattered by the ions in all directions. Consider two rays I and II, very close together, and consider only that part of the scattered beam which makes an angle θ with the surface of the crystal. Ray I strikes the upper surface at A and is reflected; ray II strikes the next layer at B and is reflected. These two rays, which are so close together that they interact with the same grain of photographic emulsion, have traveled different distances. From the figure it is evident that ray II has traveled a longer distance than ray I. If originally they started out in phase, they will now differ in phase because of the difference in paths of $CB + BD$. If this difference in path is an integral number of wavelengths $n\lambda$, the two rays

will reinforce each other. Thus we have constructive interference when

$$CD + BD = 2d \sin \theta = n\lambda. \tag{43-1}$$

Equation (43-1), called the *Bragg equation*, is of fundamental importance in determining x-ray wavelengths and in the analysis of the structure of crystals. As in the case of the diffraction grating, we refer to the bright spot when $n = 1$ as the first-order spectrum, when $n = 2$ we have the second-order spectrum, and so on. When x-rays of a particular wave-

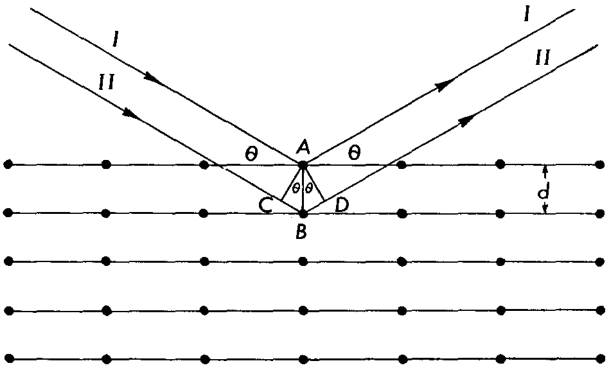


Fig. 43-7 Reflection of x-rays from atomic planes.

length are incident upon a crystal, reflection occurs only at angles where the Bragg equation is satisfied. Although the above development considered atomic planes parallel to crystal face, we may consider any set of planes drawn through the atoms of the crystal as capable of reflecting x-rays. If we consider the first-order diffraction pattern produced when a continuous spectrum of x-rays is incident upon a crystal, as in the case of the Laue pattern, we see that a bright spot is produced in diffraction for those wavelengths and angles fulfilling the conditions of the Bragg equation. If a crystal has a fourfold symmetry with respect to the direction of the incident beam, the spots produced by one set of crystal planes will appear four times in the pattern, corresponding to the symmetry of the crystal. For each set of planes there is a different value of the grating space d.

Illustrative Example. An x-ray line of wavelength 1.541 A is reflected from a set of planes in a quartz crystal in which the distance d between planes is 4.255 A. Determine the angle between the x-ray beam and the atomic planes in the second order.

From the Bragg equation we find

$$\sin \theta = \frac{n\lambda}{2d},$$

and, substituting numerical values, we have

$$\sin\theta = \frac{2 \times 1.541 \times 10^{-8}\,\text{cm}}{2 \times 4.255 \times 10^{-8}\,\text{cm}} = 0.362,$$

$$\theta = 21^\circ\,14'.$$

43-3 X-Rays and Crystal Structure

In using the Bragg equation, it is essential that we know independently either the x-ray wavelength λ or the spacing d between atomic or ionic planes. At the time of the discovery of x-ray diffraction, the crystal grating

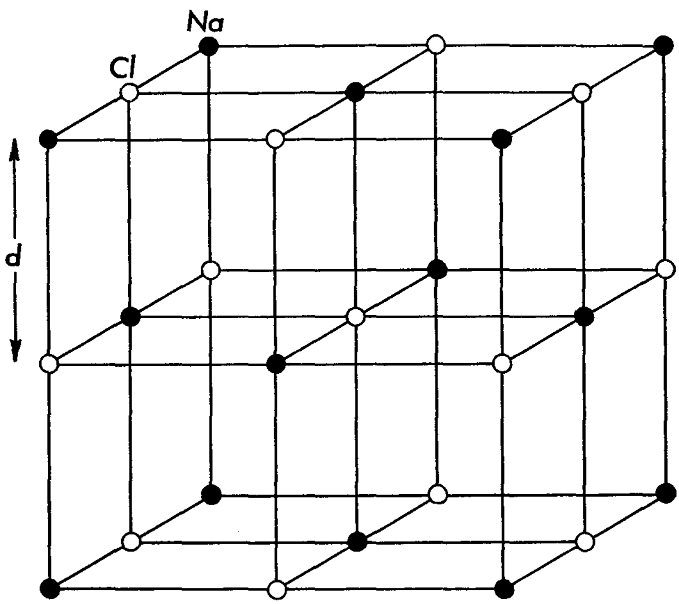


Fig. 43-8 The arrangement of sodium (Na) ions and chlorine (Cl) ions in a crystal of salt.

space of rock salt was determined from other data, and the value of d so obtained was used to measure x-ray wavelengths. Figure 43-8 shows the arrangement of the ions in a rock-salt crystal. From crystallographic studies, the crystal is known to be a cube; the centers of the ions are at the corners of the cubes. Each sodium ion (Na^+) is surrounded by 6 chlorine ions (Cl^-), and each chlorine ion is surrounded by 6 sodium ions. If d is the length of the smallest cube which can be drawn through neighboring ions, then the volume of each cube is $V = d^3$. This is the volume associated with each ion. If M is the gram molecular weight of sodium chloride

and ρ is its density, then the volume v of 1 mole of sodium chloride is

$$v = \frac{M}{\rho}.$$

Now there are $2N_0$ ions in each mole of sodium chloride, where N_0 is the Avogadro number; hence the volume associated with each ion is

$$V = \frac{v}{2N_0} = \frac{M}{2\rho N_0}.$$

The distance d between ions is therefore

$$d = \left(\frac{M}{2\rho N_0}\right)^{1/3}. \tag{43-2}$$

Let us calculate the grating space d of rock salt. The gram molecular weight M is the sum of the gram atomic weights of sodium and chlorine and is

$$M = 22.997 + 35.457 = 58.454 \text{ gm};$$

the density is $\rho = 2.164 \text{ gm/cm}^3$ at 18°C,

and $N_0 = 6.0248 \times 10^{23}$;

hence we find $d = 2.820 \times 10^{-8} \text{ cm} = 2.820$ A, at 18°C.

With the value of the grating space of rock salt known, it is now possible to measure x-ray wavelengths, using a rock salt crystal as a

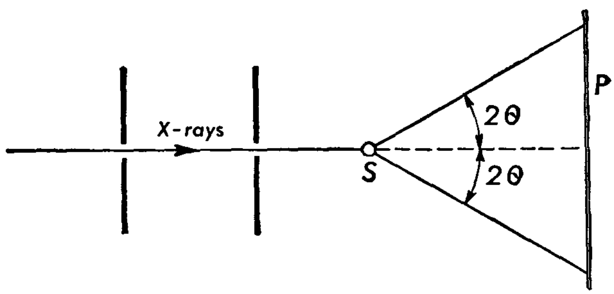


Fig. 43-9 Method of obtaining x-ray diffraction patterns using a powder.

spectrometer. The crystal used must be sufficiently large and well developed, say of the order of 1 to 2 cm in length and width. Such large crystals, sometimes called single crystals, may be found in nature and may sometimes be grown from saturated solutions or by the slow cooling of molten material in the case of metals. Once the wavelength of an x-ray beam is known, that beam may be used to determine the structure of other crystals.

The ordinary solid is not a large crystal but is made up of many very small crystals called *microcrystals*. Even when the solids are in powder form, they consist of many microcrystals. A very powerful method of x-ray analysis was developed by A. W. Hull, and independently by P. Debye and P. Scherrer. This method consists in sending a narrow pencil of x-rays of a single wavelength through a very small sample of the powder or solid, as shown in Figure 43-9. Since the powder consists of a great many

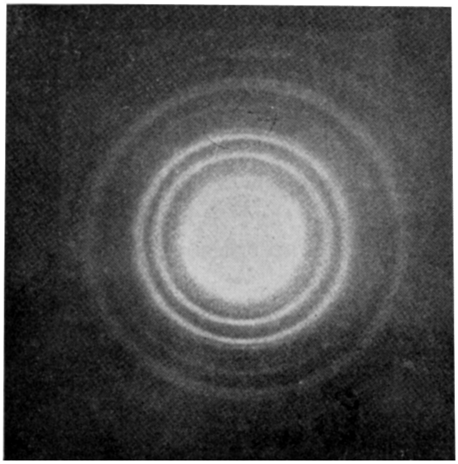


Fig. 43-10 X-ray powder diffraction pattern of aluminum. (Reproduced with the permission of A. W. Hull.)

microcrystals oriented at random, there is some probability that one of these microcrystals will be oriented so that its atomic planes make an angle θ with the incident radiation which will satisfy the Bragg equation. If the crystals are randomly oriented, the radiation reflected from this plane will be symmetrically distributed with respect to the incident beam and will form a conical shell centered at the sample S. A photographic plate, placed at right angles to the direction of the incident beam at a convenient distance from the sample, will record the diffraction pattern as a set of circles, as shown in Figure 43-10. Each circle corresponds to a particular set of atomic planes whose spacing may then be determined. The planes we draw through the atoms of the crystal lattice will have different numbers of atoms per square centimeter of plane. Furthermore, in a crystal, such as sodium chloride, it may be possible to draw planes containing atoms of one kind, as all sodium atoms or all chlorine atoms. These planes reflect x-rays with different intensities. When the resulting diffraction pattern is interpreted, each of these details must be accounted for.

In a variation of this method, the photographic plate is replaced by a photographic film bent in the form of a cylinder with the sample at its center. Holes are cut in the film so that the direct pencil of x-rays can

enter and leave this camera without blackening the film. When this film is unwrapped and developed, we obtain the type of pattern shown in Figure 43-11. The x-ray powder diffraction pattern has lines whose position and intensity are determined by the crystal itself. Thus x-ray patterns may be used to identify unknown crystals. The procedure followed is to

Fig. 43-11 X-ray powder diffraction pattern of tungsten obtained with a photographic film bent in the form of a circular cylinder. X-rays from a copper target were used in making this photograph. (From a photograph made by L. L. Wyman and supplied by A. W. Hull.)

prepare patterns from a large number of known substances, and to code these patterns in terms of the d spacings of the three most intense lines obtained in the powder pattern. The pattern obtained with an unknown substance may then be identified, in much the same way that the atomic composition of an unknown is determined from its optical spectrum.

43-4 X-Ray Spectra and Atomic Structure

When x-rays from any target are analyzed with a crystal spectrometer, the spectrum is found to consist of a series of sharp lines superimposed on a continuous background of radiation, as shown in Figure 43-12. The energy in the continuous portion of the spectrum is found to depend upon the voltage across the tube, the current passing through the tube, and the atomic number of the element which constitutes the target; the higher the atomic number, the greater is the energy, the other quantities remaining constant. The sharp lines which are superimposed on the continuous radiation are found to be characteristic of the target.

The first systematic study of the characteristic x-ray spectra of the elements was made by Moseley in 1913. He used a modification of the Bragg method in which the crystal spectrometer and the photographic plate were placed in an evacuated chamber to avoid absorption in air of the x-rays of long wavelength. Each element investigated was used as the target of an x-ray tube. He found that all the elements gave similar types of spectra; the lines emitted by each element were classified into two groups or series: a group of short wavelengths called the K series and a group of comparatively long wavelength called the L series. These two series are widely separated from one another in wavelength, as illustrated in Figure 43-13 for the case of silver. Other investigators have found two other series of still longer wavelengths in the heavier elements, classified as M series and N series. One of the important results of Moseley's work was the discovery that the square root of the frequency of the K_α lines of the

elements he investigated was proportional to the atomic number Z. Moseley's work provided certain evidence that elements were missing from

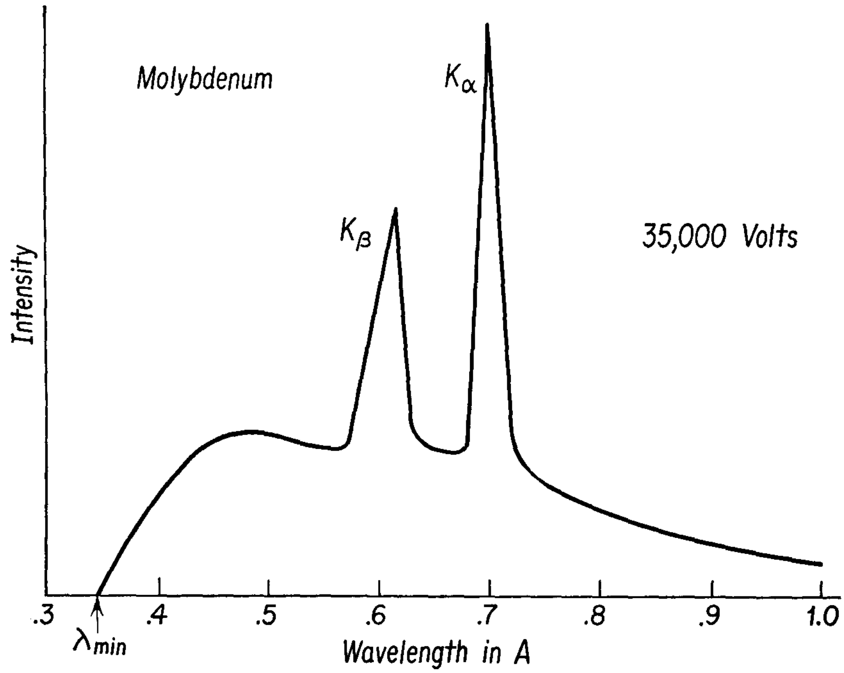


Fig. 43-12 Characteristic K_α and K_β lines superposed on the continuous x-ray spectrum of molybdenum. Note the sharp cutoff at the wavelength λ_{min}.

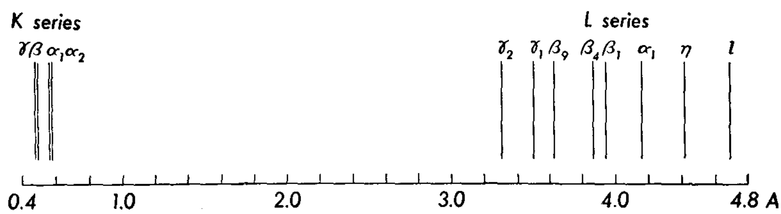


Fig. 43-13 Relative positions of the K and L x-ray series lines of silver.

the periodic table, as known at that time, and that the atomic number rather than the atomic weight was the key to the properties of the elements.

The characteristic x-ray spectra of the elements fit neatly into our theory of atomic structure. For example, the frequency of the most intense line of the K series can be written in the form

$$f = cR(Z - 1)^2\left(\frac{1}{1^2} - \frac{1}{2^2}\right), \tag{43-3}$$

in which R is the Rydberg constant, c is the velocity of light, and Z is the atomic number of the element emitting this line of frequency f. The interpretation of this equation is that this line is emitted when an electron goes from the L shell in which the principal quantum number is 2 to the K shell in which the principal quantum number is 1. All lines of the K series are associated with the transition of an electron from an outer orbit to the K shell.

To understand the appearance of the factor $(Z - 1)$ rather than Z in the equation, let us consider how the x-rays are produced. In the atoms of the elements investigated, the orbits of the principal quantum numbers 1 and 2 are completely filled. An electron from the cathode of the x-ray tube strikes the target with energy Ve, where V is the potential difference between the anode and cathode of the x-ray tube. When such an electron approaches an atom in the target, this electron may knock an electron out of the orbit for which $n = 1$. Since the other electrons in this atom are in their normal states, the electron which has been ejected from the K shell is removed outside the atom. The atom is thus ionized and left in an excited state. It is highly probable that an electron from the L shell may jump into the vacancy left behind in the K shell, thereby emitting radiation of frequency f, given by Equation (43-3). As for the factor $(Z - 1)$, since normally the orbit for which $n = 1$ has 2 electrons, when one is removed the other electron still remains in this orbit, and effectively "screens" the Z positive charges of the nucleus. Hence when an electron goes from the L shell to the K shell it moves in an electric field due to Z positive charges and 1 negative charge, or effectively $Z - 1$ positive charges.

The interpretation of the x-ray spectra is simplified by the use of an energy-level diagram such as that shown in Figure 43-14, Let us take the normal atom with all of its electrons in their normal states as our zero level of energy. When an electron is removed from the innermost orbit, a certain amount of energy must be supplied to do this; let us call this amount of energy $\mathcal{E}_K$, and we shall say that the atom is in the K state. If, however, an electron from the L shell is removed from the neutral atom, than a smaller amount of energy will be required; we shall call this amount of energy $\mathcal{E}_L$ and say that the atom is now in the L state. An atom in the M state is one in which an electron is removed from the M shell, for which $n = 3$, and so on. These energy states are plotted in the diagram.

Suppose that an atom has been raised to the K state by the removal of an electron from orbit $n = 1$. If an electron goes from orbit $n = 2$ to orbit $n = 1$, the atom will be left in the L state, and the frequency of the emitted radiation will be

$$f = \frac{\mathcal{E}_K - \mathcal{E}_L}{h}. \tag{43-4}$$

In the diagram this is called the K_α line. Similarly, the K_β line is emitted in the transition from the M shell to the K shell, and so on.

The continuous x-ray spectrum provides us with another important verification of the quantum character of electromagnetic radiation. In Figure 43-12 we note that the continuous spectrum exhibits a minimum

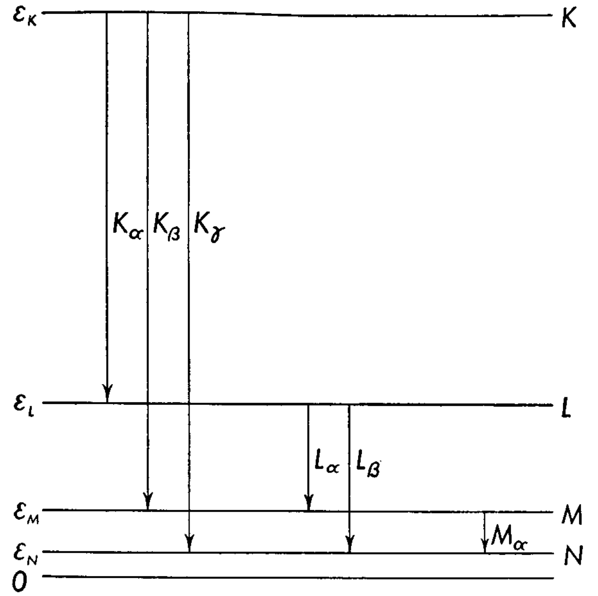


Fig. 43-14 Simplified x-ray energy-level diagram.

wavelength, λ_{min}. When the voltage across the tube is increased, the short-wavelength limit is shifted toward smaller values. The continuous spectrum results from the acceleration of the electrons when they strike the target. The energy may be radiated in a single quantum or in several quanta, but no single quantum can be emitted which has more energy than the kinetic energy of the incident electron. Thus we have

$$Ve = hf_{\text{max}} = \frac{hc}{\lambda_{\text{min}}}, \tag{43-5}$$

where V is the voltage across the x-ray tube, e is the electronic charge, and h is Planck's constant. Ve represents the energy with which an electron strikes the target. Duane and Hunt carried out a series of careful experiments on the determination of the short-wavelength limit of the continuous x-ray spectrum for various voltages across the x-ray tube and found the

value of Planck's constant to be

$$h = 6.556 \times 10^{27} \text{ erg sec},$$

in good agreement with other determinations made optically.

The information obtained by the study of optical and x-ray spectra has played an important part in establishing our present view of atomic structure, particularly the arrangement and distribution of electrons of the atom. Thus the energy-level diagram may be verified in the absorption of x-rays by means of the photoelectric effect. The interaction of x-rays with atoms is associated with the filled inner shells, while the interaction of visible light with matter is associated with the outermost electrons. The energy of the emitted photoelectron may be determined by means of a magnetic spectrometer. When the photoelectron is emitted from the K shell, we find that

$$\mathcal{E}_e = hf - \mathcal{E}_K,$$

where f is the frequency of the incident x-rays, $\mathcal{E}_K$ is the K level energy, and $\mathcal{E}_e$ is the kinetic energy of the emitted photoelectron.

43-5 Particles and Waves

As we have seen, radiation has a dual character, that of a wave motion and that of a corpuscular nature. The wave character of radiation has been amply verified by the phenomena of interference, diffraction, and polarization which we have outlined for x-rays and for visible light.

The corpuscular character of radiation was first introduced by Planck to explain the distribution of energy in the continuous spectrum of a black body at a high temperature. This idea was used by Einstein to explain the photoelectric effect and by Bohr to explain the spectrum of hydrogen, and was found applicable to the spectra of higher-atomic-number elements and to x-ray spectra. Another very convincing argument concerning the corpuscular nature of electromagnetic radiation is provided by the *Compton effect.*

From classical electromagnetic theory it is known that a beam of radiation has momentum as well as energy. As we have seen in our discussion of momentum and impulse, force is equal to the rate of change of momentum. When a light beam is incident upon a black plate and is absorbed by it, the light beam not only delivers energy to the plate but delivers momentum as well, thereby changing the momentum of the plate. The light beam thus exerts a force on the plate. This was verified experimentally by Nichols and Hull in 1903, who measured the force exerted by light from an arc lamp on a blackened disk suspended from a quartz fiber. The force per unit area, called the radiation pressure, measured by Nichols

and Hull was about 7×10^{-5} dyne/cm^2; this value was within 2 per cent of that calculated from electromagnetic theory.

If the concept of momentum applies to a light beam, it must apply to the quanta as well. It will be recalled that a quantity of energy $\mathcal{E}$ has a mass m given by

$$\mathcal{E} = mc^2, \tag{7-12b}$$

where c is the speed of light. Now the energy of a quantum of light, or a photon, is given by

$$\mathcal{E} = hf,$$

hence the mass of a photon is

$$m = \frac{hf}{c^2}. \tag{43-6}$$

Since momentum is the product of the mass of a particle by its velocity, the momentum p of a photon is

$$p = mc,$$

so that

$$p = \frac{hf}{c}. \tag{43-7}$$

Let us consider the scattering of a beam of x-rays by matter. From the point of view of electromagnetic waves, we would imagine that the electric vector of the electromagnetic wave caused the charged particles of the substance to vibrate, and that these particles would reradiate energy of the same frequency as the incident beam. The scattering would be due to the electrons of the substance, for their acceleration would be greater than the acceleration of the nucleus under the application of a given force. On examination of the wavelength of the scattered beam with an x-ray spectrometer, A. H. Compton found that the scattered beam contained not only x-rays of the original frequency f and wavelength λ but also radiation of a lower frequency f' and thus of longer wavelength λ'. The explanation of this effect, given by Compton, is that the x-ray photon makes an elastic collision with an electron and is scattered by it, and, if the electron is free to move, the photon gives up some of its energy to the electron. Using the principle of conservation of energy, we can write

$$hf = hf' + \tfrac{1}{2}mv^2, \tag{43-8}$$

where hf is the energy of the incident photon, hf' the energy of the scattered photon, and $\frac{1}{2}mv^2$ the kinetic energy of the scattered electron. Since this is a collision between two particles, we can apply the principle of conservation of momentum to this process. Figure 43-15 shows the incident photon with momentum hf/c, the photon scattered through an angle θ

with momentum hf'/c, and the electron recoiling as a result of the collision with momentum mv in a direction making an angle ϕ with the original direction of the incident radiation. From the principle of conservation of momentum, we get the equations

$$\frac{hf}{c} = \frac{hf'}{c}\cos\theta + mv\cos\phi \tag{43-9a}$$

and

$$0 = \frac{hf'}{c}\sin\theta - mv\sin\phi. \tag{43-9b}$$

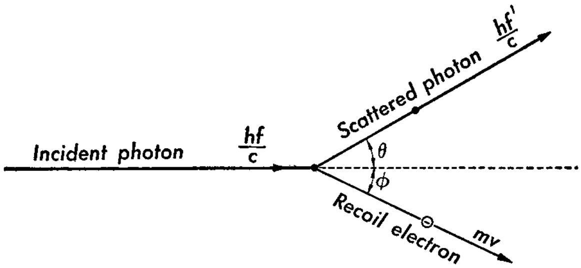


Fig. 43-15 The Compton effect; scattering of a photon and recoil of an electron as a result of a collision.

Equation (43-9a) states that the initial momentum along the original direction of the x-ray beam is equal to the components in this direction of the final momenta of the scattered photon and the recoil electron. Equation (43-9b) is a similar expression for the components of the momenta at right angles to the original direction of motion.

If the above three equations are solved for the frequency of the scattered photon, and the result is then converted to their corresponding wavelengths, the result obtained is

$$\boxed{\lambda' - \lambda = \frac{h}{mc}(1 - \cos\theta),} \tag{43-10}$$

where λ' is the wavelength of the ray scattered at an angle θ, and λ is the original wavelength. The change in wavelength depends only upon the angle of scattering and not upon the substance. The results predicted by Equation (43-10) have been amply verified by many careful experiments. The energy and momentum of the recoil electron have also been measured, and the results have been found in agreement with the predictions made from the solutions of the above equations. Once again, radiation was found to have a corpuscular character in its interaction with matter.

As has been previously stated (Section 42-6), Louis de Broglie, using the dual character of radiation as a guide, put forth the hypothesis that there should be a wave motion associated with every material particle. If m is the mass of a particle and v its speed, its momentum is $p = mv$. The wavelength λ associated with this particle should be given by the same relationship as that connecting wavelength and momentum of electromagnetic waves

$$\lambda = h/p,$$

so that

$$\lambda = h/mv. \tag{43-11}$$

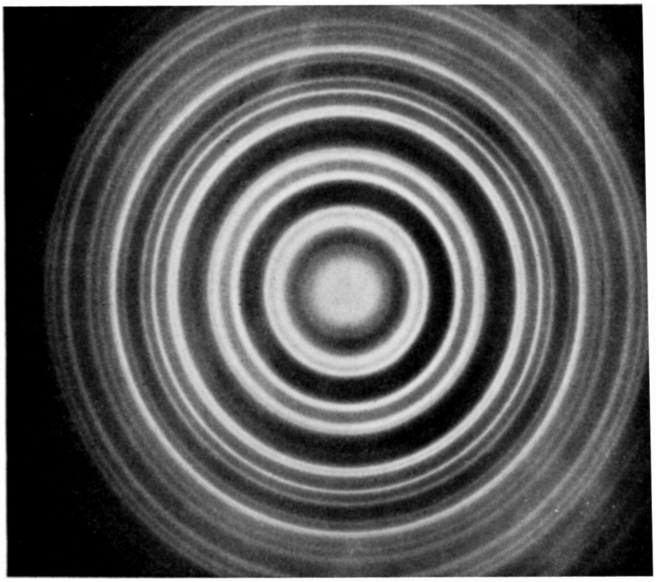


Fig. 43-16 Diffraction pattern obtained by passing a beam of electrons through gold foil; thickness of foil was about 250 A. (Reproduced from photograph by Oliver Row and N. R. Mukherjee.)

De Broglie's hypothesis was verified in two classic experiments—one performed by Davisson and Germer in 1927, and the other performed by G. P. Thomson in 1928. In the Davisson-Germer experiment, a stream of electrons was reflected from a nickel crystal. Intense maxima were observed at certain angles of reflection, which could be explained only in terms of constructive interference of electron waves from different atomic layers of the nickel crystal. The Bragg equation together with the de Broglie equation provide an explanation of the reflection maxima. Electrons, formerly thought to be wholly particlelike in character, were found to display wave properties.

In the Thomson experiment, a stream of electrons of known velocity was sent through a very thin metal foil. Since a metal foil consists of many microcrystals oriented at random, the pattern obtained is similar to that obtained in the powdered crystal x-ray diffraction experiments. From the

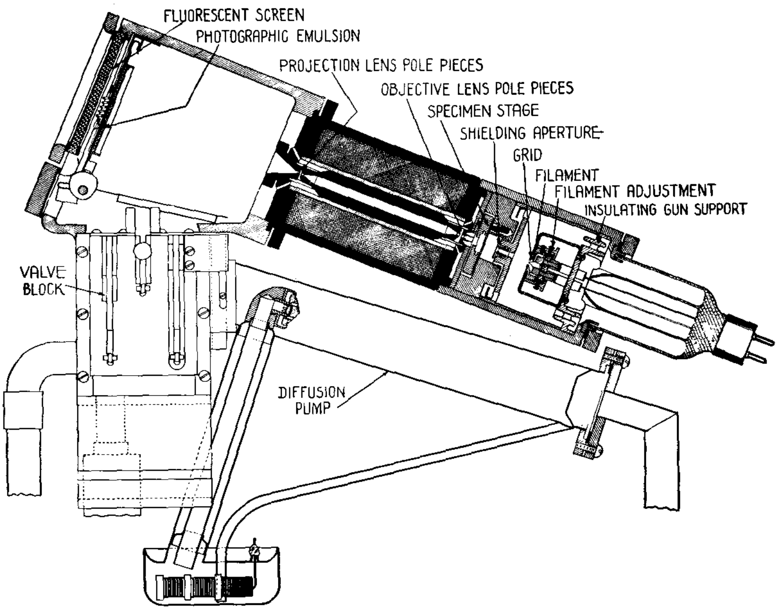


Fig. 43-17 A simplified cross section of a small electron-microscope unit. (Courtesy of RCA Laboratories.)

results of these experiments, G. P. Thomson was able to compute the grating space of metallic crystals in the thin foil, and the results agreed very well with those obtained by x-ray diffraction. The diffraction pattern obtained when a narrow beam of electrons is sent through a thin gold foil is shown in Figure 43-16.

The wave character of material particles is not limited to electrons but is characteristic of all matter. Diffraction patterns have been obtained by reflecting hydrogen and helium molecules from crystal surfaces. The wavelengths computed from the results of these experiments agreed with those calculated from Equation (43-11), where the velocity v of the molecules was determined from a knowledge of the temperature of the gas.

One of the practical results of the de Broglie hypothesis was the development of the *electron microscope*. We recall that the resolving power of a microscope is limited by the wavelength of the illuminating radiation. By using electrons of appropriate speed, the wavelength of these electron waves may be made any desired value. The electron beam may illuminate

a specimen by means of properly arranged electric and magnetic fields which form *electron lenses* and may then be focused onto a fluorescent screen or a photographic plate to produce an image of the specimen. Such an electron microscope is shown in Figure 43-17, while pictures of a specimen taken optically and with an electron microscope are shown in Figures 43-18 and 43-19. In each of the two pictures the actual resolving power is about a factor of 3 from the theoretical limit of resolution of the microscopes used. The highest resolution obtained with an electron microscope is about 20 A.

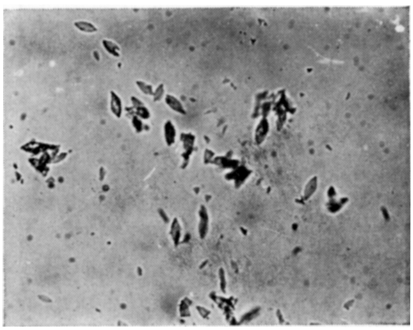


Fig. 43-18 Photograph with an optical microscope of crystals of monohydrated aluminum oxide. Magnification 500×. (Courtesy of J. Hillier, RCA Laboratories.)

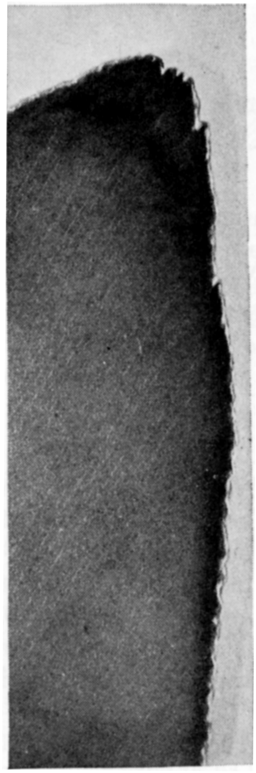


Fig. 43-19 Electron-microscope photograph of the tip of one of the crystals of Figure 43-18. Magnification 90,000×. (Courtesy of J. Hillier, RCA Laboratories.)

Illustrative Example. Find the de Broglie wavelength of an electron accelerated through a potential difference of 100 volts.

The energy acquired by an electron in passing through a potential difference of 100 volts appears as kinetic energy, so that

$$Ve = \tfrac{1}{2}mv^2.$$

Thus

$$mv = (2Vem)^{1/2}.$$

Now

$$Ve = 100 \text{ volts} \times 1.6 \times 10^{-19} \text{ coul} = 1.6 \times 10^{-17} \text{ joule},$$

or

$$Ve = 1.6 \times 10^{-10} \text{ erg}.$$

The mass of the electron is $m = 9.1 \times 10^{-28}$ gm, and $h = 6.62 \times 10^{-27}$ erg sec.

Thus
$$\lambda = \frac{h}{mv} = \frac{6.62 \times 10^{-27}}{(2 \times 1.6 \times 10^{-10} \times 9.1 \times 10^{-28})^{1/2}},$$

$$\lambda = 1.23 \times 10^{-8}\ \text{cm} = 1.23\ \text{A}.$$

This wavelength is of the order of atomic dimensions.

43-6 Heisenberg's Uncertainty Principle

An interesting interpretation of the wave-particle duality of both matter and radiant energy has been given by Heisenberg. Heisenberg's *uncertainty principle* refers to the *simultaneous* determination of the position and momentum of a particle and states that *the uncertainty* Δx *involved in the measurement of the coordinate of a particle and the uncertainty* Δp_x *involved in the simultaneous determination of its corresponding momentum are governed by the relationship*

$$(\Delta x)\ (\Delta p_x) \geqq h, \tag{43-12}$$

where h is Planck's constant.

Let us examine an idealized experiment to see how the wave concept acts as a limitation on the particle concept, giving rise to the uncertainty principle. One such idealized experiment was given by Bohr. Suppose it is desired to determine the position of an electron, using some instrument such as a microscope of very high resolving power. It may be shown that the resolving power of a microscope is given by

$$\Delta x = \frac{\lambda}{\sin \alpha}, \tag{43-13}$$

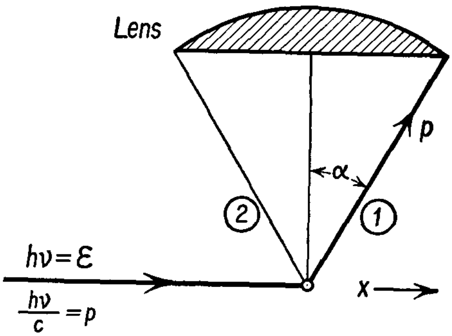
where Δx represents the distance between two points which can just be resolved by the microscope, λ is the wavelength of the light used, and α is the semivertical angle of the cone of light coming from the illuminated object into the microscope objective. The uncertainty in the determination of the x coordinate of the electron is represented by Δx. To make Δx as small as possible, we must use light of very short wavelength, or x-rays or gamma rays. The minimum amount of light that can be used is a single quantum hf. When the electron scatters this quantum into the microscope, as shown in Figure 43-20, the electron will receive some momentum from the quantum, for the process by which the radiation is scattered is the Compton effect. Since the scattered quantum can enter the microscope anywhere within the semivertical angle α, the x component of the momentum which the electron acquires is uncertain by an amount

$$\Delta p_x = p \sin \alpha = \frac{h}{\lambda} \sin \alpha,$$

where h/λ is the momentum of the quantum. The product of the uncertainties in the determination of the simultaneous values of position and momentum is therefore

$$\Delta x \, \Delta p_x = \frac{\lambda}{\sin \alpha} \times \frac{h}{\lambda} \sin \alpha = h.$$

Fig. 43-20 Schematic diagram of a gamma-ray microscope. If we neglect the change in wavelength of the scattered gamma ray, we note that the electron recoils with a momentum whose x component ranges between the limits $\pm p \sin \alpha$ about some central value, depending on whether the gamma ray entering the microscope takes path 1, path 2, or some intermediate path.



The uncertainty principle has had some striking implications in physics itself and in the philosophical extensions of physical concepts. In Newtonian mechanics the trajectory of a particle was completely determined if the initial position and momentum of the particle were known initially with any desired precision, and the forces acting on the particle could also be specified. Since the forces acting on a particle are generated by other particles through gravitation, through the electric and magnetic fields, and so on, a knowledge of the position and momentum of each particle in the universe at any one time made it possible to predict, at least in principle the position of each particle at every other time. Since each of us is composed of elementary particles, this state of affairs implied that our lives were completely predetermined. The uncertainty principle has taken away the physical foundation which was assumed to lie at the basis of this philosophy of determinism, for the wave-particle duality of both particles and waves implies that it is impossible to specify both the position and momentum of an elementary particle with sufficient precision to predict its future with certainty.

The wave-particle duality has been clarified by the principle of *complementarity* expressed by Bohr. According to Bohr, the wave and particle aspects of a phenomenon are never exhibited simultaneously. For some phenomena an electron may be interpreted as a wave, while for others it may be interpreted as a particle. The particle does not alter its characteristics, but rather, these two aspects of the electron are complementary rather than contradictory descriptions, for they represent limiting cases of a more complete description which may not yet have been achieved. In

such a description the electron is neither wave nor particle but is described mathematically by an equation or function which under some conditions is more nearly approximated by a wave, while under other conditions may be discussed as though it were a particle. We have seen evidence of this sort of thing in optics. In the phenomena of reflection and refraction, we may achieve a sufficient description of light by treating infinitely narrow rays, and as long as we are interested in the magnifications produced by lenses and mirrors and optical instruments, we may overlook the wave nature of light. In interference and diffraction phenomena we find it essential to treat light as a wave motion, and we must assert that it is impossible to restrict a light beam so that it is infinitely narrow.

Problems

43-1. A single crystal x-ray spectrometer with a quartz crystal ($d = 4.255$ A) is used to analyze the x-rays from a molybdenum target. The wavelength of the K_α line of molybdenum is 0.7078 A. (a) Determine the angle between the incident beam and the face of the crystal so that this line may be reflected in the first order. (b) Determine the angle between the incident beam and the reflected beam. (c) The atomic number of molybdenum is 42. As calculated from Equation (43-3), what is the percentage error in the calculated wavelength as compared to the experimental wavelength of this K_α line?

43-2. (a) What is the frequency of the K_α line of tungsten for which $Z = 74$? (b) What is the energy required to move a K shell electron just out of the atom? This may be inferred from Equation (43-3). (c) What voltage must be applied to a tungsten-target x-ray tube to excite the K series lines of tungsten?

43-3. A single crystal x-ray spectrometer with a calcite crystal is used to analyze the x-rays from a tube with a silver target. A very intense line is obtained when the beam makes an angle of $5°17'$ with the crystal. Determine the wavelength of this line. The grating space of calcite is 3.0294 A.

43-4. The K energy level of silver is 25,500 ev. X-rays of wavelength 0.40 A incident upon a silver foil eject electrons from the K levels of some of the silver atoms. (a) Determine the energy of the incident photons in electron volts. (b) Determine the kinetic energy of the ejected electrons in electron volts.

43-5. Assuming that Equation (43-3) may be extended to all the lines of the K series by proper adjustment of the quantum numbers, find the wavelengths of the K series lines of silver ($Z = 47$). Consult Table 5, Appendix A, to determine how many lines there are in this series.

43-6. Determine the energy, in electron volts, of the K_α line of sodium.

43-7. A high-voltage vacuum-tube rectifier with a tungsten filament and a nickel plate is used to rectify alternating voltages with a peak value of 100,000 volts. (a) Will x-rays be emitted from this tube? (b) If so, what will be the shortest wavelength generated by the tube? (c) What is the excitation energy of the K state of nickel? Assume that you may infer this value by appropriate adjustment of Equation (43-3). (d) Will the K_α line of nickel be emitted from

either the filament or the plate of this diode? (e) If a continuous spectrum of x-rays is emitted from the tube, will it be emitted from the filament or the plate?

43-8. Determine the change in wavelength produced when an x-ray photon is scattered by a substance (a) through an angle of 90° and (b) through an angle of 180°.

43-9. X-rays of wavelength 0.500 A are incident upon a block of carbon. (a) Determine the energy of the photons of the incident beam. (b) Determine the change in energy of the photons which are scattered through an angle of 90°. (c) Assuming that this scattering is due to free electrons which originally had negligible energy and momentum, determine the energy of the recoil electrons.

43-10. Referring to Problem 43-9, determine (a) the momenta of the incident photons, (b) the momenta of the photons scattered through 90°, and (c) the momenta of the recoil electrons.

43-11. (a) Determine the wavelength associated with an electron whose energy is 1,600 ev. (b) Assuming that Bragg's law holds for the waves associated with electrons, determine the angle at which a stream of electrons should be directed toward the surface of a rock salt crystal to obtain maximum reflection, if the energy of these electrons is 1,600 ev.

43-12. (a) An electron is moving with a velocity of 2×10^8 cm/sec. Determine its momentum and wavelength. (b) A proton is moving with a velocity of 2×10^8 cm/sec. Determine its momentum and wavelength.

43-13. A stream of helium gas consists of helium atoms moving with an average speed of 3×10^5 cm/sec. Determine (a) the average momentum of a helium atom and (b) the wavelength associated with such an atom.

43-14. When we know that a particle is enclosed within a cubical container of side s, we may say that its position is given by the coordinates of the center of the box, with an uncertainty in each coordinate of $s/2$. Let us suppose that an electron is imprisoned in a box whose walls are equivalent in strength to a potential of -100 volts, the inside of the box being at ground potential or at zero potential. How small can the box be so that the electron will remain trapped in the box? We may assume that the electron will leave the box when the uncertainty in its energy is equal to the potential barrier represented by the walls.

43-15. A particle rests on the level surface of a yard provided with a high wall. (a) From a classical Newtonian viewpoint, how high does the wall have to to be to keep the particle inside? (b) If the yard is 20 ft wide, and a ball has a mass of 1 gm, how high must the wall be, according to the uncertainty principle? (c) If the "yard" is 10^{-6} cm wide and the particle has a mass of 10^{-10} gm, how high must the wall be? (d) If the "yard" is 10^{-13} cm wide and the particle has a mass of 10^{-27} gm, how high must the wall be? Assume that the "wall" is built in a uniform gravitational field in which $g = 980$ cm/sec^2.

44

Stable Nuclei

44-1 Atoms and Nuclei

Atomic and nuclear physics are essentially twentieth-century developments, although these had their origins at the close of the nineteenth century. These developments followed two parallel streams, as shown in Figure 44-1, one dealing with the electronic structure of the atom and the other dealing with its nuclear structure. These two streams served to feed each other and were fed by all other branches of physics, but only a few of the more important contributions concerning the nature of radiation, matter, and energy are shown in the figure.

Before proceeding with a discussion of nuclear properties, it will be worth while to recapitulate some of the phenomena which have been discussed or inferred previously, to emphasize them for the present discussion. There are at present 102 known elements, of which the elements of atomic number above 93 are not found in nature but have been synthetically produced in the laboratory and in the atomic bomb. Each element is characterized by two numbers (a) its *atomic number* and (b) its *atomic weight.* The atomic numbers run consecutively from 1 to 102, at present, and the atomic weights range from 1 to more than 250. We recall that the atomic number was initially simply a serial number which was assigned when the elements were arranged in order of their atomic weights and chemical properties in the periodic table. The work of Moseley on the characteristic x-ray spectra demonstrated that the atomic number Z was a property of fundamental significance. We now know that the atomic number is the number of protons in the nucleus of the atom. At one time in the history of chemistry, the discovery of a new element was extremely important, and the possibility always existed that a large number of elements might be found in one place in the periodic table, as in the case of the rare earths. This possibility no longer exists. There are two scales of atomic weight in general use. The chemical system of atomic weights is based upon the assignment of atomic weight 16 to oxygen, in the isotope

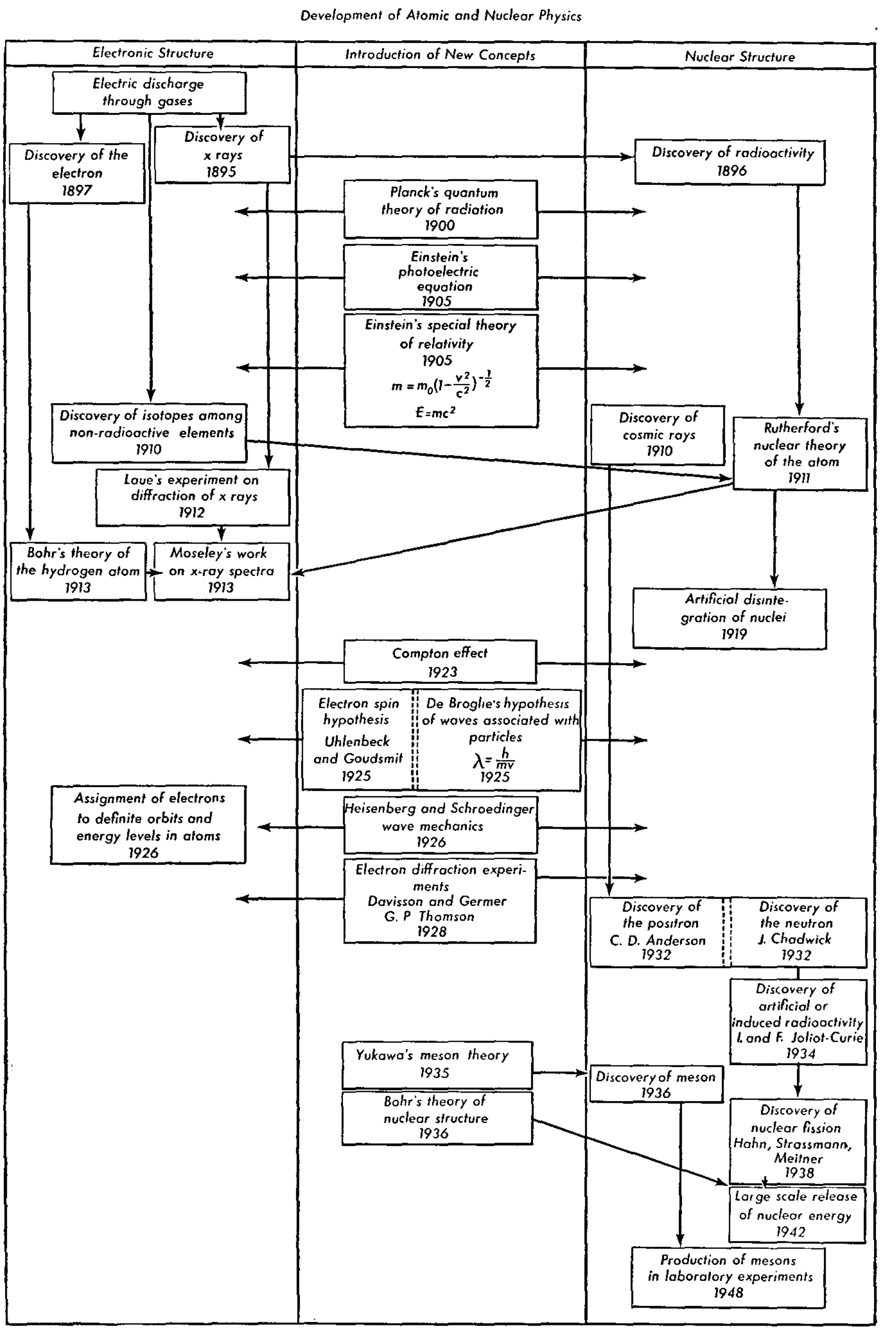


Fig. 44-1 Development of atomic and nuclear physics.

distribution which is found on the earth, while the physical scale of atomic weight is based on the assignment of atomic weight 16 to the most abundant isotope of oxygen. The difference between these scales is quite small; the ratio of the physical to the chemical scale of atomic weights being 1.000275. This difference is significant in physical measurements.

We shall take for granted that each atom consists of a positively charged nucleus of very small diameter, surrounded by an appropriate number of electrons so that the atom is electrically neutral in its normal state. The nucleus is composed of neutrons and protons, where the neutron is an electrically neutral particle of atomic weight 1.008987, and the proton is a positively charged particle of atomic weight 1.007595, whose charge is equal in magnitude to the electronic charge of $e = 4.802 \times 10^{-10}$ stcoul $= 1.602 \times 10^{-19}$ coul.

44-2 Equivalence of Mass and Energy

Until the advent of Einstein's theory of special relativity, one of the fundamental principles of physics and chemistry was the principle of the conservation of mass, which stated that the mass of an isolated system remained constant under all changes of physical states and chemical composition. Einstein recognized that mass was a form of energy, and that the principle of conservation of energy had to be enlarged to include mass. It will be recalled that the principle of conservation of energy was first formulated about 1847, when it was shown definitely that heat is a form of energy. Previous to this period there was a conservation theorem which concerned only the mechanical forms of energy, kinetic and potential energy, for mechanical transformations which occurred in frictionless systems. After the convincing work of Joule on the mechanical equivalent of heat, the concept of energy was extended to include heat. With the formulation of Maxwell's electromagnetic theory of light and the discovery of various forms of electromagnetic radiation, the principle of conservation of energy was readily extended to include electromagnetic radiation among the forms of energy. Now mass is included as one aspect of energy, along with other forms of energy.

Einstein showed that if a mass m is measured in grams or kilograms, the energy $\mathcal{E}$ of this mass can be expressed in ergs or joules by the relationship

$$\boxed{\mathcal{E} = mc^2,} \tag{44-1}$$

where c is the speed of light and is equal to 3×10^{10} cm/sec for conversion from grams to ergs, and is 3×10^8 m/sec for conversion from kilograms to joules.

Whenever a substance gains energy, as when it is heated, the substance gains in mass. Whenever a substance loses energy, as in an exothermal chemical reaction, the substance loses mass. In general, the mass change associated with ordinary chemical processes is undetectable; the energy liberated when 1 gm is converted into other forms of energy is

$$\mathcal{E} = 1 \text{ gm} \times (3 \times 10^{10})^2 = 9 \times 10^{20} \text{ ergs} = 2.25 \times 10^{13} \text{ cal},$$

while chemical processes involve energy changes of the order of 10^6 cal/mole. Thus the mass changes in such processes are of the order of 10^{-7} gm in a

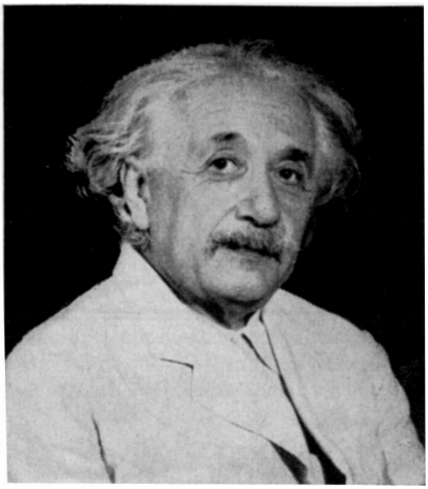


Fig. 44-2 Albert Einstein. He developed the theory of relativity and revolutionized the mode of thinking about fundamental physical problems. One consequence of this theory was the extension of the concept of energy to include mass as a form of energy. Another part of the theory gives us a new insight into gravitational phenomena. He also developed the fundamental equation of the photoelectric effect and the theory of Brownian motion. (Official U.S. Navy Photo from Acme.)

mole of substance, which is beyond the limit of measurement. It is interesting that the mass changes sought by the advocates of the caloric theory of heat, who tried to determine the weight of caloric, are today required by theory and found in experiment in nuclear reactions.

The conversion of mass into other forms of energy is believed to be going on continuously in the sun and other stars. This process is the basis for the construction of nuclear weapons and nuclear reactors. The fundamental mode of conversion of mass into energy is through changes in the nuclear constitution of atoms. An important clue to this process is the precise measurement of the masses of atoms and a comparison of these values with the masses of the constituent particles.

44-3 Positive Ions. Mass Spectrograph

An instrument designed to measure the mass of an ion is called a *mass spectrometer* or a *mass spectrograph*, depending on whether an electrical method or a photographic plate is used to record the ions. Modern mass spectrometers are instruments of very high precision which are capable of

determining atomic masses very accurately and also of determining the number and relative abundance of the *isotopes* of which the elements are composed.

The term *isotopes* was introduced by Soddy as a result of the study of the radioactive elements. It was found that several groups of elements having identical chemical properties but different atomic weights were formed in the process of radioactive disintegration. The term "isotopes" was used to designate elements which occupied the same place in the periodic table. It was subsequently found that these elements had the same atomic number but different atomic weights. Because some elements had atomic weights which differed considerably from whole numbers, it was suggested that these elements consisted of two or more different isotopes having different atomic weights. The search for isotopes among the stable elements was begun by J. J. Thomson in 1910, and the first element successfully investigated was neon, whose atomic weight 20.2 differs appreciably from a whole number. By sending the positive ions formed in a gas-discharge tube through electric and magnetic fields, Thomson determined the ratio of the charge to the mass of these ions and found that neon consists of at least two isotopes of atomic masses very close to 20 and 22. Many variations of the original method were made by later investigators to improve the accuracy of this method.

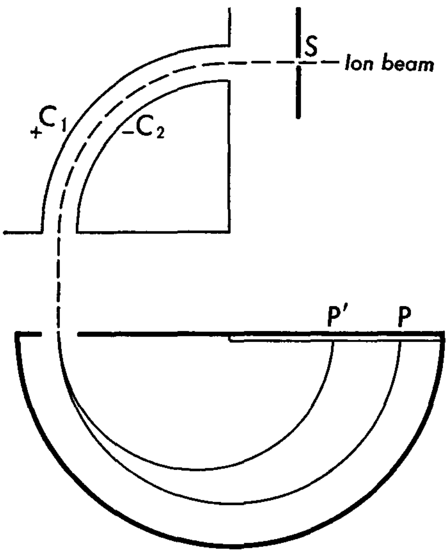


Fig. 44-3 Dempster's mass spectrograph which uses a capacitor with cylindrical plates.

In order to avoid confusion, let us introduce two new terms: (a) the *atomic mass* refers to the mass of an isotope of an element, based upon the physical scale in which the oxygen isotope of atomic mass 16.0000 is taken as the standard; this is the lightest of the three isotopes found in ordinary oxygen; (b) the *mass number* of an isotope of an element refers to the whole number which is nearest to the atomic mass of the isotope.

There are many varieties of mass spectrographs and spectrometers in use in research and industrial laboratories. The essential parts of a mass spectrograph designed by A. J. Dempster are sketched in Figure 44-3. These parts are enclosed in a vacuum chamber. Positive ions from a

convenient source pass through the narrow slit S into a radial electric field between two cylindrical plates C_1 and C_2 of a capacitor. The ions are deflected from C_1 toward C_2. Only those ions with appropriate velocity traverse the circular path between the capacitor plates. If E is the radial component of the electric field in this region, and R is the radius of the path, we require that

$$Ee = \frac{mv^2}{R},$$

where e and m are the charge and mass of an ion. The ion beam may originate in an electric discharge in a gas and may contain ions of many different velocities. The capacitor serves as a velocity selector. A short

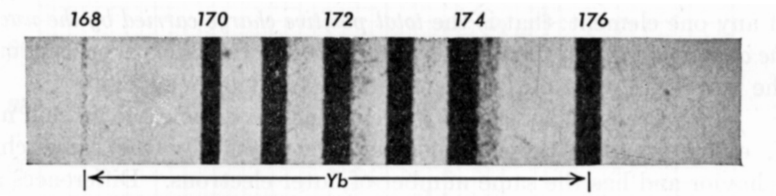


Fig. 44-4 A photograph of the isotopes of ytterbium obtained with Dempster's mass spectrograph. The mass numbers of the isotopes can be obtained from the number scale printed above the lines. (Reprinted from a photograph supplied to the author by Prof. A. J. Dempster.)

distance beyond the electric field the ions enter a uniform magnetic field which is at right angles to the plane of the paper, and, after traversing a semicircular path, strike the photographic plate PP' on which they are recorded. The radius r of the semicircular path is given by

$$\frac{mv^2}{r} = Bev,$$

where B is the magnetic flux density. From these equations we find that

$$v = \frac{ER}{Br},$$

and

$$\frac{e}{m} = \frac{ER}{(Br)^2}, \tag{44-2}$$

where all quantities are expressed in mks units.

A typical spectrogram obtained with this apparatus is illustrated in Figure 44-4, which shows the isotopes of the rare-earth element ytterbium. The mass number of each isotope is shown above the line formed by its ions on the photographic plate. Mass spectrometers are widely used in the petroleum industry as a means of following the refining and cracking proc-

esses. Here the molecular weight of large organic molecules is an important factor in the suitability of a fuel for a particular use, and this may be determined most easily with a mass spectrometer.

44-4 Masses of Isotopes and Nuclear Structure

Investigations with the mass spectrograph have established that there are about 300 different stable isotopes among the 102 known elements. The range of mass numbers runs from 1 to more than 250. The atomic masses of these isotopes differ very little from whole numbers. The number of stable isotopes per element varies from 1 for elements fluorine and gold to 10 for element tin. There is one thing which is common to all the isotopes of any one element: that is the *total positive charge carried by the nucleus of the atom*. Hence, in the neutral atom, the number of electrons surrounding the nucleus is the same for each isotope of any one element.

Since the isotopes of any one element have the same atomic number Z, every atom of the element displays essentially the same chemical behavior and has the same number of outer electrons. Differences among the atomic masses of an element must therefore be due to differences in their nuclei; the fact that the atomic masses of all isotopes are nearly integers suggests that nuclei are made up of particles (called *nucleons*) of approximately unit atomic mass. At present, two particles of nuclear size and mass nearly equal to unity are known. These are the proton and the neutron. The proton is the positively charged nucleus of the hydrogen atom of mass number 1. The hydrogen atom consists of 1 proton as nucleus and 1 electron outside the nucleus. Since the mass of the hydrogen atom is about 1,840 times the mass of the electron, practically the entire mass of the atom is due to the proton. On our scale of atomic units, the atomic mass of hydrogen is 1.00815, the mass of the electron is 0.00055, and the mass of the proton is 1.00760. The total number of neutrons and protons in a nucleus is equal to the mass number A of the isotope of an element. Thus the number of neutrons in the nucleus is $N = A - Z$, and the isotopes of any one element differ only in the number of neutrons in the nuclei of the atoms. Thus helium, the second element in the periodic table, has $Z = 2$. Its most abundant isotope has a mass number $A = 4$; its nucleus has 2 neutrons in addition to 2 protons.

An examination of the known stable isotopes is particularly interesting. Figure 44-5 is a graph of the neutron number N plotted against the proton number Z of the stable isotopes. The region of stability on this neutron-proton diagram is rather narrow. For low mass numbers $N = Z$, while for high mass numbers there are about 1.6 neutrons for each proton. Lines of constant A can be drawn at angles of 135° with the Z axis; such lines pass through *isobars*, that is, nuclei of equal mass. In general, lines of

constant A pass through one or two stable isotopes. There are only four cases of such lines passing through three stable isobars, at $A = 96, 124, 130,$

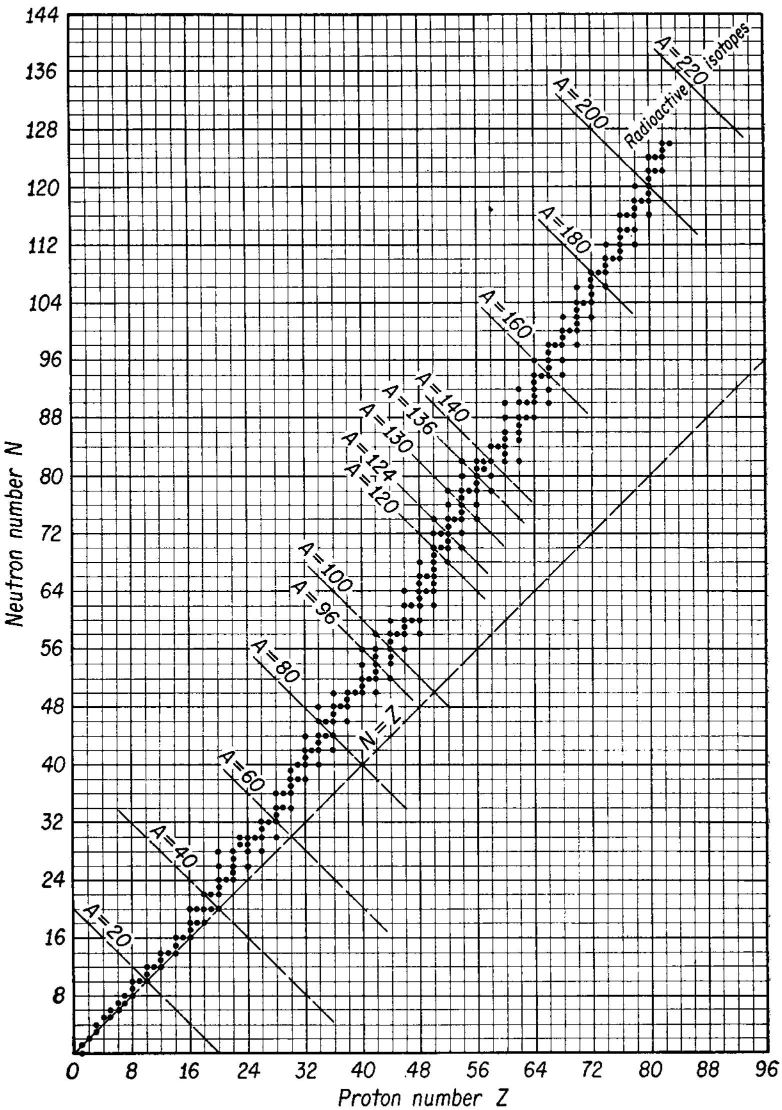


Fig. 44-5 Neutron-proton diagram of stable nuclei.

and 136. Another interesting point is that more than half of the stable nuclei have even numbers of protons and neutrons, and are referred to as even-even nuclei. About 20 per cent have even Z odd N, while about the

same number have odd Z and even N. There are only four stable nuclei which have odd Z and odd N, namely ${}_1H^2$, ${}_3Li^6$, ${}_5B^{10}$, and ${}_7N^{14}$, where the number appearing as a left-hand subscript represents Z, while the right-hand superscript represents A. In the nucleus, nature has a decided preference for even numbers. We may interpret Figure 44-5 by inferring that points to the left of the stability region have too many neutrons, while points to the right of the stability region have too few neutrons, and that such isotopes will suffer radioactive disintegration.

The particles which constitute a nucleus exert forces on each other. If the only force between nuclear particles were the Coulomb repulsive force between particles of like charge, we would expect to find no nucleus beyond hydrogen. There must be another type of force between nucleons. For want of a better name, we shall call this a *nuclear force.* The exact nature of nuclear forces is not well known and is being extensively investigated by means of high-energy particle accelerators. We may think of the nuclear force as the glue which holds the nucleus together. This glue can be furnished a nucleus without the addition of an electrical repulsive force by adding a neutron rather than a proton.

The nuclear force acts through very short distances, for experiment shows that the nucleus is a compact structure, occupying a volume which is essentially the volume of its A nucleons. The range of the electrical repulsive force between protons is much greater than the range of the nuclear force. Thus we find that the more massive nuclei which have large values of Z must have more neutrons per proton than the lighter nuclei.

It is to be expected that the ideas and concepts that proved so effective in determining the electronic structure of atoms should be carried over into nuclear physics. One of these ideas is that of *shell structure* or *level structure,* with certain shells closed owing to the stability of the system with the given number of particles, much as the stability of the rare gases is determined by the closure of atomic shells. The neutron and proton both have spin quantum numbers of $\frac{1}{2}$, like the electron, so that 2 neutrons or 2 protons can exist in a given orbital energy state. Evidence first advanced by Maria G. Mayer in 1948 showed that nuclei having 20, 50, or 82 protons or 20, 50, 82, or 126 neutrons were particularly stable. The evidence used included the number of stable isotopes in a given category. For example, there are seven stable isotopes with $N = 82$, while there is only one for $N = 81$ and one for $N = 83$. Tin, for which $Z = 50$, has the largest number of stable isotopes, namely 10, of any element. These numbers, sometimes called *magic numbers,* now have some theoretical foundation, based upon the filling of nuclear energy levels in a postulated nuclear force field.

44-5 Binding Energy of Atoms and Nuclei

Let us suppose that a neutral atom of mass number A and atomic number Z is formed by bringing together A nucleons (Z protons plus N neutrons) in the nucleus and Z electrons outside the nucleus. Since this atom is a stable structure, its total energy must be less than that of a system consisting of these same particles separated by such large distances that the effect of the nuclear forces and the electrical forces holding an atom together are negligible. The difference between the total energy of the separated particles and the total energy of the neutral atom composed of these particles is the *binding energy of the atom.* From the principle of equivalence of mass and energy, it can be concluded that the decrease in energy produced in the formation of the atom should be evidenced by a decrease in the mass of the system. The mass of the atom should be less than the sum of the masses of its constituent particles. This was first brought to light as a result of the very precise determinations of the masses of isotopes by means of the mass spectrometer.

As a simple example, let us consider the formation of deuterium ${}_1H^2$, whose nucleus consists of 1 neutron and 1 proton. The atomic mass of the deuterium atom is 2.01474 atomic mass units (abbreviated amu). The sum of the masses of the constituent particles, in atomic mass units, is

$$\begin{aligned} \text{proton mass} &= 1.00760 \text{ amu},\\ \text{neutron mass} &= 1.00899 \text{ amu},\\ \text{electron mass} &= \underline{0.00055} \text{ amu},\\ \text{total} \quad & 2.01714 \text{ amu}. \end{aligned}$$

Thus the mass of a deuterium atom is less than the sum of the masses of the constituent particles by 0.00240 amu.

To appreciate the meaning of these numbers, let us convert the atomic mass unit into more commonly used units. Now 1 amu is one sixteenth of the mass of an oxygen atom of mass number 16 and atomic mass 16. Since there are N_0 atoms in 16 gm of oxygen, 1 atom has a mass of 16 gm/N_0, so that 1 amu, which is one sixteenth of this, is simply

$$1 \text{ amu} = \frac{1 \text{ gm}}{N_0} = \frac{1 \text{ gm}}{6.025 \times 10^{23}},$$

from which $$1 \text{ amu} = 1.66 \times 10^{-24} \text{ gm}.$$

Using the relationship between mass and energy given in Equation (44-1), we find that

$$1 \text{ amu} = 1.49 \times 10^{-3} \text{ erg}.$$

Another convenient unit of energy is the electron volt where

$$1 \text{ ev} = 1.60 \times 10^{-12} \text{ erg}.$$

Thus

$$1 \text{ amu} = 931.2 \times 10^{6} \text{ ev} = 931.2 \text{ Mev},$$

where Mev stands for million electron volts.

The binding energy of the deuterium atom which has been determined to be 0.00240 amu is therefore equal to 2.23 Mev. In order to separate deuterium into its constituent particles, an amount of energy equal to 2.23 Mev must be supplied. This may be done by irradiating deuterium with gamma rays. When these rays have energy greater than 2.23 Mev, it is observed that deuterium is decomposed into its components, that is, into 1 neutron and 1 proton. The binding energy of the electron is very nearly equal to the energy required to ionize a hydrogen atom, or 13.6 ev. This energy is negligible compared to the total binding energy of deuterium. Practically the entire binding energy is that of the two nucleons in the deuterium nucleus. This calculation gives us an indication of the relative magnitudes of the energies involved in atomic and nuclear processes. Normal chemical reactions which deal with the outer electrons of an atom are associated with energies of the order of 10 ev/atom. In nuclear reactions the energies involved are of the order of millions of electron volts per nucleon. While the nucleus does not participate in a chemical reaction, it exercises a controlling influence on any chemical process through its determination of the valency of the atom. Conversely, the state of chemical combination exercises little or no influence on nuclear processes.

The binding energy $\mathcal{E}_B$ of a nucleus is given by

$$\mathcal{E}_B = ZM_H + NM_n - M, \tag{44-3}$$

where M_H is the mass of a hydrogen atom, M_n is the mass of the neutron, and M is the mass of the atom as determined in the mass spectrograph. The mass of a hydrogen atom rather than the mass of a proton is used in order to correct for the number of electrons. If the total binding energy of the nucleus is divided by A, the mass number, we obtain the average binding energy per nucleon, which is plotted in Figure 44-6. Here we see that the average binding energy per nucleon is about 8 Mev for elements of mass number 20 or greater. Of course the binding energy of the hydrogen nucleus is zero. We note that the binding energy per nucleon is about 7.6 Mev for uranium, while it is about 8.7 Mev for manganese. When a uranium nucleus is split into several smaller parts, this difference in binding

energy results in the liberation of about 1 Mev/nucleon, or about 200 Mev/nucleus. We see that a given mass of uranium liberates millions of times as much energy through nuclear processes as through chemical processes. Thus nuclear fuels are millions of times more compact than chemical fuels. A greater amount of energy is liberated when protons and neutrons are converted into helium, for here the binding energy is about 7.2 Mev/nucleon. This is the basis of the sun's energy and of thermonuclear processes, such as the hydrogen bomb.

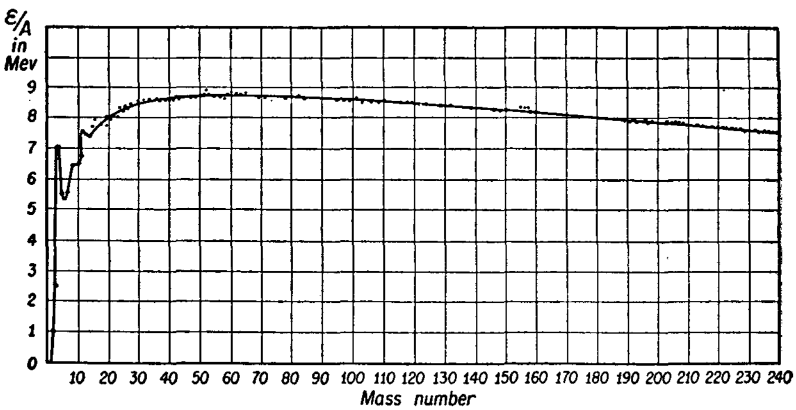


Fig. 44-6 Average binding energy per nucleon as a function of the mass number A.

In Figure 44-6 we note the presence of binding-energy peaks in the light elements at intervals of $\Delta A = 4$. If we imagine that there are energy levels in the nucleus for both neutrons and protons, we would expect these levels to be filled by pairs of neutrons and pairs of protons. According to the patterns established in the atom, a filled level is a more stable structure than an unfilled level. Thus we expect a preference for even numbers of neutrons and protons, and for systems which contain multiples of the helium nucleus. This is indeed the case until we have accumulated sufficient protons in the nucleus for the electrical repulsive force to become a dominating factor. In this connection it is interesting that the isotope $_4Be^8$ is unstable and is not found in nature, for the saturation of nuclear forces is so great in helium that a combination of two helium nuclei is not a stable form. These characteristics of the nucleus and the fact that a relatively constant binding energy per nucleon is observed in the heavier nuclei indicate that each nucleon is bound only to its near neighbors, and that the nuclear force is a short-range force.

The atomic masses of the isotopes are given in Table 3, Appendix A.

Problems

44-1. Calculate the difference between the binding energy of a nucleus of C^{12} and the sum of the binding energies of three alpha particles. (An alpha particle is a He nucleus.)

44-2. Using the data given in Table 3, Appendix A, calculate the binding energy of the last neutron in a $_3Li^7$ nucleus (a) in atomic mass units and (b) in million electron volts. (c) What is the voltage which must be applied across an x-ray tube in order that the x-rays emitted have sufficient energy to decompose this nucleus into $_3Li^6$ + a neutron?

44-3. Singly charged lithium ions of mass numbers 6 and 7, liberated from a heated anode, are accelerated by means of a difference of potential of 400 volts between the anode and the cathode, and then pass through a hole in the cathode into a uniform magnetic field perpendicular to their direction of motion. If the magnetic induction is 800 gausses, determine the radii of the paths of these ions.

44-4. Calculate the mass energy of a proton, in million electron volts.

44-5. Uranium isotopes of mass numbers 235 and 238 are to be separated from a piece of uranium by using a mass spectrometer which will deflect them through 180° into two collectors 4.0 cm apart. If the singly charged ions have energies of 2,000 ev when entering the magnetic field, calculate (a) the magnetic induction necessary to achieve this separation and (b) the radii of the paths of the ions.

44-6. Calculate, in electron volts, the electrostatic potential energy of two protons when their centers are 1.5×10^{-12} cm apart.

44-7. From Figure 44-6 calculate the energy released per atom when a uranium nucleus for which $A = 235$ splits into two fission products whose mass numbers are 72 and 163, respectively.

44-8. Let us suppose that a gram of uranium $A = 238$ and $Z = 92$ splits into fragments, each of mass number 119 and atomic number 46, and that the resulting material is a gas. Assuming that no energy is lost from the system and that the temperature is sufficiently high that each of the 92 electrons of a uranium atom acts as a free particle, calculate the final temperature of the resulting gas.

44-9. (a) How much energy, in calories, is liberated when 1 gm of hydrogen is converted into helium? (b) The heat of combustion of bituminous coal is approximately 10,000 Btu/lb. In this fusion process, how much coal is energetically equivalent to 1 gm of hydrogen?

45

Natural Radioactivity

45-1 The Discovery of Radioactivity: Evidence from the Nucleus

The discovery of an important phenomenon usually leads to other important discoveries. The discovery of x-rays by Roentgen in 1895 led to the discovery of radioactivity by Becquerel in 1896. In the gas type of x-ray tube used by Roentgen, the glass walls of the tube were observed to fluoresce. Becquerel was interested in determining whether there was any relationship between the fluorescence of the glass of an x-ray tube and the phosphorescence of certain salts which were irradiated by ordinary light. One of the salts used by Becquerel was the double sulphate of uranium and potassium. He wrapped a photographic plate in very thick black paper, placed a crystal of the uranium salt on it, and exposed the whole thing to sunlight. Later, on developing the plate, he found the silhouette of the crystals on the negative; he interpreted this as produced by radiations coming from the crystal. He also performed other experiments in which he placed various absorbing materials between the uranium salt and the photographic plate; in each case, upon developing the plate, he found the shadow of the absorbing material imaged on the plate; he interpreted this as being due to the absorption of the radiation by the substance which had been placed between the uranium salt and the photographic plate. The crowning experiment was the one in which he decided not to irradiate the salt with light from an external source but to determine whether the salt was itself the source of these radiations. For these experiments he built a light tight box that would hold a photographic plate at the bottom. In one experiment some uranium salt crystals were placed on the photographic plate; later, on developing this plate, Becquerel obtained the silhouettes of the individual crystals. In another experiment he put a piece of aluminum between the uranium crystals and the photographic plate, and, on developing the plate, he again found the silhouettes of the crystals, but they were of decreased intensity due to the absorption of some of the radiations by the aluminum. From these experiments Becquerel concluded

that the radiations came from the uranium salt, and that the external light had no influence on them.

Becquerel then proceeded with a series of experiments to determine the origin of these radiations. He used a variety of compounds of uranium in these experiments and came to the conclusion that the radiations were emitted by the uranium in the compounds. Some of the compounds used were phosphorescent; others were not. Becquerel also discovered a most important property of these radiations; namely, they could cause the dis-

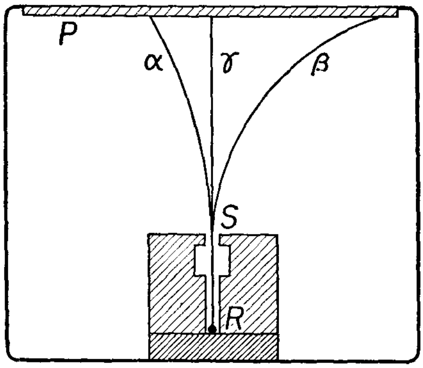


Fig. 45-1 Paths of the rays from a radioactive substance R in a magnetic field. The magnetic field is perpendicular to the plane of the paper and directed into the paper.

charge of electrically charged bodies. This made it possible to investigate the phenomenon quantitatively with the aid of ionization chambers and electroscopes or electrometers.

Using an ionization method, Mme. Curie made quantitative measurements on the activity of uranium salt and demonstrated that this activity was directly proportional to the mass of uranium, showing that radioactivity was an atomic rather than a molecular phenomenon. In 1898 M. and Mme. Curie discovered two new radioactive elements, polonium and radium, in a systematic chemical analysis of uranium pitchblende. The activity of radium was found to be more than a million times that of an equal mass of uranium. Many more radioactive substances have been discovered since then.

Early work by Rutherford on the penetrating power of the radiation from uranium salt showed that the radiation consisted of two components: a very soft radiation easily absorbed in matter, which he called *alpha rays,* and a more penetrating type called *beta rays.* Subsequent work showed that radioactive substances emitted a third type of radiation called *gamma rays.* When a radioactive source is placed in an evacuated chamber provided with a photographic plate, as shown in Figure 45-1, and a strong magnetic field is applied at right angles to the plane of the paper, three distinct lines are found on the photographic plate. The line deflected to

the left in the figure is produced by positively charged alpha rays. The line deflected to the right is produced by negatively charged beta rays. The undeviated line is produced by gamma rays, which are undeflected by the magnetic field. Experiments on interference and diffraction of the gamma rays, similar to those performed with x-rays, show that gamma rays are electromagnetic radiations of very short wavelengths of the order of 0.1 A or less. The alpha particles making up the alpha rays are known to be helium nuclei, and the beta rays are known to consist of electrons. These radiations are spontaneously emitted by radioactive nuclei; that is, radioactivity is a nuclear rather than an atomic process.

45-2 Radioactive Transformations. Half Life

The rate at which a particular radioactive material disintegrates is, with only minor exceptions, independent of physical and chemical conditions. In general, we find that it is impossible to predict which of the atoms in a given collection will disintegrate at a particular time, but that it is possible to make statistical predictions that a certain number of atoms will disintegrate in a given time interval. This implies that each atom has equal probability λ of disintegrating in a unit time, and that the probability that this atom will disintegrate in a time interval dt is given by $\lambda\, dt$. If there are N atoms of a particular species present in a given sample, the total number of atoms we would expect to disintegrate in the time interval dt is the product of the number of atoms present by the probability that each one will disintegrate, or

$$dN = -\lambda N\, dt, \tag{45-1}$$

where dN is the number of disintegrating atoms, or the number whose species is changed, in the time interval dt. Integrating this expression, we find

$$\boxed{N = N_0 e^{-\lambda t},} \tag{45-2}$$

where N_0 represents the number of atoms present at time $t = 0$, N is the number remaining at time t, and λ is a constant for the particular process called the *disintegration constant.* Equation (45-2) shows that the number of atoms of a given radioactive substance decreases exponentially with time. Half of the material will have disintegrated at the end of a time T, called the *half life,* which can be determined by setting $N = N_0/2$, and $t = T$ in Equation (45-2), yielding

$$\lambda T = \log_e 2 = 0.693. \tag{45-3}$$

At the end of one half life, one half of the original material remains in the

sample; at the end of two half lives, one quarter of the original material will still be in existence, and so on.

The half lives vary considerably among the radioactive elements. Radium, for example has a half life of 1,620 years, while that of radon is 3.82 days. In general, we find that the process of alpha emission exhibits the longest half lives, ranging up to billions of years. The process of beta emission is of intermediate lifetime, while the process of gamma-ray emission has half lives ranging from months to intervals of 10^{-10} sec. The half life of a radioactive substance may be determined by measuring its *disintegration rate dN/dt*. This may be measured with an ion chamber and an electrometer by measuring the ion currents generated by the emitted radiations, or by means of a *Geiger counter*.

One interesting application of the concept of radioactive disintegration lies in the field of *radiocarbon dating*. Carbon 14 is a radioactive isotope of carbon with a half life of 5,580 years. In terms of the earth's age of several billions of years, the half life of carbon 14 is very small, and any of this isotope originally present would be very much depleted. It is believed that cosmic radiation incident upon the earth's atmosphere continually replenishes the supply of carbon 14, by the interaction of neutrons liberated in the atmosphere by cosmic rays, with atmospheric nitrogen. Carbon is a constituent of all living matter. When a living organism dies, its intake of carbon 14 from the atmosphere ceases, and the percentage of this isotope in the carbon constituent of organic matter diminishes exponentially after death. It is thus possible to determine the age of archaeological and geological samples containing carbon by determining the percentage of carbon 14 in them and by assuming that the rate of production of this isotope has remained reasonably constant. Wherever possible, age determinations by radiocarbon dating are compared with other reliable methods as a check, and it has been found that carbon dating has been very satisfactory. It is now possible to use carbon dating with reasonable accuracy for age determinations up to about 30,000 years.

45-3 Radioactive Series

In the years following the discovery of radioactivity of uranium by Becquerel in 1896, many elements and isotopes of elements were found to be radioactive. Most of the naturally radioactive isotopes were found to be genetically related and fitted into one of three *radioactive series*. These are known as the *uranium series*, the *thorium series*, and the *actinium series*. Radium, for example, is a member of the uranium series. This series starts with the isotope of uranium, ${}_{92}U^{238}$, which has a half life of 4.50×10^9 years. The other members of the series are formed through a succession of alpha-particle and/or beta-particle emissions, as shown in Figure 45-2.

The emission of an alpha particle by an isotope having given values of Z and A results in the production of a new isotope of atomic number $Z - 2$ and mass number $A - 4$. The emission of a beta particle leaves the mass number unchanged but increases the atomic number of the product nucleus to $Z + 1$.

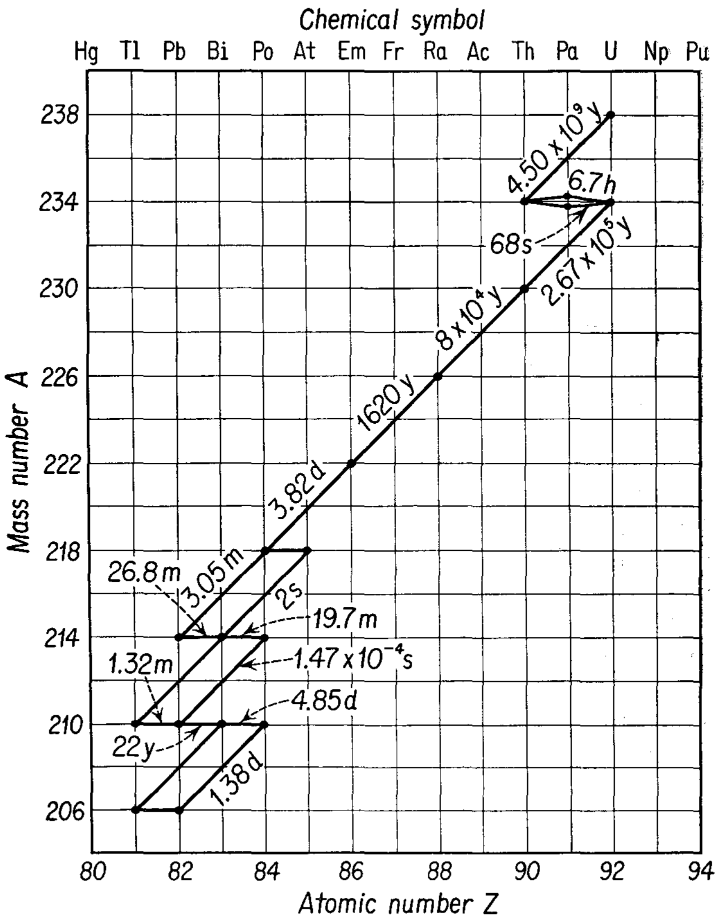


Fig. 45-2 The naturally radioactive uranium series. The half lives of the disintegrations are expressed in either years (y), days (d), hours (h), minutes (m), or seconds (s).

Each of the three naturally radioactive series terminates with a nonradioactive isotope of lead. In the uranium series this isotope has a mass number of 208, in the actinium series its mass number is 207, and in the thorium series its mass number is 208.

The half lives of the radioactive isotopes of the uranium series are

indicated in Figure 45-2. If a sample of ore containing uranium is analyzed chemically, and the percentage of uranium and lead is measured, it is possible to calculate from the ratio of uranium to lead how long a time must have elapsed for the radioactive production of this amount of lead. This value can be used as an indication of the age of the earth. From such measurements the age of the earth is known to lie between 3 and 6 billion years.

In our discussion of radioactive decay, it was tacitly assumed that a given isotope will have only one mode of disintegration. However, a glance at Figure 45-2 will show many interesting cases of *branching*. For example, the isotope of polonium with $Z = 84$ and $A = 218$, also known as radium A, will decay most often (in 99.96 per cent of the cases) with the emission of an alpha particle forming the product nucleus $_{82}Pb^{214}$. However, in a few cases (0.04 per cent) the nuclei will emit beta particles, forming the product nucleus $Z = 85$ and $A = 218$, an isotope of astatine. We interpret branching as indicating that a given nucleus may decay in two ways with different probabilities; the probability for disintegration by the first mode is λ_1, and the probability for disintegration in the second mode is λ_2. In this case the total probability for disintegration by any mode is $\lambda = \lambda_1 + \lambda_2$. The isotope disintegrates according to the disintegration constant λ, but the branching ratios or the percentage of the disintegrations which take place in a particular branch is determined by the constants, λ_1 and λ_2.

When a radioactive decay chain is in equilibrium, the number of atoms of any particular isotope remains substantially constant. Thus if the chain or series consists of isotopes A, B, C, and so on, which are formed by successive disintegrations, the number of atoms of isotope B remains constant, implying that the number of B atoms formed from A is equal to the number of disintegrating B atoms, in any time interval. From Equation (45-1) this implies

$$\lambda_A N_A = \lambda_B N_B = \lambda_C N_C, \tag{45-4}$$

and we may determine the disintegration constant of a long-lived member of the chain by chemically isolating one of the shorter-lived members C. The disintegration constant λ_C may be measured by means of an ion chamber, and N_C may be determined with an analytical balance and a knowledge of the atomic weight of this isotope. If, then, element A is separated chemically and weighed, the number N_A of A atoms present may be determined, and λ_A may then be calculated from Equation (45-4).

45-4 Radioactive Disintegration by Alpha-Particle Emission

The alpha particle is the nucleus of the helium atom. This was suggested by experimental determinations of the ratio of charge to mass determined

by Rutherford and co-workers in an apparatus similar to a mass spectrograph. To make the identification certain Rutherford and Royds, in 1909, carried out a spectrographic analysis by collecting the alpha particles emitted by the radioactive element radon into a spectroscopic tube which had previously been thoroughly evacuated. Using a spectroscope, they examined the light emitted by this tube under electrical excitation, and observed the spectrum of helium. This spectroscopic evidence proves conclusively that alpha particles are helium nuclei. The alpha particles became helium atoms by the capture of electrons from the glass walls of the tube.

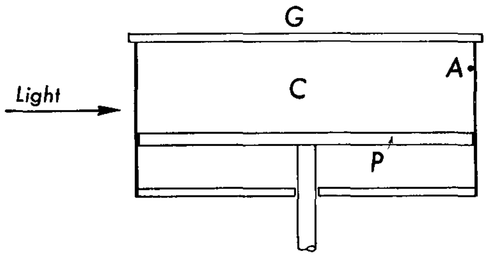


Fig. 45-3 Schematic diagram of a Wilson cloud chamber.

The energy of the alpha particles emitted from a particular isotope may be determined by measuring the radius of the circular orbit which the particles follow when they pass into a transverse magnetic field. When this is done, it is found that the alpha particles are emitted in a *line spectrum;* that is, the alpha particles are emitted with characteristic energies. This is interpreted as meaning that each alpha particle is emitted in a transition from a discrete state of the parent nucleus to a discrete state of the product nucleus. The energy of alpha particles may also be determined by measuring the *range* of the particle in a gas, such as air, using a *Wilson cloud chamber,* as shown in Figure 45-3. This apparatus consists essentially of a cylinder C containing a gas saturated with water vapor, and a piston P which may be lowered very rapidly to produce a sudden expansion of the gas in the chamber. As a result of this expansion, the gas is cooled and becomes supersaturated with water vapor. If there are any ions present in the gas, the water vapor will condense on these ions, forming small droplets. If a source of alpha particles is placed inside the chamber at A, then, in their passage through the gas in the chamber, the alpha particles will ionize gas molecules along their paths. During each expansion of the gas, water droplets form on the ions, showing the path of each individual alpha particle. Typical alpha-ray tracks are shown in Figure 45-4. These tracks are, in general, straight lines almost up to the end of the range. Occasionally, a track is bent sharply, or else it branches off into two tracks. These are usually ascribed to collisions with nuclei of the gas.

An important relationship exists between the range of alpha particles and the disintegration constant of the emitter, known as the *Geiger-Nuttall* law, which is usually written in the form

$$\log R = A \log \lambda + B, \tag{45-5}$$

where A is a constant which has practically the same value for each of the three radioactive series, and B is a constant which has a different value for each series.

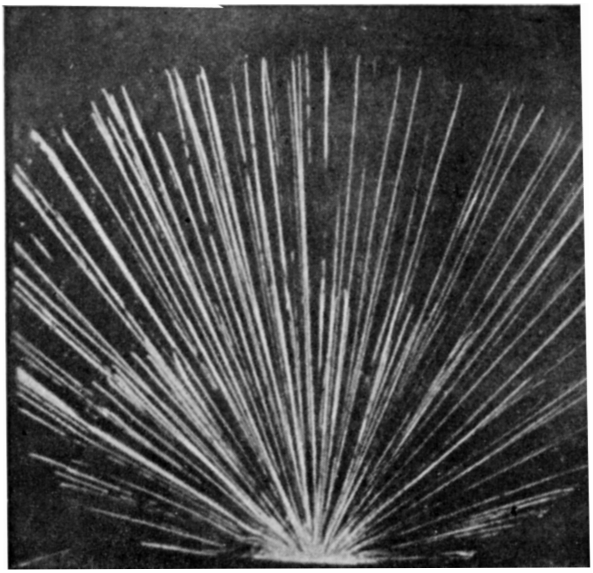


Fig. 45-4 Tracks of α particles from thorium $(C + C')$ in a Wilson cloud chamber, showing two distinct ranges. (From Rutherford, Chadwick and Ellis, *Radiations from Radioactive Substances*. By permission of The Macmillan Company, publishers.)

The radioactive disintegration of a nucleus by alpha-particle emission was first successfully explained, in terms of the penetration of a potential barrier, by Gamow, Condon, and Gurney in 1928. The alpha particle emitted from a nucleus in a given transition comes out of each nucleus with the same kinetic energy, but this energy is much smaller than the height of the barrier. The barrier arises from the Coulomb repulsive force upon an alpha particle. At the radius of the nucleus, given by

$$R = r_0 A^{1/3}, \tag{45-6}$$

where R is the nuclear radius, and the radius parameter $r_0 = 1.4 \times 10^{-13}$ cm, the potential energy of an alpha particle would be about 30 Mev in U238, while the kinetic energy of the emitted alpha particles is only 4.2

Mev. From a classical point of view, the alpha particle should never leave the nucleus, for it is surrounded by high potential walls which it cannot surmount, as shown in Figure 45-5. Thus, classically, there should be no such thing as alpha particle emission. When the alpha particle is considered as a wave, incident upon the barrier in one of its reflections back and forth within the nucleus, its passage through the barrier is determined by considerations similar to the transmission of light through a film. The thickness of the film seen by the alpha particle depends upon its energy $\mathcal{E}_\alpha$.

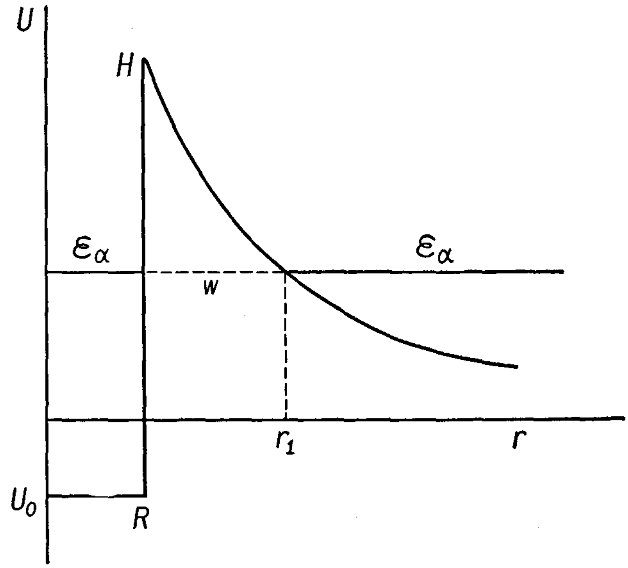


Fig. 45-5 Potential barrier presented by the nuclear charge and nuclear forces to an alpha particle. An alpha particle of energy $\mathcal{E}_\alpha$ sees a barrier of width w and height H

Particles of high energy see a thin film which they pass through quite readily, while particles of low energy see a thick film, for which the probability of penetration is small. Using these wave mechanical ideas and the appropriate equations, Gamow, Condon, and Gurney were able to account quantitatively for the Geiger-Nuttall law.

The transmission of alpha particles through a nuclear barrier is analogous to the transmission of light through a thin air film. Let us suppose that light is incident upon the face of a totally reflecting prism, as shown in Figure 45-6. When the hypotenuse of the prism is placed in contact with a second prism whose hypotenuse is slightly curved, we would expect light to be transmitted through the region of contact, but the question arises as to whether there should be any light passing through the thin air film because of the total internal reflection expected in the prism I. When the

light passing through both prisms is examined, we see a spot of light surrounded by interference maxima, indicating that there has been transmission of light through the air film, just as there is penetration of alpha particles through a potential barrier. We must understand that the concept of total internal reflection of light is a ray concept, analogous to

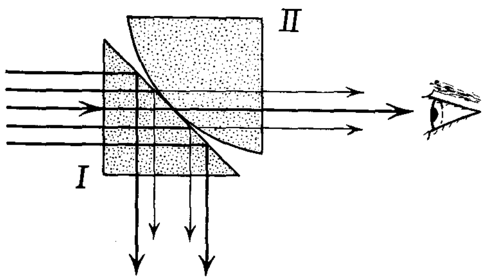


Fig. 45-6 Light partially transmitted through a thin film, though incident upon the face of a "totally reflecting" prism, *I*.

the particle concept of an alpha particle. When the problem of total internal reflection is examined from the wave point of view, the wave amplitude does not diminish to very small intensity for a distance of several wavelengths outside the total reflecting surface, and if a second glass surface is interposed, as in the case of prism II, some of the light is transmitted.

45-5 Scattering of Alpha Particles by Nuclei

An alpha particle is a comparatively massive particle possessing a considerable amount of kinetic energy. Its mass is about 7,500 times that of an electron, and its kinetic energy may be of the order of several million electron volts. It thus forms an ideal projectile for investigating the properties of atoms. Beginning in 1910, Rutherford and his co-workers, Geiger and Marsden, undertook a series of experiments in which alpha particles of known energy were fired at thin metallic foils, as shown in Figure 45-7. Most of the alpha particles went straight through the foils without deviation, many others were deviated, or scattered through small angles, but a few were deviated through angles greater than 90°; that is, they were scattered backward toward the side facing the incident beam.

Alpha particles carry a positive charge equivalent to twice the charge of an electron. Because of their large energy and great mass, the only way to account for the backward scattering of the alpha particles was to assume that they came very close to another massive charged particle. Furthermore, since most of the alpha particles incident on the foil went through it with little or no deviation, the massive charged particles must be very small in comparison with the distance between them, as indicated in Figure 45-8. On the basis of these experiments, Rutherford, in 1911, proposed his

nuclear theory of the atom. These experiments were then extended to determine the nuclear charge and the nuclear radius, and to produce nuclear disintegrations artificially by bombarding the nuclei with alpha particles. It must be emphasized that Rutherford's alpha-particle-scattering experi-

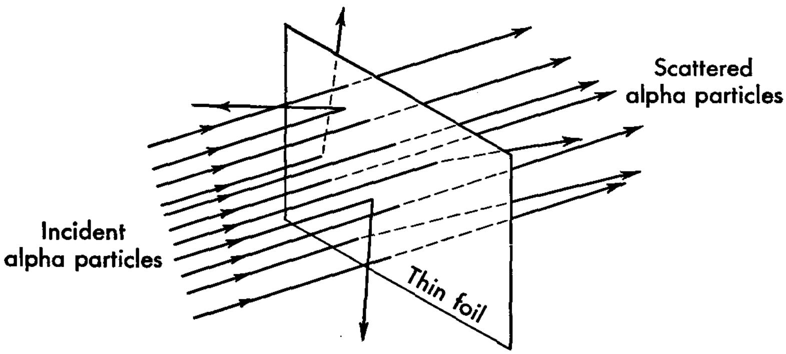


Fig. 45-7 Alpha particles directed against a thin foil. Most of them go through the foil either without any deviation or with very slight deviations from the original direction. A few alpha particles, however, are deviated through very large angles.

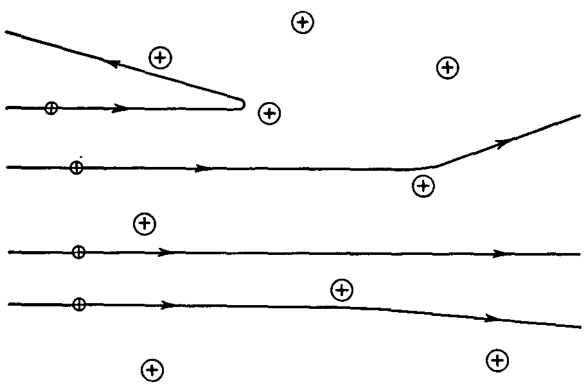


Fig. 45-8 Schematic diagram showing the paths of some alpha particles through a thin foil. The distances between nuclei (larger circles) are much greater than those shown here. A close approach of an alpha particle to a nucleus will therefore be a very rare event.

ment represented the beginning of nuclear physics and the nuclear theory of the atom. It was upon this foundation that the Bohr theory of the atom was laid. The technique of bombarding the nucleus with particles of high energy to explore its properties was introduced by Rutherford and forms the basis of the present-day methods for exploring nuclear forces and other nuclear properties. The higher the energy of the bombarding particle, the

farther into the nucleus the particle can penetrate, so that we may obtain greater knowledge of the nuclear structure. It is for this reason that physicists are continually engaged in the construction of accelerators capable of producing particles of higher and higher energies.

45-6 Beta-Ray Spectra

The beta particles, or electrons, emitted in the radioactive disintegration of a nucleus are most commonly investigated by measurement of the radii of curvature of their paths in a magnetic field of known induction B. One such arrangement is shown in Figure 45-9. In this arrangement beta rays from the source C are bent around by the magnetic field and are focused on the aperture O. The number of particles entering the aperture O is counted

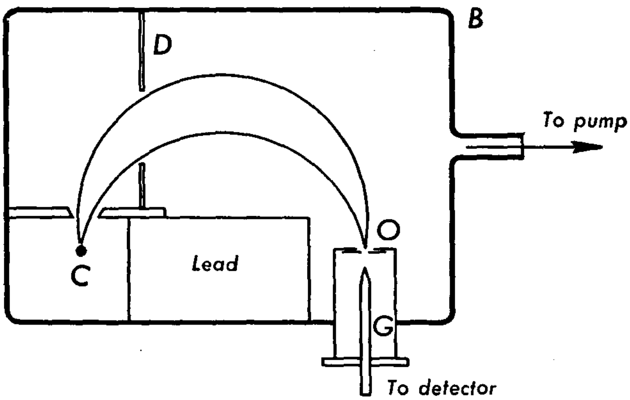


Fig. 45-9 Variable field magnetic spectrometer.

at a given value of the magnetic induction B. By repeating this determination at a number of different values of B, the number of particles in a given momentum interval can be determined as a function of the momentum.

The results of these experiments show that there are apparently two distinct types of beta-ray spectra, one a sharp line spectrum, and the other a continuous spectrum. The sharp line spectrum has been shown to consist of *internal conversion* electrons. A nucleus in an excited state may descend to the ground state either by the emission of a gamma ray, as we will see in the next section, or by a radiationless transfer of its energy to an atomic electron. The latter process is called internal conversion and takes place by the emission of a K, L, M, or N electron from the electronic shells of the atom. The continuous spectrum is that produced by the electrons emitted by the nucleus in the process of beta decay. A typical beta spectrum is shown in Figure 45-10, where the number of beta particles having a

given energy is plotted as the ordinate, and the energy of these particles, expressed in million electron volts, is plotted as abscissa. The beta-ray spectrum of an isotope differs remarkably from the other spectra characteristic of the same element in that all other characteristic spectra are line spectra, while the beta-ray spectrum is a continuous spectrum. Since the spectrum is presumably emitted in the transition of a nucleus of atomic number Z and mass number A in a definite energy state to a definite energy

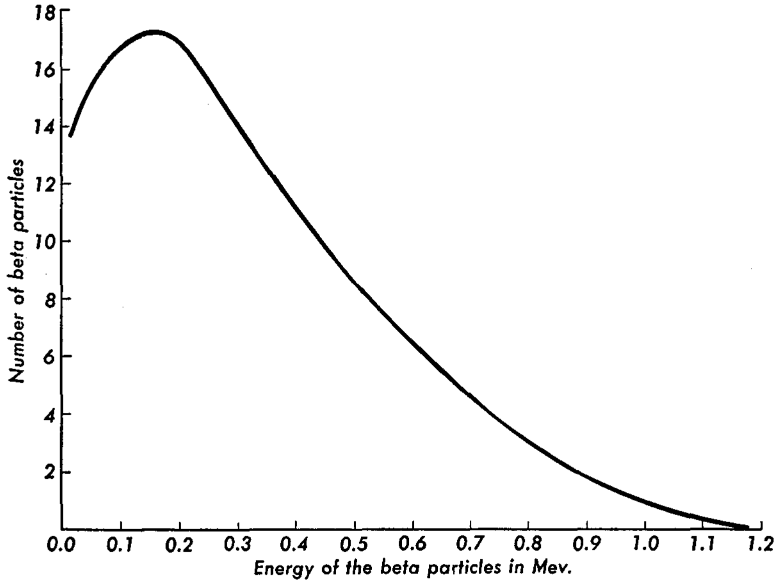


Fig. 45-10 Distribution of energy among beta particles emitted in the beta decay of bismuth, $A = 210$ (radium E).

state of a nucleus of atomic number $Z + 1$ and mass number A, it was difficult to understand why the spectrum was a continuous one.

It was originally believed that the beta particles were emitted with discrete energy, and that these particles lost energy on passing out of the sample by collision with the atoms of the sample. Calorimetric measurements showed that this was not so, and that the mean energy per disintegration corresponded to the mean energy of the spectrum rather than its end-point energy. A way out of the dilemma was suggested by Pauli in 1931. Pauli proposed that the actual disintegration energy corresponded to the end-point energy in the beta-ray spectrum, and that a second particle was emitted simultaneously, along with the beta particle. This particle, called the *neutrino*, was assumed to be an electrically neutral particle of very small mass. The disintegration energy of the nucleus is then shared among

the electron, the neutrino, and the product nucleus in such a way as to conserve energy and momentum. This implies that the spectrum is continuous, and the total energy of the particles involved in the process may be shared in any way consistent with the conservation laws.

When it was first proposed, the neutrino seemed an interesting fiction, invented to account for the missing energy in beta decay. As time passed, evidence accumulated to lend substance to the neutrino concept. In 1934 Fermi developed a theory of beta decay based on the neutrino which accounted for the shape of the beta-ray spectrum. Recently, Reines and Cowan (1953) have detected the neutrino in a very elaborate experiment utilizing the copious supply of neutrinos produced during the operation of a nuclear reactor. Very little detailed information is known about the neutrino. For example, its mass is known to be less than 0.05 per cent of the mass of the electron. At the present time the neutrino is required in order to obtain consistent explanations of many nuclear processes classed as weak interactions, and there seems to be little doubt that such a particle really exists. Our present theory of beta decay can be represented by the equation

$$n \text{ (in nucleus)} \rightarrow p \text{ (in nucleus)} + \beta + \nu;$$

that is, a neutron n in the nucleus decomposes into a proton p in the nucleus, and in the process an electron β and a neutrino ν are ejected from the nucleus. Although electrons are emitted by the nucleus in beta decay, there are no electrons in the nucleus.

In addition to the emission of an electron, there are two other processes involving neutrino's which may take place when a nucleus disintegrates; these are also classed as beta-decay processes. In one of these processes a positively charged particle, called a *positron,* is emitted from the nucleus. The positron has the same mass and the same numerical value of charge as the electron. Like the electron, the positron is emitted as a continuous spectrum along with a neutrino. Positrons are emitted by those isotopes whose nuclear charge is too great for their mass, in the nuclear stability diagram of Figure 44-5. A second mode of reduction of the nuclear charge by an unstable isotope is the process of *electron capture.* Here a nucleus absorbs one of its atomic electrons and ejects a neutrino. When the absorbed electron is in the K shell, the process is called K capture. We may visualize the processes of positron emission and electron capture as being incidental to the transformation of a proton in the nucleus into a neutron.

45-7 Gamma-Ray Emission. Isomerism

We have previously noted that the line spectra found in a beta-ray spectrometer may be attributed to the process of internal conversion, in which a

nucleus in an excited state makes a transition to a state of lower energy by ejecting an atomic electron. Accompanying these internal conversion electrons, it is found that gamma rays are emitted from the radioactive source; that is to say, some nuclei alter their states by the emission of conversion electrons, while in other cases the transition is effected by the emission of gamma rays. If the emitted gamma ray is of frequency f, the kinetic energy of the conversion electron is $\mathcal{E}$, and the K shell excitation energy as measured in x-ray absorption is $\mathcal{E}_K$, it is found that K conversion electrons obey the relationship

$$\mathcal{E} = hf - \mathcal{E}_K, \tag{45-7}$$

and similarly for the L, M, and N conversion electrons. By a study of the conversion electrons emitted from a radioactive source, we may determine both the gamma-ray energy associated with a particular transition and the element responsible for the transition, from the characteristic K-L and L-M shell energy differences.

Many gamma-ray transitions have measurable lifetimes, ranging from 10^{-10} sec to several months. Excited states which have measurable lifetimes are called *isomeric states.* One of the strong evidences for a nuclear shell structure is the existence of "islands of isomerism" in the N-P diagram of the radioactive isotopes just prior to the closure of a shell at one of the magic numbers. For example, there are 29 known cases of isomerism in odd A nuclei where either the odd proton number or the odd neutron number, as the case may be, lies between 39 and 49. As the magic number 50 is crossed, there are no cases of isomerism until we begin to approach the next magic number 82.

The nucleus must be in an excited state in order to emit gamma rays. This excitation may be accomplished in several ways. When a sample of a substance is placed in a nuclear reactor and bombarded with neutrons, some of the nuclei of the sample may capture neutrons. Some of the binding energy of a neutron to the nucleus is promptly emitted in the form of gamma radiation and is known as "capture gamma rays"; if the product nucleus has an isomeric state, the sample will contain some nuclei in this state for an appreciable time, and these may be subsequently studied. In a similar way it is possible to raise a nucleus to an excited state by irradiating a sample with high-energy x-rays. Nuclei may also be excited by collision with charged particles of sufficient energy, by a process known as *Coulomb excitation.* When a parent nucleus disintegrates by alpha or beta emission, these transitions may take place to an excited state of the product nucleus, with the subsequent emission of gamma rays and conversion electrons.

When all available information about the radiation emitted from a particular source is in hand, it is the problem of the nuclear spectroscopist

to attempt to assemble this information into an energy-level diagram which accounts for all the observed transitions. One such case is shown in Figure 45-11 for thorium C ($_{83}Bi^{212}$), a member of the thorium family of

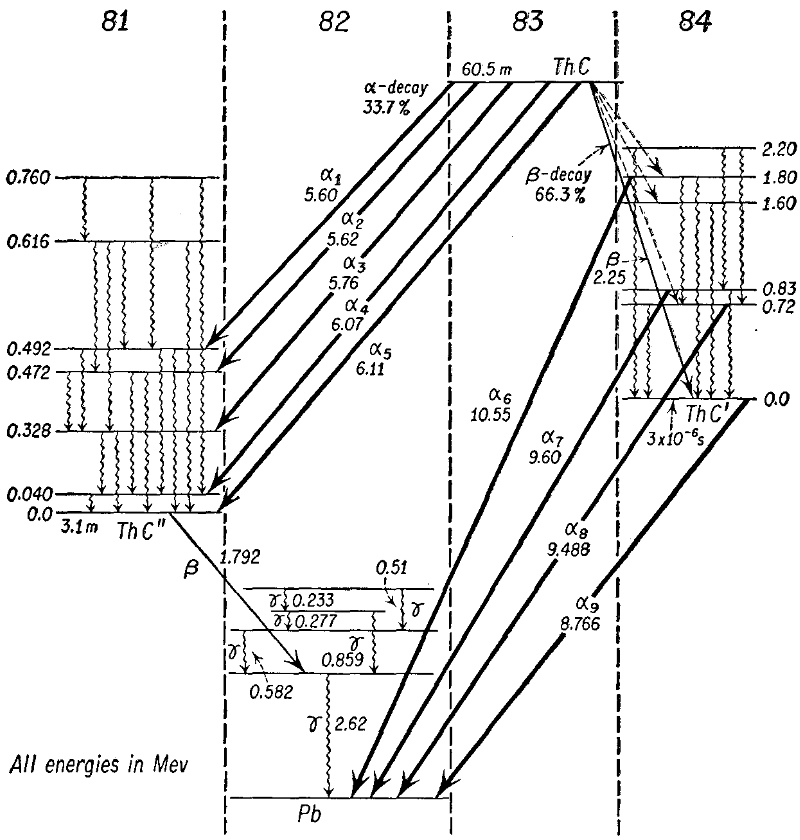


Fig. 45-11 Nuclear energy-level diagram for the disintegration of Th C. (After *Nuclear Data*, National Bureau of Standards, Circular 499.)

naturally radioactive elements. At the present state in the study of the nucleus, such diagrams are still being assembled, and no theory of nuclear structure comparable in scope to the Bohr theory of the atom has yet been devised.

45-8 Interaction of Radiation with Matter

We have studied some of the processes by which electromagnetic radiation interacts with matter. Thus light may be scattered by the free electrons

of a substance, generating the blue of the sky. In the limit of low energies of the incident photon, the scattering process may be treated as a wave phenomenon, in which the electric vector of the incident wave is thought to accelerate the electrons of a system and cause them to reradiate at the frequency of the incident radiation. When the wavelength gets sufficiently short, as in the case of x-rays, we note that the particlelike characteristics of electromagnetic radiation predominate, and it is necessary to supplement

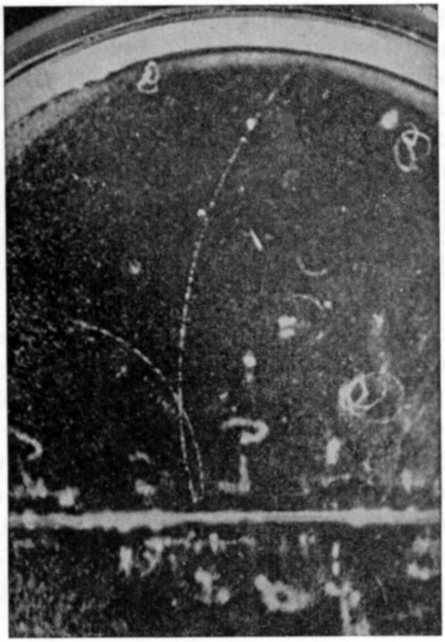


Fig. 45-12 Cloud-chamber photograph of the paths of a pair of oppositely charged particles, an electron and a positron, formed by the disintegration of a 5.7 Mev gamma-ray photon in its passage through a sheet of lead 0.002 inch thick. Magnetic field of 1,680 gausses is directed into the paper. (Photograph by H. R. Crane.)

the wave picture with a quantum picture, as in the Compton effect, in which the scattering process is treated as an elastic collision between a photon and a free electron. Looking toward the problem of low-energy scattering from the vantage point of the Compton effect, we see that these are two complementary views of the same scattering process, involving the scattering of electromagnetic waves by free electrons, in an elastic collision.

Electromagnetic waves interact with matter inelastically in the photoelectric effect, in that some of the incident energy of the radiation is absorbed by the atom or nucleus in the process of freeing the emitted particle. At low energies of the order of a few electron volts, electrons are

liberated from the outer shell of an atom or from the free electrons of a metal; at higher energies of the order of thousands of electron volts, as in x-rays, the electrons from the inner shells of an atom may be liberated; at extremely high energies of the order of millions of electron volts, the radiation may interact with nuclei, producing the nuclear photoeffects, involving the emission of protons or neutrons, or simply the excitation of a nucleus to a state of higher energy.

Another process of importance at high energies is *pair production*, in which a gamma-ray photon of energy greater than 1.02 Mev interacts with matter to form an electron-positron pair. A cloud-chamber photograph of such an interaction is shown in Figure 45-12. The determination of this threshold energy in itself constitutes a verification of the mass-energy equation, for mc^2 for the electron is 5.11×10^5 ev, according to this equation, so that we must not expect to find an electron-positron pair materialized from a photon unless the energy of the incident photon is greater than $2mc^2$, or 1.02 Mev. The process is also a verification of the law of conservation of charge, which we discussed in connection with electric currents. Since the electric charge of an incident photon is zero, the net charge of the materialized matter must also be zero. In the process of pair production, an electron and positron of equal and opposite charge are produced. When the incident photon has energy greater than 1.02 Mev, the excess energy is shared by the electron and positron as kinetic energy.

The relative importance of each of these processes varies with energy and with the atomic number of the material being irradiated. At low photon energies the photoelectric effect contributes most to the absorption coefficient, at intermediate energies the Compton effect predominates, while at high energies pair production is most important.

Problems

45-1. The half life of U238 is 4.50×10^9 years. (a) Compute the disintegration constant in sec^{-1}. (b) Calculate the number of alpha particles per second emitted by 1 gm of this isotope.

45-2. The half life of radon is 3.82 days. (a) Compute its disintegration constant in sec^{-1}. (b) Calculate the number of alpha particles emitted per second by 1 cm^3 of radon gas at 1 atm pressure and 0°C. Neglect the radiations from the product nuclei.

45-3. A *curie* is defined as the quantity of any radioactive isotope in which the number of disintegrations per second is 3.700×10^{10}. Using the data of Problem 45-1, how many curies are present in 1 gm of U238?

45-4. A hospital has 1 gm of radium $_{88}\text{Ra}^{226}$ which has a half life of 1,620 years in equilibrium with its decay products. Radium is a member of the uranium series, illustrated in Figure 45-2. What is the mass of radon $_{86}\text{Rn}^{222}$ present in the container?

45-5. A 1-millicurie source of radon is measured again after a period of 2 days. The half life of radon is 3.82 days. What is the activity of this source, in curies, at the time of the measurement?

45-6. What is the height of the Coulomb barrier experienced by an alpha particle in $_{92}U^{238}$?

45-7. When an alpha particle is emitted from a nucleus, the law of conservation of momentum requires that the nucleus must recoil. Not all of the available energy is delivered to the alpha particle as kinetic energy, for the nucleus acquires kinetic energy as well. The mass of radium is 226.10309 amu, the mass of radon is 222.09397 amu, and the mass of helium is 4.00388 amu. In the disintegration of radium into radon plus an alpha particle (a) what is the mass difference, in atomic mass units, between radium and its products? (b) The emitted alpha particle has a kinetic energy of 4.80 Mev. What is the kinetic energy, in million electron volts, of the recoil nucleus?

45-8. A gamma ray is emitted from Cs^{137} of energy 661.77 kev. In x-ray investigations it is found that K photoelectrons are emitted from cesium when irradiated with x-rays of wavelength 0.34473 A, and L photoelectrons are emitted when the wavelength is 2.17245 A or shorter. What are the energies of K and L internal conversion electrons emitted in the above transition in cesium?

46

Nuclear Reactions

46-1 Special Relativity

One of the most important developments of twentieth-century physics was the formulation of the special theory of relativity. This theory was an outgrowth of the failure of all attempts to show that the motion of the source of light relative to the observer had any effect on the speed of light. It is impossible to account for these experimental findings of Michelson and Morley, and others, on the basis of classical mechanics and electromagnetic theory. In 1905, Albert Einstein put forth the suggestion that all experimental findings would be clarified if it were assumed that *the speed of light is a constant and is independent of the relative motion of the source and the observer.* This statement forms the first postulate of the special, or restricted, theory of relativity. The second postulate of the theory is that *all systems which are in uniform motion relative to one another are equally valid frames of reference, and all fundamental physical laws must have the same mathematical forms in each of these reference frames.* Einstein expressed the viewpoint that all motion was relative motion, that there was no absolute coordinate frame, and that it was impossible to distinguish between a state of rest and a state of uniform translational motion by any physical experiment whatever. Thus, if the statement that the velocity of light was 3×10^{10} cm/sec was a fundamental physical law, every observer in uniform translational motion who measures the velocity of light must obtain this value, regardless of the motion of the source of light.

Let us consider some of the immediate implications of the first postulate. Suppose that we have two observers on two coordinate frames which are in relative motion with respect to each other. For simplicity we will orient the axes parallel to each other and will call one of these the unprimed frame and the other the primed frame. Let us suppose that the primed frame moves in the x direction with velocity v as seen in the unprimed frame. At the instant the origin of the two frames overlap, we cause a pulse of light to be emitted. From the first postulate both observers must

see the light spreading out as a spherical Huygens wave. The equation of this sphere as seen by the observer in the unprimed frame is

$$x^2 + y^2 + z^2 = c^2t^2, \tag{46-1}$$

which is the equation of a sphere of radius ct. At the moment we can assume nothing whatever about the primed frame except that the velocity of light is c. Thus we write that the Huygens wave front, as seen in the primed frame, may be represented by the equation

$$x'^2 + y'^2 + z'^2 = c^2t'^2. \tag{46-2}$$

We seek to find a set of equations representing the coordinate transformations between the primed and the unprimed frames. Such a transformation must involve only the first powers of the coordinates, for otherwise a single point in one system might become two or more points in the second system. Furthermore, we have no right to assume that the time t measured in the unprimed frame is the same as the time t' measured in the primed frame. When these conditions are applied, it is found that the appropriate transformation equations are

$$x' = \gamma(x - vt),\ y' = y,\ z' = z,\ t' = \gamma\left(t - \frac{vx}{c^2}\right), \tag{46-3}$$

where

$$\gamma = \left(1 - \frac{v^2}{c^2}\right)^{-1/2}.$$

This result may be verified by substituting into Equation (46-2) to obtain Equation (46-1). Equations (46-3) are called the Lorentz transformation equations. These equations imply that an event which takes place in the unprimed frame at coordinates x, y, z, and t will be reported by an observer in the primed frame as taking place in his frame at coordinates x', y', z', and t'. We assume that both observers have meter sticks and clocks with which to make measurements, and that these sticks and clocks are in agreement with each other when set side by side. The transformation equations work in both directions. To find the coordinates in the unprimed frame of an event taking place in the primed frame, we may solve Equations (46-3), or, more simply, we note that the unprimed frame is moving in the $-x$ direction, with velocity $-v$, with respect to the primed frame, so that we obtain

$$x = \gamma(x' + vt'),\ y = y',\ z = z',\ t = \gamma\left(t' + \frac{vx'}{c^2}\right). \tag{46-4}$$

In the Lorentz transformation equations we note that time is reduced to the status of a coordinate, that it has no special properties.

While we cannot derive all the important results of relativity theory, two important consequences may be simply obtained. These are the *Lorenz-Fitzgerald contraction* of length, and *time dilatation*.

Let us suppose that a rod of length l is at rest in the unprimed system, and that the coordinates of its end points are given by x_2 and x_1 such that $l = x_2 - x_1$. An observer in the primed system who wishes to determine the length of the rod measures the coordinates of the ends of the rod at the same time t', as read on his clocks, and finds that the rod is of length $l' = x_2' - x_1'$. We may determine the relationship between l and l' by applying Equations (46-4). Thus we note that

$$x_2 - x_1 = \gamma(x_2' + vt') - \gamma(x_1' + vt'),$$

or

$$l = \gamma l'. \tag{46-5}$$

Since the speed of the coordinate frames with respect to each other is less than the velocity of light, γ is greater than 1, and we note that the length l' is less than the length l. The observer in the moving coordinate frame sees the rod as contracted, as compared to the length seen by the observer in the frame in which the rod was at rest. Note that we have used the transformation equations in the form of Equations (46-4) rather than Equations (46-3) for we knew that the measurements were made at the same time t' in the primed frame. We have no right to assume that the observations were made at the same time t in the unprimed frame, for the measurements were made at two different points in the primed frame. From Equations (46-4) this implies that the observer in the unprimed frame will infer that the measurements were made at two different times t_1 and t_2.

Suppose a clock is located at a fixed point x_1 in the unprimed system, and that this clock is used to measure the time interval Δt between two events which occur at times t_1 and t_2, such that $\Delta t = t_2 - t_1$. Applying transformation Equations (46-3), we find that an observer in the primed system would observe that the time interval between the two events was $\Delta t' = t_2' - t_1'$ such that

$$t_2' - t_1' = \gamma\left(t_2 - \frac{vx_1}{c^2}\right) - \gamma\left(t_1 - \frac{vx_1}{c^2}\right),$$

or

$$\Delta t' = \gamma\, \Delta t. \tag{46-6}$$

Again γ is greater than 1, so that the observer in the moving frame will claim that the time interval between the two events is greater than the interval recorded by the observer who has no relative motion with respect to the clock. This is called time dilatation. Since the Lorentz transformation equations are based upon the finite velocity of light, the same for all observers, these strange effects of length contraction and time dilatation are associated with the fact that, in actual measurement, information is obtained by means of light signals. When an object is moving with speeds approaching the velocity of light, we must examine with great care what we mean by length and time.

Among the results of relativity theory, we find that the simple rules for the vector addition of velocity must be modified. Thus we may no longer write

$$\mathbf{v} = \mathbf{w} + \mathbf{u}$$

for the vector addition of two velocities, but rather we find that, if the two velocities are directed along the same line, the resultant velocity is given by

$$v = \frac{w + u}{1 + \dfrac{uw}{c^2}} \cdot \tag{46-7}$$

The momentum $\mathbf{p}$ of a particle of mass m moving with velocity $\mathbf{v}$ is given by

$$\mathbf{p} = \frac{m\mathbf{v}}{\left(1 - \dfrac{v^2}{c^2}\right)^{1/2}} = \gamma m\mathbf{v}. \tag{46-8a}$$

The de Broglie wavelength to be associated with a particle of mass m moving with velocity v and momentum p is given by

$$\lambda = \frac{h}{p} = \frac{h}{\gamma m v} \cdot \tag{46-8b}$$

The energy of a moving particle is given by

$$\mathcal{E} = \gamma mc^2 = \frac{mc^2}{\left(1 - \dfrac{v^2}{c^2}\right)^{1/2}}, \tag{46-9}$$

and the kinetic energy $\mathcal{E}_k$ of a moving particle is the difference between its energy in motion and its mass energy, or

$$\mathcal{E}_k = \gamma mc^2 - mc^2 = mc^2(\gamma - 1), \tag{46-10}$$

which reduces to $\mathcal{E}_k = \frac{1}{2}mv^2$ in the case of slowly moving objects.

Equation (46-9) implies that the limiting speed for the motion of a particle is c; that is, any particle moving with speed c would have infinite energy.

If we solve the equation defining γ for v^2 we find

$$v^2 = c^2\left(1 - \frac{1}{\gamma^2}\right);$$

squaring Equation (46-8a) and substituting this result, we obtain

$$p^2 = \gamma^2 m^2 v^2 = m^2c^2(\gamma^2 - 1).$$

We may solve this equation for γ^2 to find

$$\gamma^2 = \left(1 + \frac{p^2}{m^2c^2}\right).$$

Substituting this result into the square of Equation (46-9), we find

$$\mathcal{E}^2 = \gamma^2 m^2 c^4 = p^2c^2 + m^2c^4. \tag{46-11}$$

Equation (46-11) is extremely interesting, for it illustrates the unity between mass particles and quanta. The relation between energy and momentum for particles of zero rest mass reduces to the familiar relation for quanta $\mathcal{E} = pc$, while a particle whose momentum is zero has energy given by the mass-energy equation.

One further result is of interest to us at this time. We have seen that a charged particle whose velocity is transverse to a field of flux density B moves in a circular path whose radius r is given by $Be = (mv)/r$, where e is the charge of the particle, and m is its mass. At high velocities we must replace mv by the relativistic expression for momentum to find

$$Be = \frac{\gamma mv}{r}. \tag{46-12}$$

The relationships and concepts of the theory of special relativity are presented in the preceding paragraphs to indicate some of the fundamental changes in our concept structure and in the equations of mechanics which must be introduced in the study of rapidly moving particles. These equations reduce to more familiar forms at the velocities of ordinary experience. These equations must be used in the analysis of the data obtained with a beta-ray spectrometer and in the analysis of nuclear radiations. Here the equations find ample experimental verification. Perhaps the most detailed verification of relativity theory lies in its application to the design, construction, and successful operation of modern high-energy accelerators.

46-2 Particle Accelerators

There are many different types of devices designed and built to accelerate particles to high energies. We have already described one of these, the betatron, which accelerates electrons to energies up to 300 Mev. Using other methods, electrons can now be accelerated up to energies of about 2 Bev (billion electron volts), and protons can be accelerated up to about 6 Bev. Designs and experiments are now in progress for a device to accelerate protons to 100 Bev.

The forerunner of this development of particle accelerators is the *cyclotron*, which was originally designed and built by E. O. Lawrence and M. S. Livingston in 1931. It consists essentially of a short, hollow, metal

cylinder, divided into two sections A and B, as shown in Figures 46-1 and 46-2. The cylinder is placed between the poles of a large electromagnet.

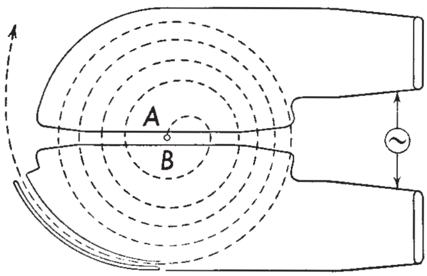


Fig. 46-1 The paths of particles in a cyclotron.

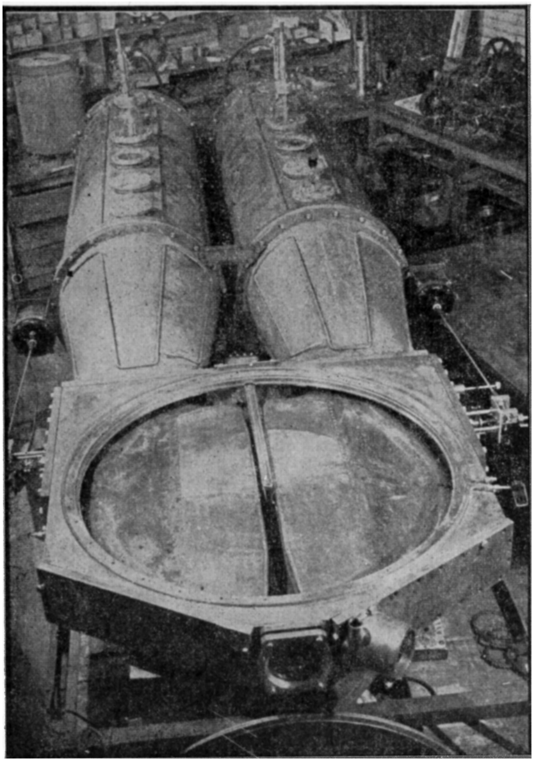


Fig. 46-2 Shop assembly photograph of the M.I.T. cyclotron chamber showing the construction of the chamber and dees. (From the *Journal of Applied Physics*, January, 1944. Courtesy of the Radioactivity Center at the Massachusetts Institute of Technology.)

Some of the magnets built have pole pieces from 30 to 60 in. in diameter, and a modified form of cyclotron in use in California has a diameter of

184 in. The magnetic field is perpendicular to the plane of the figure, that is, along the axis of the cylinder. The two half cylinders are connected to the terminals of a high-frequency a-c circuit, so that the charge on each half changes a few million times a second. The magnetic field causes ions, emitted from an ion source at the center of the cylinders, to travel in a

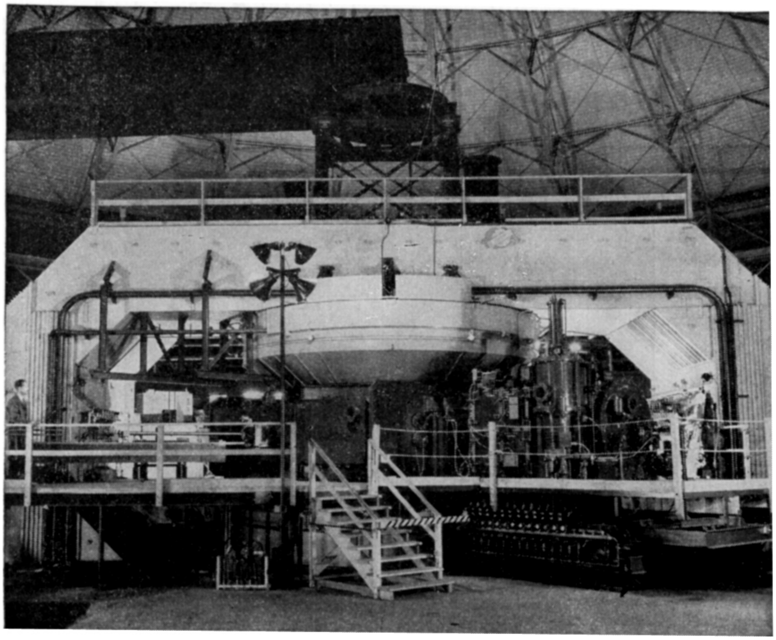


Fig. 46-3 The 184-in. Berkeley cyclotron at the Radiation Laboratory of the University of California at Berkeley. The vacuum chamber is in place between the pole pieces of the electromagnet. The tube extending into the chamber at the left carries the target, which is bombarded by the high-energy ions in the cyclotron chamber. (Photograph supplied by Prof. R. I. Thornton, Radiation Laboratory, University of California, Berkeley, Calif.)

circle of radius r given by Equation (46-12). Each time the particle nears the gap, the potential of the half cylinders, or *dees*, as they are called, is reversed, so that the particle is accelerated through the gap. If the timing is just right, the potential change is always synchronized with the passage of the particle through the gap, and the particle may be accelerated to high velocities. The potential difference between the sections A and B may have any value from about 10,000 to 200,000 volts. The particle emerging from the cyclotron may have an energy of several million volts due to the successive accelerations it experiences. The cyclotron is thus a low-voltage source of high-energy particles.

From Equation (46-12) we see that the radius of the orbit of the particle is given by

$$r = \frac{\gamma m v}{Be}.$$

The maximum magnetic induction B obtainable with the use of iron is about 20,000 gausses, so that the radius of the pole face of the magnet presents a limit on the maximum velocity which can be achieved with a cyclotron. The time t required for a particle to traverse a semicircular path is

$$t = \frac{\pi r}{v} = \frac{\pi \gamma m}{Be}.$$

As long as the velocity of the ion is considerably less than c, γ is close to 1, and the time required to traverse the semicircle is independent of the speed of the particle. Thus the cyclotron may be driven by a fixed-frequency oscillator. When the velocity is such that γ is appreciably different from 1, it is necessary to vary the frequency of the oscillator to maintain the proper timing. The frequency is varied in synchronism with the changes in t, and the accelerator is then called a *synchro-cyclotron.* The synchro-cyclotron of the University of California is shown in Figure 46-3.

There are many other important types of accelerators which cannot be discussed here, for the principles of their operation are beyond the scope of this book. The particles accelerated are always charged particles and include electrons, protons, deuterons, helium ions, and ions of more massive atoms.

46-3 Nuclear Reactions with Protons, Deuterons, and Neutrons

With the high-energy particles now available, many different types of nuclear reactions have been produced. For example, when lithium of mass 7 is bombarded with protons, the compound nucleus thus formed breaks up into two alpha particles having an energy of about 8.63 Mev each. The reaction is

$${}_3\mathrm{Li}^7 + p \rightarrow ({}_4\mathrm{Be}^8) \rightarrow {}_2\mathrm{He}^4 + {}_2\mathrm{He}^4 + Q;$$

the reaction energy Q is 17.28 Mev, which checks very well with the value obtained from the difference in the masses of the initial and final nuclei. Sometimes the proton captured by the lithium nucleus remains with the compound nucleus, beryllium, and the excess energy is emitted in the form of a gamma ray of 17 Mev energy. The ${}_4\mathrm{Be}^8$ nucleus subsequently decays into two alpha particles with a half life which is probably less than 10^{-14} sec. A reaction in which a proton is absorbed and a gamma ray is emitted is known as a (p, γ) reaction.

Another interesting reaction occurs when high-energy deuterons bombard deuterium. Two different reactions have been observed:

$$_1H^2 + {_1H^2} \rightarrow (_2He^4) \rightarrow {_1H^3} + {_1H^1} + Q,$$

and

$$_1H^2 + {_1H^2} \rightarrow (_2He^4) \rightarrow {_2He^3} + {_0n^1} + Q.$$

The first of these reactions results in the production of *tritium*, an isotope of hydrogen of mass number 3, identified with the aid of the Wilson cloud chamber. The value of Q for this reaction was determined from the range of the particles in air, yielding a value of $Q = 4.03$ Mev. With the value of Q, and the known masses of hydrogen and deuterium, the mass of tritium can be determined very accurately. This is at present the most accurate method for determining the mass of H^3. This method is now widely used for determining the atomic masses of isotopes formed in nuclear reactions for which accurate Q-values are known. The second of the above reactions results in the production of neutrons and is one of the simplest methods for obtaining neutrons of known energy. Neutrons have been observed for comparatively low values of incident deuteron energies, as low as 6 kev. The neutron yield increases with reaction energy. The value of Q for this reaction has been found to be $Q = 3.18$ Mev.

When nitrogen is bombarded with neutrons, a radioactive isotope of carbon of mass number 14 is produced, according to the following reaction:

$$_7N^{14} + {_0n^1} \rightarrow (_7N^{15}) \rightarrow {_6C^{14}} + {_1p^1}.$$

The carbon isotope produced in this reaction has a half life of 5,580 years. This same reaction takes place in the atmosphere, induced by neutrons liberated by cosmic rays, and is responsible for the existence of radiocarbon dating.

46-4 Nuclear Cross Section

The concept of a nuclear cross section is a very useful one in nuclear physics. In general, we do not refer to the geometrical cross section, as though the nucleus were a sphere; rather, we use the term "cross section" as a *measure of the probability of the occurrence of a given process.* Thus, if there are n nuclei per unit volume of a given substance, a foil of thickness t will contain nt nuclei per unit of surface area of the foil. The probability that a single particle will strike one of these nuclei if it is directed at a unit area is simply the fraction of the surface area of the foil occupied by nuclei. Thus if σ is the nuclear cross section, the probability that an incident particle will strike a nucleus is σnt. If an incident beam whose flux is N particles per unit area per second strikes the foil, the number of interactions with nuclei, per second, will be N_i, according to the equation

$$N_i = N\sigma nt. \tag{46-13}$$

Equation (46-13) may be taken as the defining equation for the cross section for a particular process. Thus we have scattering cross sections, capture cross sections, and so on. A large number of nuclear cross sections are of the order of 10^{-24} cm^2, so that it has been convenient to use a separate name for this area. The name adopted is a *barn;* that is, 1 barn = 10^{-24} cm^2. Nuclear forces are short-range forces, so that the scattering cross section for fast neutrons might be expected to reflect the physical size of the nucleus, and, indeed, this cross section is one of the means for determining the nuclear radius as given in Equation (45-6).

46-5 Nuclear Fission

All the nuclear disintegrations described thus far have concerned the emission of comparatively light particles, such as electrons, positrons, protons, and alpha particles. A new type of process known as *nuclear fission* was discovered by Hahn and Strassman in 1939 in a series of experiments in which uranium was bombarded by neutrons. Chemical analysis of the products of disintegration showed the presence of barium, $Z = 56$, and lanthanum, $Z = 57$, and other elements of medium atomic weights. The interpretation of these results is that when a neutron is captured by a uranium nucleus, the compound nucleus formed becomes unstable and disintegrates into two particles of intermediate masses; for example, if one of the particles is barium, $Z = 56$, the other particle is krypton, $Z = 36$. Cloud-chamber photographs, as in Figure 46-4, have verified this hypothesis.

The particles which are produced in the fission of uranium have energies of the order of 200 Mev. The source of this energy is the difference in mass between the reacting particles, the neutron and the uranium nucleus, and the product or fission particles. There is a decrease in mass of about 0.1 per cent in this process; thus, in the nuclear fission of 1 kg of uranium, there is a decrease in mass of about 1 gm, and this corresponds to about 25×10^6 kw hr of energy.

The masses of the fission products are found to be those of unstable isotopes; that is, they have many more neutrons than the stable isotopes of the corresponding elements. One of the first questions investigated was the manner in which these unstable fission products disintegrated, particularly whether any of the excess neutrons were emitted in this process. Early experiments showed that between 2 and 3 neutrons were emitted per nuclear fission. The process can now be represented schematically, as shown in Figure 46-5: when a neutron is captured by a uranium nucleus of mass number 238, a new isotope of mass number 239 is formed; in the process of nuclear fission the latter splits into two isotopes of intermediate masses, say barium and krypton, with the prompt emission of 2 neutrons. A variety of other pairs of nuclei may be produced in the fission process,

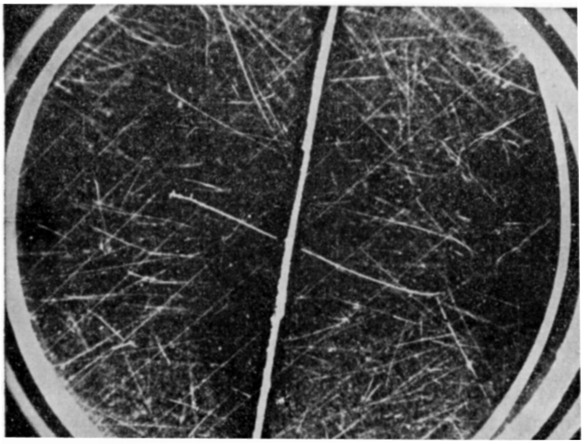


Fig. 46-4 Cloud-chamber photograph showing the fission of uranium. The foil in the center of the cloud chamber is coated with uranium and bombarded by neutrons. The tracks of the two heavy fission particles can be seen coming from the foil where a uranium atom has undergone fission as a result of the capture of a neutron. (From a photograph by J. K. Boggild, K. K. Brostom, and T. Lauritsen.)

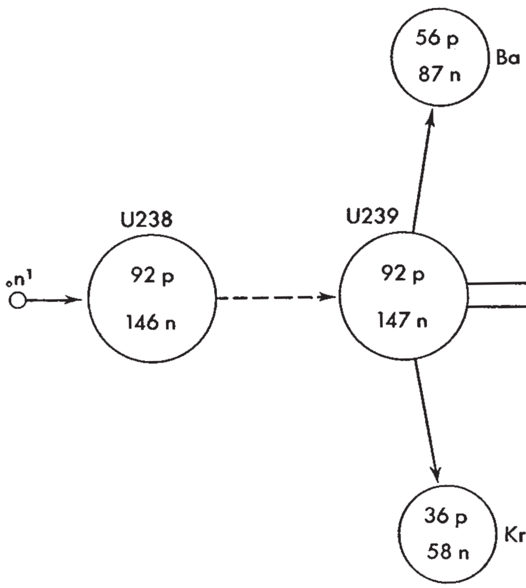


Fig. 46-5 Nuclear fission of uranium. A fast neutron is captured by a nucleus of uranium of mass number 238 forming uranium 239; the latter splits into two comparatively massive particles, in the above case krypton and barium, with the simultaneous emission of two fast neutrons.

all of them radioactive, most of them decaying to a stable isotope by the emission of beta rays; gamma rays are also emitted by many of these isotopes.

In addition to uranium, thorium, $Z = 90$, and protoactinium, $Z = 91$, have been found to be fissionable by the capture of neutrons, and a new element, plutonium, $Z = 94$, is also fissionable by the capture of neutrons. Fission may also occur spontaneously, by excitation of a nucleus with high-energy gamma rays, and by the bombardment of heavy nuclei with protons, deuterons, or alpha particles. The fission cross section varies with the energy and type of incident particles. In the following sections, we shall discuss only neutron-induced nuclear fission.

46-6 A Nuclear Chain Reaction

The concept of a nuclear chain reaction is very simple: if a single nuclear fission process involving the capture of one neutron results in the release of energy and simultaneously the release of more than one neutron, it should be possible to so arrange the mass of fissionable material to ensure the capture of the newly released neutrons. Or, stated another way, the mass of fissionable material should be so arranged that at any one place the number of new neutrons produced should be equal to the number of free neutrons originally present at that place. The ratio of these two numbers of neutrons is called the *multiplication factor* K. If $K = 1$, the chain reaction will be self-sustaining; if K is less than 1, the process will ultimately come to a halt; if K is greater than 1, the neutron density will increase and may lead to an explosive reaction. A mass of fissionable material so arranged that the multiplication factor is equal to or greater than 1 constitutes a *nuclear reactor*.

In order to be able to design a nuclear reactor, it is essential to know the conditions under which neutrons are captured by nuclei and the conditions under which such capture of neutrons results in the fission of the product nuclei. We shall restrict this discussion to the fission of uranium. Ordinary uranium consists of 3 isotopes: one of mass number 238, another of mass number 235, and a third of mass number 234. The most abundant of these is U238—about 99.3 per cent abundance. The amount of U234 in ordinary uranium is negligible. U235 constitutes about 0.7 per cent of ordinary uranium. Experiments show that U238 is fissionable only if it captures *fast* neutrons, that is, neutrons having energies of 1 Mev or greater. On the other hand, U235 is fissionable with neutrons of any speed, and the fission cross sector is particularly high with *slow* neutrons, that is, neutrons having energies corresponding to the thermal energies at ordinary temperatures. These energies are much less than 1 ev.

The neutrons released in nuclear fission have a wide range of energies.

In the case of the fission of U235, these energies extend up to about 17 Mev, with a maximum number having energies of about 0.75 Mev. If such neutrons are captured by other uranium nuclei, they produce nuclear fission. However, not every collision between a fast neutron and a uranium nucleus results in capture of the neutron; the collision may simply produce a decrease in the energy of the neutron. Thereafter the probability of its capture will be very small; additional collisions will produce further reductions in the energy of the neutrons. At some particular values of energy, the neutron will be readily captured by U238, but such a capture does not result in nuclear fission. Instead, the newly formed isotope of uranium, U239, emits a gamma-ray photon and then becomes radioactive, emitting a beta ray with a half life of 23 min. The nuclear reaction equations are

$$_{92}\mathrm{U}^{238} + {_0n^1} \rightarrow (_{92}\mathrm{U}^{239}) \rightarrow {_{92}\mathrm{U}^{239}} + \text{gamma ray},$$

then
$$_{92}\mathrm{U}^{239} \rightarrow {_{93}\mathrm{Np}^{239}} + \beta^-. \qquad T = 23 \text{ min.} \tag{46-14}$$

The new element thus formed, called neptunium, Np, is itself radioactive, emitting a beta particle with a half life of 2.3 days. The product nucleus formed in this reaction is plutonium, Pu, of atomic number 94. The reaction in which this is formed is

$$_{93}\mathrm{Np}^{239} \rightarrow {_{94}\mathrm{Pu}^{239}} + \beta^-. \qquad T = 2.3 \text{ days.}$$

It is followed by
$$_{94}\mathrm{Pu}^{239} \rightarrow {_{92}\mathrm{U}^{235}} + {_2\mathrm{He}^4} \qquad T = 24{,}000 \text{ years.} \tag{46-15}$$

The isotope of plutonium is radioactive, emitting an alpha particle, but it has a very long half life—24,000 years. In this sense, it is a comparatively stable element. It will be noted that neptunium and plutonium are *transuranic* elements, that is, elements with atomic numbers greater than that of uranium. When plutonium disintegrates with the emission of an alpha particle, the resulting nucleus is U235. The plutonium isotope formed in the above process is fissionable by the capture of neutrons of any energy and is thus similar to U235 as far as the fission process is concerned. Since it is chemically different from uranium, it can be separated more readily from the uranium metal than the uranium isotope of mass number 235.

If ordinary uranium is to be used in a nuclear reactor, it is essential to avoid loss of neutrons by nonfission capture. Since slow neutrons can produce fission in U235, and since the probability of capture varies inversely with the speed of the neutron, one method of ensuring its fissionable capture is to slow down the neutrons very rapidly to thermal energies. This is done with the aid of a *moderator*, that is, a light element in which the probability of nuclear capture of a neutron is negligible, but in which collisions between neutrons and nuclei will cause a rapid decrease of the energy of the

neutron. Deuterium and carbon are two elements suitable for use as moderators.

The first nuclear reactor, or *uranium pile* as it is sometimes called, was operated successfully in Chicago on December 2, 1942; it was built under the direction of E. Fermi and operated by groups headed by W. H. Zinn and H. L. Anderson. A schematic diagram of the construction of a graphite-moderated uranium pile is shown in Figure 46-6. Rods of uranium metal are embedded in blocks of graphite; rods of boron metal are inserted at various places in the pile to control the flux of neutrons; boron nuclei

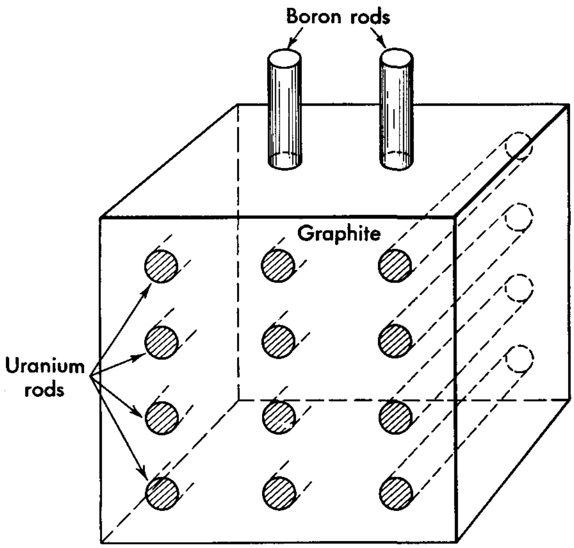


Fig. 46-6 Schematic diagram of a uranium pile. Cylindrical rods of uranium are embedded in a large mass of graphite which acts as a moderator to slow down neutrons. Boron rods, which are inserted into the pile, control its rate of activity.

capture neutrons very readily. No special source of neutrons is needed to start this pile; there are always neutrons present from cosmic rays, or from spontaneous fission, to start the nuclear reactor. The mode of its operation can be understood by referring to Figure 46-7. Suppose that a neutron is captured by a uranium nucleus, so that fission results and that two new neutrons are released with energies of about 1 Mev each. These neutrons then make several collisions with nuclei of the moderator, graphite, until their energies are reduced to thermal energies. Whenever one of these slow neutrons is captured by U235, fission will again occur with the release of, say, 2 neutrons. Some neutrons may be lost through the surface of the reactor; one way to reduce this loss is to make the reactor very large; the increase in the surface area is proportional to the square

of its linear dimension, while the volume is proportional to the cube of the linear dimension. Other neutrons may be lost through capture by impurities or through nonfissionable capture by U238. But if $K = 1$, the reaction will be self-sustaining. To prevent the multiplication factor from becoming excessive, boron rods are inserted to various depths in the pile to absorb the excess neutrons. One other control factor may be mentioned here; that is, not all of the neutrons are emitted promptly in nuclear fission; a small percentage of the neutrons are delayed, some by 0.01 sec, others by as much as 1 min.

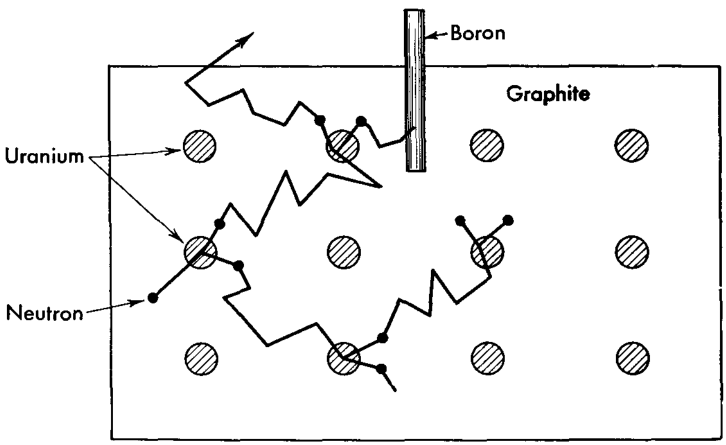


Fig. 46-7 Schematic diagram of the action of a neutron in a uranium pile based on the assumption that each fission process yields two neutrons. The shaded circles represent rods of uranium; the small circular dots represent neutrons. Sudden changes in direction of the neutron path are due to collisions with nuclei of the moderator graphite.

A whole new field of nuclear science and engineering has been opened as the result of the discovery of nuclear fission and following the successful construction of the first nuclear reactor. Nuclear reactors designed for many different purposes are now in operation throughout the world. Some are used as sources of energy for power plants; others are used for experimental purposes. A nuclear reactor is one of the best sources of neutrons for use in physical, chemical, and biological experiments. It is also a source of radioactive isotopes for medical and industrial uses. A nuclear reactor may also be designed as a military weapon known as an *atomic bomb* or *A-bomb*. The latter is a type of nuclear reactor in which the multiplication factor K is greater than 1. It may consist of uranium containing a large percentage of U235 or of plutonium 239. If the mass of fissionable material is less than a certain *critical* amount, K will be less than 1, and there will be no chain reaction. If the mass is built up rapidly so that the total

exceeds the critical mass, a very fast chain reaction will be produced. One of the problems in exploding an atomic bomb is to hold the material together for a sufficient time, probably several millionths of a second, so that a large quantity of the material will take part in the fission process. It has been estimated that the energy released in an atomic bomb is sufficient to raise the temperature of this material to several million degrees and produce pressures upon explosion of perhaps a few million atmospheres. In addition, great quantities of radioactive materials and gamma rays are produced.

46-7 The New Particles of Physics

There were only two fundamental particles known in physics before 1932, the proton and the electron. It was then believed that all matter in the universe was composed of these two particles. The picture changed when the neutron was discovered in 1932, followed shortly by the discovery of the positron in 1934. Among the theories proposed to explain nuclear forces and nuclear phenomena was one put forward by H. Yukawa in 1935 in which it was assumed that a nuclear field of force, called a *meson field*, exists between nucleons. Furthermore, this field has particles, called *mesons*, associated with it in a manner analogous to the association of photons with an electromagnetic field. The Yukawa theory predicted that the mass of the mesons should be intermediate between the mass of an electron and that of a proton. This Yukawa particle or meson could have a positive charge, a negative charge, or zero charge, the charge being equivalent to that of one electron.

Many such particles have since been discovered; the first one, known as a *mu meson* (μ meson) or *muon*, was discovered in 1937 by S. H. Neddermyer and C. D. Anderson and independently by J. C. Street and E. C. Stevenson. All the others were discovered after 1947. Most of these new particles were discovered in the study of *cosmic rays*, a type of very energetic radiation which reaches the earth's atmosphere from outside, penetrates the atmosphere, and, in so doing, reacts with atmospheric particles giving rise to many nuclear disintegrations. The cosmic rays as observed at various places in the atmosphere will usually consist of a combination of *primary cosmic rays*, which consist almost entirely of protons and other nuclei, plus the secondary radiations or particles produced by the interactions of the primaries with matter. With the development of high-energy particle accelerators, many of the new particles can be produced and studied under controlled conditions.

The particle which reacts most strongly with nuclei and is assumed to be the Yukawa particle is the *pi meson* (π meson) or *pion*. This was first discovered by Lattes, Occhialini, Powell, and D. H. Perkins in 1947 in

high-altitude cosmic-ray investigations using special photographic plates. In 1948, E. Gardner and M. Lattes, using the 184-in. Berkeley cyclotron, bombarded a carbon target with 380-Mev alpha particles and showed that pi mesons are emitted by the carbon nuclei as a result of this bombardment.

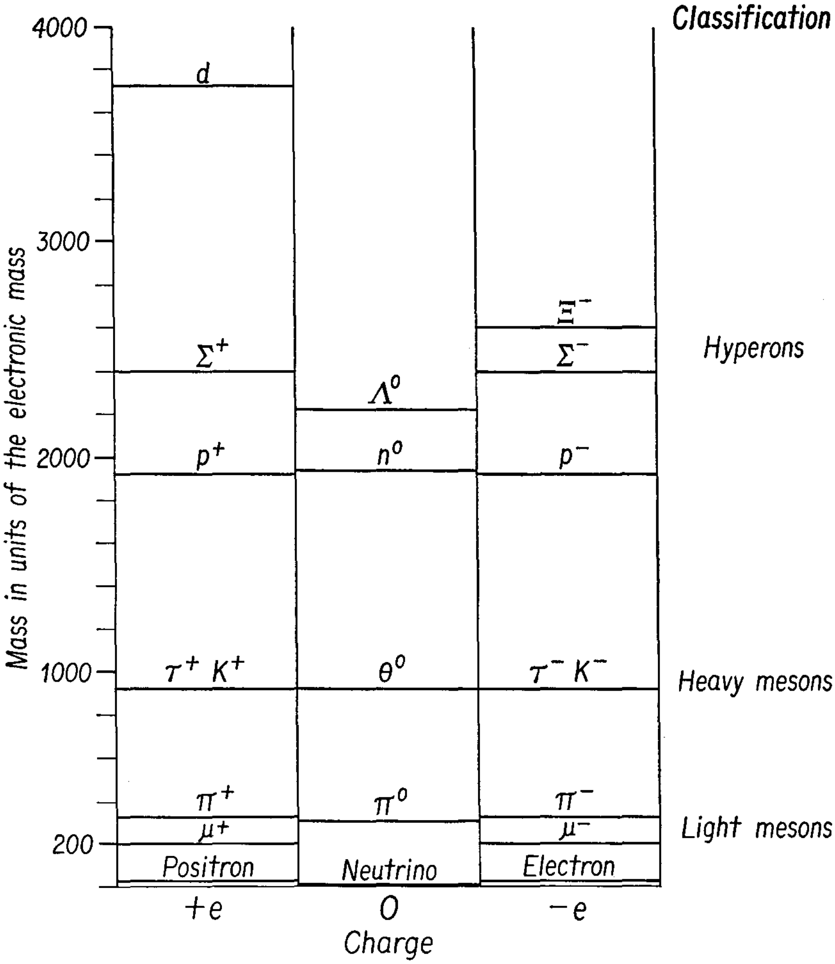


Fig. 46-8 Diagram showing the masses and charges of the so-called elementary particles; the mass of the electron is taken as unity.

The mesons were bent in circular paths by the magnetic field of the cyclotron, the positively charged pions traveling in one direction, say clockwise, the negative ones counterclockwise. They were detected by special photographic plates placed at suitable positions along these paths. The existence of a *neutral pi meson* (π^0 meson) was verified in other experiments.

The pi and mu mesons are only two of the types of particles discovered. Figure 46-8 shows the other new particles known at present (1958). The mass of each particle is shown on the vertical scale for charge $+e$, 0, and $-e$. The unit of mass is the mass of one electron. The particle most recently discovered (1955) is the *negative proton*, that is, a particle with a negative

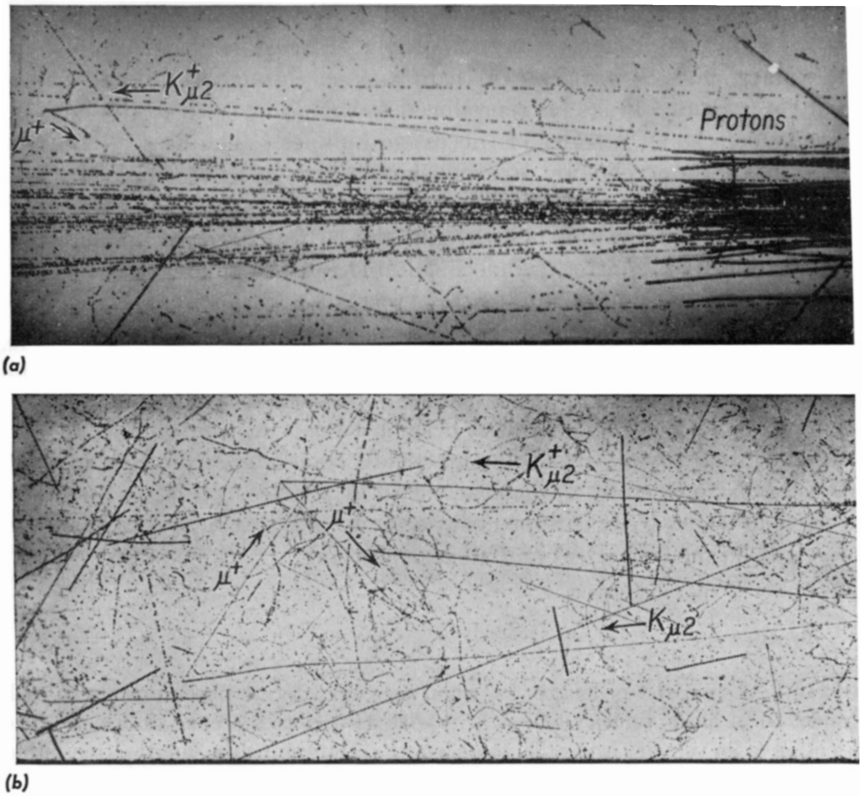


Fig. 46-9 (a) Photographs of tracks of charged particles in a bubble chamber containing propane. The short, heavy tracks are produced by protons which stop in the chamber. The disintegration of a K^+ particle into a μ^+ meson is also shown. (b) Photograph showing the paths of two K^+ particles in a bubble chamber; each of these particles stops in the chamber and decays into a μ^+ meson. (Photographs courtesy of Donald A. Glaser.)

charge and a mass equal to that of the proton. All of these particles, with the exception of the electron, proton, and deuteron, are unstable; that is, they either disintegrate when in the free state or combine with other particles. For example, the neutron is radioactive with a half life of about 12 min, decaying into a proton, electron, and neutrino. The positron combines with an electron and the energy of the two, which is essentially

the rest mass energy $2m_0c^2$, is converted into the energy of one or more, usually two, gamma-ray photons. The charged pi meson disintegrates into a charged mu meson and a neutrino in about 10^{-8} sec, while the mu meson disintegrates into an electron and two neutrinos with an average lifetime of about 2 μsec.

One of the newest devices for studying high-energy reactions is the bubble chamber first developed by Donald A. Glaser in 1953. (See Section 17-7.) Figure 46-9 shows two photographs of tracks made by some of the elementary charged particles in their passage through the liquid of the chamber. Figure 46-9(a) is reproduced to show that each of these tracks actually consists of a series of successive bubbles. The heavy tracks on the right are those made by protons whose paths end in the chamber. An interesting event, the disintegration of a heavy meson, a K^+-meson into a μ^+-meson, is shown in this photograph, and two such events are shown in Figure 46-9(b); the latter are more suitable for measurements. In order to conserve linear momentum in this process, another particle must have been emitted simultaneously with the μ-meson. This second particle must be a neutral particle since it leaves no track in the chamber; it may be a neutrino. Because it disintegrates into two particles, a μ-meson and a neutral particle, this K-meson is designated as a $K\mu_2$ particle.

The discovery of these new particles, sometimes called elementary particles, has opened up a whole new field of physics, coming to be known as particle physics. It is hoped that a study of particle physics will shed new light on nuclear forces and nuclear process.

Problems

46-1. A thin rod of length L_0, when measured by an observer at rest with respect to it, has a velocity of $v = 3c/4$ with respect to a second observer. The direction of its velocity is parallel to its length. Determine the length L measured by the second observer.

46-2. A small particle is in the form of a sphere of radius R_0 when at rest. Determine its shape as seen by an observer if this particle is moving in the x direction with a velocity $c/2$ with respect to this observer.

46-3. Two charged particles are emitted by a substance in opposite directions, each moving with a velocity $v = 0.9c$ with respect to the emitting substance. Determine the velocity of one particle relative to the other.

46-4. The K conversion electron from Cs^{137} produces a sharp line in its beta-ray spectrum for which $Br = 3{,}381$ gausses cm. The binding energy of the K electron is 37.44 kev. (a) Determine the velocity of the electron. (b) Determine the kinetic energy of the electron. (c) Determine the energy of this gamma ray emitted in a transition between the same two levels.

46-5. A gamma-ray photon from Cs^{137}, when incident upon a piece of uranium, ejects photoelectrons from its K shell. These photoelectrons follow

a circular path in a beta-ray spectrometer for which $Br = 3{,}083$ gausses cm. The binding energy of a K electron in uranium is 115.59 kev. Determine (a) the velocity of the photoelectrons, (b) the kinetic energy of the photoelectrons, and (c) the energy of the gamma-ray photons.

46-6. Verify the validity of the Lorentz transformation equations.

46-7. By expanding γ according to the binomial expansion, show that the relativistic expression for the kinetic energy of a moving particle reduces to the classical expression in the limit of low velocities.

46-8. Let us assume that a cyclotron will operate with a fixed-frequency oscillator as long as γ is no more than 1.01. (a) What is the maximum energy to which electrons can be accelerated in a cyclotron? (b) Protons? (c) Alpha particles?

46-9. The cyclotron of the Nobel Institute of Physics has a pole face of 88.5 in. diameter, a maximum flux density of 18,000 gausses, and an oscillator frequency of 8.7×10^6 cycles/sec. Find the maximum energy to which (a) protons, (b) deuterons, and (c) helium nuclei can be accelerated with this cyclotron. Neglect relativistic effects.

46-10. The Oak Ridge reactor has a flux of thermal neutrons of 10^{12} neutrons/cm^2 sec. The activation cross section of Te^{126} is 90 millibarns, in a process in which this isotope absorbs a neutron to become radioactive Te^{127}. If 1 gm of Te^{126} is placed in the Oak Ridge reactor, how many grams of Te^{127} are manufactured in each second?

46-11. (a) Write the nuclear reaction equation for the fission process illustrated in Figure 46-5. (b) Determine the difference in mass between the initial components of the reaction and its final products. (c) Express the result of part (b) in million electron volts. (d) How much energy is released in this reaction? Assume the following atomic masses: Ba = 142.955; Kr = 93.935.

Appendices

Tables 1 and 2 are from *Heath's Logarithmic and Trigonometric Tables,* revised by E. J. Oglesby, and are reprinted by permission of D. C. Heath & Company.

Appendix A

TABLES

Table 1. Four-Place Logarithms of Ordinary Numbers

N	0	1	2	3	4	5	6	7	8	9
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	5611	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
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
N	0	1	2	3	4	5	6	7	8	9

Table 1. Four-Place Logarithms of Ordinary Numbers (Cont.)

N	0	1	2	3	4	5	6	7	8	9
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	9689	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
N	0	1	2	3	4	5	6	7	8	9

Table 2. Four-Place Values of Functions and Radians

Degrees	Radians	Sin	Cos	Tan	Cot	Sec	Csc		
0° 00′	.0000	.0000	1.0000	.0000	——	1.000	——	1.5708	90° 00′
10	029	029	000	029	343.8	000	343.8	679	50
20	058	058	000	058	171.9	000	171.9	650	40
30	.0087	.0087	1.0000	.0087	114.6	1.000	114.6	1.5621	30
40	116	116	.9999	116	85.94	000	85.95	592	20
50	145	145	999	145	68.75	000	68.76	563	10
1° 00′	.0175	.0175	.9998	.0175	57.29	1.000	57.30	1.5533	89° 00′
10	204	204	998	204	49.10	000	49.11	504	50
20	233	233	997	233	42.96	000	42.98	475	40
30	.0262	.0262	.9997	.0262	38.19	1.000	38.20	1.5446	30
40	291	291	996	291	34.37	000	34.38	417	20
50	320	320	995	320	31.24	001	31.26	388	10
2° 00′	.0349	.0349	.9994	.0349	28.64	1.001	28.65	1.5359	88° 00′
10	378	378	993	378	26.43	001	26.45	330	50
20	407	407	992	407	24.54	001	24.56	301	40
30	.0436	.0436	.9990	.0437	22.90	1.001	22.93	1.5272	30
40	465	465	989	466	21.47	001	21.49	243	20
50	495	494	988	495	20.21	001	20.23	213	10
3° 00′	.0524	.0523	.9986	.0524	19.08	1.001	19.11	1.5184	87° 00′
10	553	552	985	553	18.07	002	18.10	155	50
20	582	581	983	582	17.17	002	17.20	126	40
30	.0611	.0610	.9981	.0612	16.35	1.002	16.38	1.5097	30
40	640	640	980	641	15.60	002	15.64	068	20
50	669	669	978	670	14.92	002	14.96	039	10
4° 00′	.0698	.0698	.9976	.0699	14.30	1.002	14.34	1.5010	86° 00′
10	727	727	974	729	13.73	003	13.76	981	50
20	756	756	971	758	13.20	003	13.23	952	40
30	.0785	.0785	.9969	.0787	12.71	1.003	12.75	1.4923	30
40	814	814	967	816	12.25	003	12.29	893	20
50	844	843	964	846	11.83	004	11.87	864	10
5° 00′	.0873	.0872	.9962	.0875	11.43	1.004	11.47	1.4835	85° 00′
10	902	901	959	904	11.06	004	11.10	806	50
20	931	929	957	934	10.71	004	10.76	777	40
30	.0960	.0958	.9954	.0963	10.39	1.005	10.43	1.4748	30
40	989	987	951	992	10.08	005	10.13	719	20
50	.1018	.1016	948	.1022	9.788	005	9.839	690	10
6° 00′	.1047	.1045	.9945	.1051	9.514	1.006	9.567	1.4661	84° 00′
10	076	074	942	080	9.255	006	9.309	632	50
20	105	103	939	110	9.010	006	9.065	603	40
30	.1134	.1132	.9936	.1139	8.777	1.006	8.834	1.4573	30
40	164	161	932	169	8.556	007	8.614	544	20
50	193	190	929	198	8.345	007	8.405	515	10
7° 00′	.1222	.1219	.9925	.1228	8.144	1.008	8.206	1.4486	83° 00′
10	251	248	922	257	7.953	008	8.016	457	50
20	280	276	918	287	7.770	008	7.834	428	40
30	.1309	.1305	.9914	.1317	7.596	1.009	7.661	1.4399	30
40	338	334	911	346	7.429	009	7.496	370	20
50	367	363	907	376	7.269	009	7.337	341	10
8° 00′	.1396	.1392	.9903	.1405	7.115	1.010	7.185	1.4312	82° 00′
10	425	421	899	435	6.968	010	7.040	283	50
20	454	449	894	465	6.827	011	6.900	254	40
30	.1484	.1478	.9890	.1495	6.691	1.011	6.765	1.4224	30
40	513	507	886	524	6.561	012	6.636	195	20
50	542	536	881	554	6.435	012	6.512	166	10
9° 00′	.1571	.1564	.9877	.1584	6.314	1.012	6.392	1.4137	81° 00′
		Cos	Sin	Cot	Tan	Csc	Sec	Radians	Degrees

Table 2. Four-Place Values of Functions and Radians (Cont.)

Degrees	Radians	Sin	Cos	Tan	Cot	Sec	Csc		
9° 00′	.1571	.1564	.9877	.1584	6.314	1.012	6.392	1.4137	**81° 00′**
10	600	593	872	614	197	013	277	108	50
20	629	622	868	644	084	013	166	079	40
30	.1658	.1650	.9863	.1673	5.976	1.014	6.059	1.4050	30
40	687	679	858	703	871	014	5.955	1.4021	20
50	716	708	853	733	769	015	855	992	10
10° 00′	.1745	.1736	.9848	.1763	5.671	1.015	5.759	1.3963	**80° 00′**
10	774	765	843	793	576	016	665	934	50
20	804	794	838	823	485	016	575	904	40
30	.1833	.1822	.9833	.1853	5.396	1.017	5.487	1.3875	30
40	862	851	827	883	309	018	403	846	20
50	891	880	822	914	226	018	320	817	10
11° 00′	.1920	.1908	.9816	.1944	5.145	1.019	5.241	1.3788	**79° 00′**
10	949	937	811	974	066	019	164	759	50
20	978	965	805	.2004	4.989	020	089	730	40
30	.2007	.1994	.9799	.2035	4.915	1.020	5.016	1.3701	30
40	036	.2022	793	065	843	021	4.945	672	20
50	065	051	787	095	773	022	876	643	10
12° 00′	.2094	.2079	.9781	.2126	4.705	1.022	4.810	1.3614	**78° 00′**
10	123	108	775	156	638	023	745	584	50
20	153	136	769	186	574	024	682	555	40
30	.2182	.2164	.9763	.2217	4.511	1.024	4.620	1.3526	30
40	211	193	757	247	449	025	560	497	20
50	240	221	750	278	390	026	502	468	10
13° 00′	.2269	.2250	.9744	.2309	4.331	1.026	4.445	1.3439	**77° 00′**
10	298	278	737	339	275	027	390	410	50
20	327	306	730	370	219	028	336	381	40
30	.2356	.2334	.9724	.2401	4.165	1.028	4.284	1.3352	30
40	385	363	717	432	113	029	232	323	20
50	414	391	710	462	061	030	182	294	10
14° 00′	.2443	.2419	.9703	.2493	4.011	1.031	4.134	1.3265	**76° 00′**
10	473	447	696	524	3.962	031	086	235	50
20	502	476	689	555	914	032	039	206	40
30	.2531	.2504	.9681	.2586	3.867	1.033	3.994	1.3177	30
40	560	532	674	617	821	034	950	148	20
50	589	560	667	648	776	034	906	119	10
15° 00′	.2618	.2588	.9659	.2679	3.732	1.035	3.864	1.3090	**75° 00′**
10	647	616	652	711	689	036	822	061	50
20	676	644	644	742	647	037	782	032	40
30	.2705	.2672	.9636	.2773	3.606	1.038	3.742	1.3003	30
40	734	700	628	805	566	039	703	974	20
50	763	728	621	836	526	039	665	945	10
16° 00′	.2793	.2756	.9613	.2867	3.487	1.040	3.628	1.2915	**74° 00′**
10	822	784	605	899	450	041	592	886	50
20	851	812	596	931	412	042	556	857	40
30	.2880	.2840	.9588	.2962	3.376	1.043	3.521	1.2828	30
40	909	868	580	994	340	044	487	799	20
50	938	896	572	.3026	305	045	453	770	10
17° 00′	.2967	.2924	.9563	.3057	3.271	1.046	3.420	1.2741	**73° 00′**
10	996	952	555	089	237	047	388	712	50
20	.3025	979	546	121	204	048	356	683	40
30	.3054	.3007	.9537	.3153	3.172	1.049	3.326	1.2654	30
40	083	035	528	185	140	049	295	625	20
50	113	062	520	217	108	050	265	595	10
18° 00′	.3142	.3090	.9511	.3249	3.078	1.051	3.236	1.2566	**72° 00′**
		Cos	Sin	Cot	Tan	Csc	Sec	Radians	Degrees

Table 2. Four-Place Values of Functions and Radians (Cont.)

Degrees	Radians	Sin	Cos	Tan	Cot	Sec	Csc		
18° 00′	.3142	.3090	.9511	.3249	3.078	1.051	3.236	1.2566	**72° 00′**
10	171	118	502	281	047	052	207	537	50
20	200	145	492	314	018	053	179	508	40
30	.3229	.3173	.9483	.3346	2.989	1.054	3.152	1.2479	30
40	258	201	474	378	960	056	124	450	20
50	287	228	465	411	932	057	098	421	10
19° 00′	.3316	.3256	.9455	.3443	2.904	1.058	3.072	1.2392	**71° 00′**
10	345	283	446	476	877	059	046	363	50
20	374	311	436	508	850	060	021	334	40
30	.3403	.3338	.9426	.3541	2.824	1.061	2.996	1.2305	30
40	432	365	417	574	798	062	971	275	20
50	462	393	407	607	773	063	947	246	10
20° 00′	.3491	.3420	.9397	.3640	2.747	1.064	2.924	1.2217	**70° 00′**
10	520	448	387	673	723	065	901	188	50
20	549	475	377	706	699	066	878	159	40
30	.3578	.3502	.9367	.3739	2.675	1.068	2.855	1.2130	30
40	607	529	356	772	651	069	833	101	20
50	636	557	346	805	628	070	812	072	10
21° 00′	.3665	.3584	.9336	.3839	2.605	1.071	2.790	1.2043	**69° 00′**
10	694	611	325	872	583	072	769	1.2014	50
20	723	638	315	906	560	074	749	985	40
30	.3752	.3665	.9304	.3939	2.539	1.075	2.729	1.1956	30
40	782	692	293	973	517	076	709	926	20
50	811	719	283	.4006	496	077	689	897	10
22° 00′	.3840	.3746	.9272	.4040	2.475	1.079	2.669	1.1868	**68° 00′**
10	869	773	261	074	455	080	650	839	50
20	898	800	250	108	434	081	632	810	40
30	.3927	.3827	.9239	.4142	2.414	1.082	2.613	1.1781	30
40	956	854	228	176	394	084	595	752	20
50	985	881	216	210	375	085	577	723	10
23° 00′	.4014	.3907	.9205	.4245	2.356	1.086	2.559	1.1694	**67° 00′**
10	043	934	194	279	337	088	542	665	50
20	072	961	182	314	318	089	525	636	40
30	.4102	.3987	.9171	.4348	2.300	1.090	2.508	1.1606	30
40	131	.4014	159	383	282	092	491	577	20
50	160	041	147	417	264	093	475	548	10
24° 00′	.4189	.4067	.9135	.4452	2.246	1.095	2.459	1.1519	**66° 00′**
10	218	094	124	487	229	096	443	490	50
20	247	120	112	522	211	097	427	461	40
30	.4276	.4147	.9100	.4557	2.194	1.099	2.411	1.1432	30
40	305	173	088	592	177	100	396	403	20
50	334	200	075	628	161	102	381	374	10
25° 00′	.4363	.4226	.9063	.4663	2.145	1.103	2.366	1.1345	**65° 00′**
10	392	253	051	699	128	105	352	316	50
20	422	279	038	734	112	106	337	286	40
30	.4451	.4305	.9026	.4770	2.097	1.108	2.323	1.1257	30
40	480	331	013	806	081	109	309	228	20
50	509	358	001	841	066	111	295	199	10
26° 00′	.4538	.4384	.8988	.4877	2.050	1.113	2.281	1.1170	**64° 00′**
10	567	410	975	913	035	114	268	141	50
20	596	436	962	950	020	116	254	112	40
30	.4625	.4462	.8949	.4986	2.006	1.117	2.241	1.1083	30
40	654	488	936	.5022	1.991	119	228	054	20
50	683	514	923	059	977	121	215	1.1025	10
27° 00′	.4712	.4540	.8910	.5095	1.963	1.122	2.203	1.0996	**63° 00′**
		Cos	Sin	Cot	Tan	Csc	Sec	Radians	Degrees

Table 2. Four-Place Values of Functions and Radians (Cont.)

Degrees	Radians	Sin	Cos	Tan	Cot	Sec	Csc		
27° 00′	.4712	.4540	.8910	.5095	1.963	1.122	2.203	1.0996	**63° 00′**
10	741	566	897	132	949	124	190	966	50
20	771	592	884	169	935	126	178	937	40
30	.4800	.4617	.8870	.5206	1.921	1.127	2.166	1.0908	30
40	829	643	857	243	907	129	154	879	20
50	858	669	843	280	894	131	142	850	10
28° 00′	.4887	.4695	.8829	.5317	1.881	1.133	2.130	1.0821	**62° 00′**
10	916	720	816	354	868	134	118	792	50
20	945	746	802	392	855	136	107	763	40
30	.4974	.4772	.8788	.5430	1.842	1.138	2.096	1.0734	30
40	.5003	797	774	467	829	140	085	705	20
50	032	823	760	505	816	142	074	676	10
29° 00′	.5061	.4848	.8746	.5543	1.804	1.143	2.063	1.0647	**61° 00′**
10	091	874	732	581	792	145	052	617	50
20	120	899	718	619	780	147	041	588	40
30	.5149	.4924	.8704	.5658	1.767	1.149	2.031	1.0559	30
40	178	950	689	696	756	151	020	530	20
50	207	975	675	735	744	153	010	501	10
30° 00′	.5236	.5000	.8660	.5774	1.732	1.155	2.000	1.0472	**60° 00′**
10	265	025	646	812	720	157	1.990	443	50
20	294	050	631	851	709	159	980	414	40
30	.5323	.5075	.8616	.5890	1.698	1.161	1.970	1.0385	30
40	352	100	601	930	686	163	961	356	20
50	381	125	587	969	.675	165	951	327	10
31° 00′	.5411	.5150	.8572	.6009	1.664	1.167	1.942	1.0297	**59° 00′**
10	440	175	557	048	653	169	932	268	50
20	469	200	542	088	643	171	923	239	40
30	.5498	.5225	.8526	.6128	1.632	1.173	1.914	1.0210	30
40	527	250	511	168	621	175	905	181	20
50	556	275	496	208	611	177	896	152	10
32° 00′	.5585	.5299	.8480	.6249	1.600	1.179	1.887	1.0123	**58° 00′**
10	614	324	465	289	590	181	878	094	50
20	643	348	450	330	580	184	870	065	40
30	.5672	.5373	.8434	.6371	1.570	1.186	1.861	1.0036	30
40	701	398	418	412	560	188	853	1.0007	20
50	730	422	403	453	550	190	844	977	10
33° 00′	.5760	.5446	.8387	.6494	1.540	1.192	1.836	.9948	**57° 00′**
10	789	471	371	536	530	195	828	919	50
20	818	495	355	577	520	197	820	890	40
30	.5847	.5519	.8339	.6619	1.511	1.199	1.812	.9861	30
40	876	544	323	661	501	202	804	832	20
50	905	568	307	703	1.492	204	796	803	10
34° 00′	.5934	.5592	.8290	.6745	1.483	1.206	1.788	.9774	**56° 00′**
10	963	616	274	787	473	209	781	745	50
20	992	640	258	830	464	211	773	716	40
30	.6021	.5664	.8241	.6873	1.455	1.213	1.766	.9687	30
40	050	688	225	916	446	216	758	657	20
50	080	712	208	959	437	218	751	628	10
35° 00′	.6109	.5736	.8192	.7002	1.428	1.221	1.743	.9599	**55° 00′**
10	138	760	175	046	419	223	736	570	50
20	167	783	158	089	411	226	729	541	40
30	.6196	.5807	.8141	.7133	1.402	1.228	1.722	.9512	30
40	225	831	124	177	.393	231	715	483	20
50	254	854	107	221	385	233	708	454	10
36° 00′	.6283	.5878	.8090	.7265	1.376	1.236	1.701	.9425	**54° 00′**
		Cos	Sin	Cot	Tan	Csc	Sec	Radians	Degrees

Table 2. Four-Place Values of Functions and Radians (Cont.)

Degrees	Radians	Sin	Cos	Tan	Cot	Sec	Csc		
36° 00′	.6283	.5878	.8090	.7265	1.376	1.236	1.701	.9425	**54° 00′**
10	312	901	073	310	368	239	695	396	50
20	341	925	056	355	360	241	688	367	40
30	.6370	.5948	.8039	.7400	1.351	1.244	1.681	.9338	30
40	400	972	021	445	343	247	675	308	20
50	429	995	004	490	335	249	668	279	10
37° 00′	.6458	.6018	.7986	.7536	1.327	1.252	1.662	.9250	**53° 00′**
10	487	041	969	581	319	255	655	221	50
20	516	065	951	627	311	258	649	192	40
30	.6545	.6088	.7934	.7673	1.303	1.260	1.643	.9163	30
40	574	111	916	720	295	263	636	134	20
50	603	134	898	766	288	266	630	105	10
38° 00′	.6632	.6157	.7880	.7813	1.280	1.269	1.624	.9076	**52° 00′**
10	661	180	862	860	272	272	618	047	50
20	690	202	844	907	265	275	612	.9018	40
30	.6720	.6225	.7826	.7954	1.257	1.278	1.606	.8988	30
40	749	248	808	.8002	250	281	601	959	20
50	778	271	790	050	242	284	595	930	10
39° 00′	.6807	.6293	.7771	.8098	1.235	1.287	1.589	.8901	**51° 00′**
10	836	316	753	146	228	290	583	872	50
20	865	338	735	195	220	293	578	843	40
30	.6894	.6361	.7716	.8243	1.213	1.296	1.572	.8814	30
40	923	383	698	292	206	299	567	785	20
50	952	406	679	342	199	302	561	756	10
40° 00′	.6981	.6428	.7660	.8391	1.192	1.305	1.556	.8727	**50° 00′**
10	.7010	450	642	441	185	309	550	698	50
20	039	472	623	491	178	312	545	668	40
30	.7069	.6494	.7604	.8541	1.171	1.315	1.540	.8639	30
40	098	517	585	591	164	318	535	610	20
50	127	539	566	642	157	322	529	581	10
41° 00′	.7156	.6561	.7547	.8693	1.150	1.325	1.524	.8552	**49° 00′**
10	185	583	528	744	144	328	519	523	50
20	214	604	509	796	137	332	514	494	40
30	.7243	.6626	.7490	.8847	1.130	1.335	1.509	.8465	30
40	272	648	470	899	124	339	504	436	20
50	301	670	451	952	117	342	499	407	10
42° 00′	.7330	.6691	.7431	.9004	1.111	1.346	1.494	.8378	**48° 00′**
10	359	713	412	057	104	349	490	348	50
20	389	734	392	110	098	353	485	319	40
30	.7418	.6756	.7373	.9163	1.091	1.356	1.480	.8290	30
40	447	777	353	217	085	360	476	261	20
50	476	799	333	271	079	364	471	232	10
43° 00′	.7505	.6820	.7314	.9325	1.072	1.367	1.466	.8203	**47° 00′**
10	534	841	294	380	066	371	462	174	50
20	563	862	274	435	060	375	457	145	40
30	.7592	.6884	.7254	.9490	1.054	1.379	1.453	.8116	30
40	621	905	234	545	048	382	448	087	20
50	650	926	214	601	042	386	444	058	10
44° 00′	.7679	.6947	.7193	.9657	1.036	1.390	1.440	.8029	**46° 00′**
10	709	967	173	713	030	394	435	999	50
20	738	988	153	770	024	398	431	970	40
30	.7767	.7009	.7133	.9827	1.018	1.402	1.427	.7941	30
40	796	030	112	884	012	406	423	912	20
50	825	050	092	942	006	410	418	883	10
45° 00′	.7854	.7071	.7071	1.000	1.000	1.414	1.414	.7854	**45° 00′**
		Cos	Sin	Cot	Tan	Csc	Sec	Radians	Degrees

Table 3. Isotopic Masses

Mass No.	Atomic No.	Element	Atomic Mass
1	0	n	1.008987
1	1	H	1.008145
2	1	H	2.014741
3	1	H	3.016997
3	2	He	3.016977
4	2	He	4.003879
5	2	He	5.0137
5	3	Li	5.0136
6	2	He	6.020833
6	3	Li	6.01697
7	3	Li	7.01822
7	4	Be	7.01916
8	3	Li	8.02502
8	4	Be	8.00785
8	5	B	8.0264
9	4	Be	9.01503
9	5	B	9.01620
10	4	Be	10.01677
10	5	B	10.016110
10	6	C	10.0206
11	5	B	11.012811
11	6	C	11.01495
12	6	C	12.003844
12	7	N	12.0227
13	6	C	13.007505
13	7	N	13.00988
14	6	C	14.00767
14	7	N	14.007550
15	6	C	15.0143
15	7	N	15.004902
15	8	O	15.0078
16	7	N	16.0109
16	8	O	16.000000
17	7	N	17.0139
17	8	O	17.004533
17	9	F	17.007505
18	8	O	18.004883
18	9	F	18.006651
19	8	O	19.0091
19	9	F	19.004444
19	10	Ne	19.007952

Table 3. Isotopic Masses (Cont.)

Mass No.	Atomic No.	Element	Atomic Mass
20	9	F	20.006350
20	10	Ne	19.998772
21	10	Ne	21.000504
21	11	Na	21.004286
22	10	Ne	21.998382
22	11	Na	22.001409
23	10	Ne	23.001768
23	11	Na	22.997055
23	12	Mg	23.001453
24	11	Na	23.998568
24	12	Mg	23.992628
25	12	Mg	24.993745
26	12	Mg	25.990802
27	12	Mg	26.992876
27	13	Al	26.990109
28	13	Al	27.990760
28	14	Si	27.985825
29	14	Si	28.985705
30	14	Si	29.983307
31	14	Si	30.985140
31	15	P	30.983619
32	15	P	31.984016
32	16	S	31.982274
33	15	P	32.982166
33	16	S	32.981941
34	16	S	33.978709
35	17	Cl	34.980064
36	18	A	35.97926
37	17	Cl	36.977675
38	18	A	37.97491
39	19	K	38.97606
40	18	A	39.975148
40	20	Ca	39.97545
41	19	K	40.97490
42	20	Ca	41.97216
43	20	Ca	42.97251
44	20	Ca	43.96924
45	21	Sc	44.97010
46	22	Ti	45.96697
47	22	Ti	46.96668
48	20	Ca	47.96778

Table 3. Isotopic Masses (Cont.)

Mass No.	Atomic No.	Element	Atomic Mass
48	22	Ti	47.96317
49	22	Ti	48.96358
50	22	Ti	49.96077
50	24	Cr	49.96210
51	23	V	50.96052
52	24	Cr	51.95707
53	24	Cr	52.95772
54	24	Cr	53.9563
54	26	Fe	53.95704
55	25	Mn	54.95581
56	26	Fe	55.95272
57	26	Fe	56.95359
58	26	Fe	57.9520
58	28	Ni	57.95345
59	27	Co	58.95182
60	27	Co	59.95250
60	28	Ni	59.94901
61	28	Ni	60.94907
62	28	Ni	61.94681
63	29	Cu	62.94926
64	28	Ni	63.94755
64	30	Zn	63.94955
65	29	Cu	64.94835
66	30	Zn	65.94722
67	30	Zn	66.94815
68	30	Zn	67.94686
70	30	Zn	69.94779
70	32	Ge	69.9447
74	32	Ge	73.9426
74	34	Se	73.9439
75	33	As	74.9432
76	32	Ge	75.9433
79	35	Br	78.944
81	35	Br	80.943
82	36	Kr	81.938
84	36	Kr	83.938
85	37	Rb	84.931
86	38	Sr	85.93533
87	37	Rb	86.9295
88	38	Sr	87.93374
94	42	Mo	93.9343

Table 3. Isotopic Masses (Cont.)

Mass No.	Atomic No.	Element	Atomic Mass
98	42	Mo	97.93610
102	46	Pd	101.9375
104	46	Pd	103.93655
105	46	Pd	104.9384
106	46	Pd	105.9368
106	48	Cd	105.93984
108	46	Pd	107.93801
108	48	Cd	107.93860
110	46	Pd	109.93965
110	48	Cd	109.93857
111	48	Cd	110.93978
112	48	Cd	111.93885
113	48	Cd	112.94061
113	49	In	112.94045
114	48	Cd	113.93997
115	49	In	114.94040
115	50	Sn	114.94014
116	48	Cd	115.94202
116	50	Sn	115.93927
117	50	Sn	116.94052
118	50	Sn	117.93978
119	50	Sn	118.94122
120	50	Sn	119.94059
120	52	Te	119.94288
122	50	Sn	121.94249
122	52	Te	121.94193
123	52	Te	122.94368
124	50	Sn	123.94490
124	52	Te	123.94278
124	54	Xe	123.94578
125	52	Te	124.94460
126	52	Te	125.94420
126	54	Xe	125.94476
127	53	I	126.94528
128	52	Te	127.94649
128	54	Xe	127.94446
129	54	Xe	153.94601
130	52	Te	129.94853
130	54	Xe	129.94501
131	54	Xe	130.94673
132	54	Xe	131.94615

Table 3. Isotopic Masses (Cont.)

Mass No.	Atomic No.	Element	Atomic Mass
134	54	Xe	133.94803
136	54	Xe	135.95046
136	56	Ba	135.9488
137	56	Ba	136.9502
138	56	Ba	137.9498
140	58	Ce	139.9489
141	59	Pr	140.9514
142	58	Ce	141.9537
144	60	Nd	143.9560
150	60	Nd	149.9687
176	72	Hf	175.9923
178	72	Hf	177.9936
180	72	Hf	180.0029
181	73	Ta	181.0031
182	74	W	182.0033
183	74	W	183.0059
184	74	W	184.0052
194	78	Pt	194.0256
196	78	Pt	196.02744
205	82	Pb	205.04559
206	81	Tl	206.04702
206	82	Pb	206.04519
207	81	Tl	207.04934
207	82	Pb	207.04725
208	81	Tl	208.05290
208	82	Pb	208.04754
208	83	Bi	208.04968
209	81	Tl	209.05778
209	82	Pb	209.05398
209	83	Bi	209.05325
209	84	Po	209.05496
210	81	Tl	210.06264
210	82	Pb	210.05622
210	83	Bi	210.05614
210	84	Po	210.05488
211	82	Pb	211.06196
211	83	Bi	211.06047
211	84	Po	211.05927
212	82	Pb	212.06487
212	83	Bi	212.06345
212	84	Po	212.06094

Table 3. Isotopic Masses (Cont.)

Mass No.	Atomic No.	Element	Atomic Mass
212	85	At	212.06079
213	83	Bi	213.06824
213	83	Po	213.06696
214	82	Pb	214.07362
214	83	Bi	214.07225
214	84	Po	214.06852
214	85	At	214.06955
215	84	Po	215.07392
215	85	At	215.07313
216	84	Po	216.07617
216	85	At	216.07586
216	86	Em	216.07358
217	85	At	217.07979
217	86	Em	217.07939
218	84	Po	218.08407
218	85	At	218.08369
218	86	Em	218.08017
218	87	Fr	218.08108
219	86	Em	219.08527
219	87	Fr	219.08501
220	86	Em	220.08693
220	87	Fr	220.08706
220	88	Ra	220.08567
221	87	Fr	221.09057
221	88	Ra	221.09060
222	86	Em	222.09397
222	88	Ra	222.09116
222	89	Ac	222.09342
223	87	Fr	223.09697
223	88	Ra	223.09559
223	89	Ac	223.09615
224	88	Ra	224.09703
224	89	Ac	224.09769
224	90	Th	224.09743
225	88	Ra	225.10102
225	89	Ac	225.10081
225	90	Th	225.10170
226	88	Ra	226.10309
226	90	Th	226.10193
226	91	Pa	226.10494
227	88	Ra	227.10723

Table 3. Isotopic Masses (Cont.)

Mass No.	Atomic No.	Element	Atomic Mass
227	89	Ac	227.10666
227	90	Th	227.10642
227	91	Pa	227.10710
228	88	Ra	228.11005
228	89	Ac	228.11005
228	90	Th	228.10685
228	91	Pa	228.10823
228	92	U	228.10863
229	90	Th	229.11021
229	91	Pa	229.11088
229	92	U	229.11258
230	90	Th	230.11206
230	91	Pa	230.11441
230	92	U	230.11222
231	90	Th	231.11628
231	91	Pa	231.11607
231	93	Np	231.11776
232	90	Th	232.11852
232	91	Pa	232.11768
232	92	U	232.11650
232	94	Pu	232.11973
233	90	Th	233.12198
233	91	Pa	233.12027
233	92	U	233.11937
234	90	Th	234.12394
234	91	Pa	234.12281
234	92	U	234.12115
234	94	Pu	234.12269
235	92	U	235.12517
236	94	Pu	236.12667
237	92	U	237.13010
237	93	Np	237.12932
238	92	U	238.13232
238	93	Np	238.13255
238	94	Pu	238.13106
239	92	U	239.13704
239	93	Np	239.13620
239	94	Pu	239.13494
239	95	Am	239.13568
240	96	Cm	240.13744
241	94	Pu	241.13909

Table 3. Isotopic Masses (Cont.)

Mass No.	Atomic No.	Element	Atomic Mass
241	95	Am	241.13919
242	95	Am	242.14215
242	96	Cm	242.14160

Collins, T. L., A. O. Nier, and W. H. Johnson, Jr., *Phys. Rev.*, 86, 408, 1952.
Duckworth, H. E., C. L. Kegley, J. M. Olson, and G. S. Stanford, *Phys. Rev.*, 83, 1114, 1951.
Halsted, R. E., *Phys. Rev.*, 88, 660, 1952.
Hays, E. E., P. I. Richards, and S. A. Goudsmit, *Phys. Rev.*, 84, 824, 1951.
Hornyak, W. F., T. Lauritsen, P. Morrison, and W. A. Fowler, *Rev. Modern Phys.*, 22, 291, 1950.
Ogata, K., and H. Matsuda, *Phys. Rev.*, 89, p. 27, 1953.
Stern, N. O., *Revs. Modern Phys.*, 21, 316, 1949.

Table 4. Atomic Weights of the Elements

Element	Symbol	Atomic Number	Atomic Weight[a]
Actinium	Ac	89	227
Aluminum	Al	13	26.98
Americium	Am	95	[243]
Antimony	Sb	51	121.76
Argon	A	18	39.944
Arsenic	As	33	74.91
Astatine	At	85	[210]
Barium	Ba	56	137.36
Berkelium	Bk	97	[249]
Beryllium	Be	4	9.013
Bismuth	Bi	83	209.00
Boron	B	5	10.82
Bromine	Br	35	79.916
Cadmium	Cd	48	112.41
Calcium	Ca	20	40.08
Californium	Cf	98	[249]
Carbon	C	6	12.011
Cerium	Ce	58	140.13
Cesium	Cs	55	132.91
Chlorine	Cl	17	35.457
Chromium	Cr	24	52.01
Cobalt	Co	27	58.94
Copper	Cu	29	63.54
Curium	Cm	96	[245]
Dysprosium	Dy	66	162.51
Erbium	Er	68	167.27
Europium	Eu	63	152.0
Fluorine	F	9	19.00
Francium	Fr	87	[223]
Gadolinium	Gd	64	157.26
Gallium	Ga	31	69.72
Germanium	Ge	32	72.60
Gold	Au	79	197.0
Hafnium	Hf	72	178.50
Helium	He	2	4.003
Holmium	Ho	67	164.94
Hydrogen	H	1	1.0080
Indium	In	49	114.82
Iodine	I	53	126.91
Iridium	Ir	77	192.2
Iron	Fe	26	55.85
Krypton	Kr	36	83.80
Lanthanum	La	57	138.92
Lead	Pb	82	207.21
Lithium	Li	3	6.940
Lutetium	Lu	71	174.99
Magnesium	Mg	12	24.32
Manganese	Mn	25	54.94
Mendelevium	Mv	101	[256]
Mercury	Hg	80	200.61

[a] A value given in brackets denotes the mass number of the most stable known isotope. From *Journal of the American Chemical Society*, Vol. 78, p. 3235, 1956.

Table 4. Atomic Weights of the Elements (Cont.)

Element	Symbol	Atomic Number	Atomic Weight[a]
Molybdenum	Mo	42	95.95
Neodymium	Nd	60	144.27
Neon	Ne	10	20.183
Neptunium	Np	93	[237]
Nickel	Ni	28	58.71
Niobium (Columbium)	Nb	41	92.91
Nitrogen	N	7	14.008
Osmium	Os	76	190.2
Oxygen	O	8	**16**
Palladium	Pd	46	106.4
Phosphorus	P	15	30.975
Platinum	Pt	78	195.09
Plutonium	Pu	94	[242]
Polonium	Po	84	210
Potassium	K	19	39.100
Praseodymium	Pr	59	140.92
Promethium	Pm	61	[145]
Protactinium	Pa	91	231
Radium	Ra	88	226.05
Radon	Rn	86	222
Rhenium	Re	75	186.22
Rhodium	Rh	45	102.91
Rubidium	Rb	37	85.48
Ruthenium	Ru	44	101.1
Samarium	Sm	62	150.35
Scandium	Sc	21	44.96
Selenium	Se	34	78.96
Silicon	Si	14	28.09
Silver	Ag	47	107.880
Sodium	Na	11	22.991
Strontium	Sr	38	87.63
Sulfur	S	16	32.066
Tantalum	Ta	73	180.95
Technetium	Tc	43	[99]
Tellurium	Te	52	127.61
Terbium	Tb	65	158.93
Thallium	Tl	81	204.39
Thorium	Th	90	232.05
Thulium	Tm	69	168.94
Tin	Sn	50	118.70
Titanium	Ti	22	47.90
Tungsten	W	74	183.86
Uranium	U	92	238.07
Vanadium	V	23	50.95
Xenon	Xe	54	131.30
Ytterbium	Yb	70	173.04
Yttrium	Y	39	88.92
Zinc	Zn	30	65.38
Zirconium	Zr	40	91.22

[a] A value given in brackets denotes the mass number of the most stable known isotope. From *Journal of the American Chemical Society*, Vol. 78, p. 3235, 1956.

Table 5. Distribution of Electrons in the Atoms

X-Ray Notation		K	L		M			N			
Quantum Numbers n, l		1,0	2,0	2,1	3,0	3,1	3,2	4,0	4,1	4,2	4,3
Element	Atomic Number Z										
H	1	1									
He	2	2									
Li	3	2	1								
Be	4	2	2								
B	5	2	2	1							
C	6	2	2	2							
N	7	2	2	3							
O	8	2	2	4							
F	9	2	2	5							
Ne	10	2	2	6							
Na	11				1						
Mg	12				2						
Al	13		Neon		2	1					
Si	14		Configuration		2	2					
P	15		10 electron		2	3					
S	16		core		2	4					
Cl	17				2	5					
A	18				2	6					
K	19							1			
Ca	20							2			
Sc	21						1	2			
Ti	22						2	2			
V	23						3	2			
Cr	24						5	1			
Mn	25						5	2			
Fe	26			Argon			6	2			
Co	27			Configuration			7	2			
Ni	28			18 electron			8	2			
Cu	29			core			10	1			
Zn	30						10	2			
Ga	31						10	2	1		
Ge	32						10	2	2		
As	33						10	2	3		
Se	34						10	2	4		
Br	35						10	2	5		
Kr	36						10	2	6		

Table 5. Distribution of Electrons in the Atoms (Cont.)

X-Ray Notation		K L M	N			O			P			Q	
Quantum Numbers n, l		1 2 3	4,0 4,1	4,2	4,3	5,0	5,1	5,2	6,0	6,1	6,2	7,0	7,1
Element	Atomic Number Z												
Rb	37					1							
Sr	38					2							
Y	39	Krypton		1		2							
Zr	40	Configuration		2		2							
Nb	41	36 electron		4		1							
Mo	42	core		5		1							
Ma	43			6		1							
Ru	44			7		1							
Rh	45			8		1							
Pd	46			10									
Ag	47					1							
Cd	48					2							
In	49	Palladium				2	1						
Sn	50	Configuration				2	2						
Sb	51	46 electron				2	3						
Te	52	core				2	4						
I	53					2	5						
Xe	54					2	6						
Cs	55	Xenon Configuration							1				
Ba	56	54 electron core							2				
La	57					2	6	1	2				
Ce	58				1	2	6	1	2				
Pr	59				2	2	6	1	2				
Nd	60				3	2	6	1	2				
Pm	61				4	2	6	1	2				
Sm	62	Shells			5	2	6	1	2				
Eu	63	1,0 to 4,2			6	2	6	1	2				
Gd	64	contain			7	2	6	1	2				
Tb	65	46 electrons			8	2	6	1	2				
Dy	66				9	2	6	1	2				
Ho	67				10	2	6	1	2				
Er	68				11	2	6	1	2				
Tm	69				13	2	6	0	2				
Yb	70				14	2	6	0	2				
Lu	71				14	2	6	1	2				

Table 5. Distribution of Electrons in the Atoms (Cont.)

X-Ray Notation		K L M N	O			P			Q	
Quantum Numbers n, l		1 2 3 4	5,0 5,1	5,2	5,3	6,0	6,1	6,2	7,0	7,1
Element	Atomic Number Z									
Hf	72			2		2				
Ta	73			3		2				
W	74			4		2				
Re	75			5		2				
Os	76			6		2				
Ir	77			7		2				
Pt	78	Shells		9		1				
Au	79	1,0 to 5,1		10		1				
Hg	80	contain		10		2				
Tl	81	68 electrons		10		2	1			
Pb	82			10		2	2			
Bi	83			10		2	3			
Po	84			10		2	4			
At	85			10		2	5			
Rn	86			10		2	6			
Fr	87	Radon Configuration							1	
Ra	88	86 electron core							2	
Ac	89					2	6	1	2	
Th	90				1	2	6	1	2	
Pa	91				2	2	6	1	2	
U	92				3	2	6	1	2	
Np	93				4	2	6	1	2	
Pu	94				5	2	6	1	2	
Am	95				6	2	6	1	2	
Cm	96				7	2	6	1	2	
Bk	97				8	2	6	1	2	
Cf	98				9	2	6	1	2	
E	99				10	2	6	1	2	
Fm	100				11	2	6	1	2	
Mv	101				12	2	6	1	2	

Table 6. Periodic Table of the Elements

	I	II	III	IV	V	VI	VII	VIII
1	1 H 1.0080							2 He 4.003
2	3 Li 6.940	4 Be 9.013	5 B 10.82	6 C 12.011	7 N 14.008	8 O 16	9 F 19.00	10 Ne 20.183
3	11 Na 22.991	12 Mg 24.32	13 Al 26.98	14 Si 28.09	15 P 30.975	16 S 32.066	17 Cl 35.457	18 A 39.944
4	19 K 39.100	20 Ca 40.08	21 Sc 44.96	22 Ti 47.90	23 V 50.95	24 Cr 52.01	25 Mn 54.94	26 Fe 27 Co 28 Ni 55.85 58.94 58.71
	29 Cu 63.54	30 Zn 65.38	31 Ga 69.72	32 Ge 72.60	33 As 74.91	34 Se 78.96	35 Br 79.916	36 Kr 83.80
5	37 Rb 85.48	38 Sr 87.63	39 Y 88.92	40 Zr 91.22	41 Nb 92.91	42 Mo 95.95	43 Tc [99]	44 Ru 45 Rh 46 Pd 101.1 102.91 106.4
	47 Ag 107.880	48 Cd 112.41	49 In 114.82	50 Sn 118.70	51 Sb 121.76	52 Te 127.61	53 I 126.91	54 Xe 131.30
6	55 Cs 132.91	56 Ba 137.36	57–71 Rare Earths*	72 Hf 178.50	73 Ta 180.95	74 W 183.86	75 Re 186.22	76 Os 77 Ir 78 Pt 190.2 192.2 195.09
	79 Au 197.0	80 Hg 200.61	81 Tl 204.39	82 Pb 207.21	83 Bi 209.00	84 Po 210	85 At [210]	86 Rn 222
7	87 Fr [223]	88 Ra 226.05	89–102 Actinide† Series					

*Rare Earth or Lanthanide Series.

57 La 138.92	58 Ce 140.13	59 Pr 140.92	60 Nd 144.27	61 Pm [145]
62 Sm 150.35	63 Eu 152.0	64 Gd 157.26	65 Tb 158.93	66 Dy 162.51
67 Ho 164.94	68 Er 167.27	69 Tm 168.94	70 Yb 173.04	71 Lu 174.99

† Actinide Series.

89 Ac 227	90 Th 232.05	91 Pa 231	92 U 238.07
93 Np [237]	94 Pu [242]	95 Am [243]	96 Cm [245]
97 Bk [249]	98 Cf [249]	99 E (Einsteinium)	100 Fm (Fermium)
101 Mv [256]	102 (Nobelium)		

Appendix B

THE GREEK ALPHABET

Lower-case Letter	*Capital Letter*	*Name of Letter*
α	A	alpha
β	B	beta
γ	Γ	gamma
δ	Δ	delta
ϵ	E	epsilon
ζ	Z	zeta
η	H	eta
θ	Θ	theta
ι	I	iota
κ	K	kappa
λ	Λ	lambda
μ	M	mu
ν	N	nu
ξ	Ξ	xi
o	O	omicron
π	Π	pi
ρ	P	rho
σ, s	Σ	sigma
τ	T	tau
υ	Υ	upsilon
ϕ	Φ	phi
χ	X	chi
ψ	Ψ	psi
ω	Ω	omega

Appendix C

Answers to Odd-Numbered Problems

Chapter 1

1–1. No answer.
1–3. (a) $(5, 53°)$; (b) $(7.1, 315°)$; (c) $(13.4, 153°25')$.
1–5. (2.5 m, 4.3 m).
1–7. $R = 10.7$ m; $(x, R) = 73°45'$; $(y, R) = 62°15'$; $(z, R) = 32°55'$.
1–9. $\mathbf{A} + \mathbf{B}$: $(7.07, 171°50')$; $\mathbf{A} - \mathbf{B}$: $(9.06, 276°20')$.
1–11. (a) $(4.51, 145°45')$; (b) $(24.1, 357°5')$.
1–13. No answer.
1–15. 5.04, $79°35'$.

Chapter 2

2–1. (a) 40 mi/hr; (b) 58.7 ft/sec.
2–3. (a) 7.75 mi/hr; (b) 3.9 min.
2–5. 253 mi/hr, $9°5'$ E of N.
2–7. 4.9 ft/sec^2.
2–9. 4.28 ft/sec^2.
2–11. (a) 5.71 ft/sec^2; (b) 20.5 sec.
2–13. (a) 71.6 ft/sec; (b) 75.5 ft/sec.
2–15. (a) 4 sec; (b) 120 ft from base; (c) 132 ft/sec, $\theta = -76°50'$.
2–17. 402 cm/sec, 598 cm/sec.
2–19. (a) 7.95 km; (b) 30,000 cm/sec, $\theta = -30°$.
2–21. (a) 0.24 ft/sec^2; (b) 4,950 ft.
2–23. (a) 6.67 m/sec; (b) 5.2 m/sec; (c) 1.38 m.
2–25. (a) 5.76 ft; (b) 57.6 ft/sec, $\theta = 19°30'$.
2–27. No answer.
2–29. (a) 38 m/sec, $\theta = 71°35'$; (b) 12 m/sec^2 (y-direction).
2–31. (a) 1.63 sec; (b) 56.4 ft; (c) 143.5 ft.
2–33. (a) -32 ft/sec; (b) 0.
2–35. No answer.
2–37. No answer.
2–39. No answer.

Chapter 3

3–1. 15 lb.
3–3. (a) 69.3 lb; (b) 138.5 lb.
3–5. (a) 230 lb; (b) 230 lb.
3–7. (a) 575 lb; (b) 1,150 lb.
3–9. 175 lb.
3–11. (a) 90 lb; (b) 150 lb.
3–13. (a) 49 lb; (b) 130 lb; (c) 88 lb; (d) 62 lb.
3–15. (a) 8 lb; (b) 0.061; (c) 65.9 lb.

Chapter 4

4–1. 6 lb ft.
4–3. (a) 1,225 lb; (b) 285 lb; (c) 980 lb.
4–5. 2.5 ft from 60 lb wt.
4–7. (a) 1,200 lb; (b) 2,000 lb.
4–9. (a) 115 lb; (b) 693 lb.
4–11. 0.56.
4–13. $0.67a$, $0.29a$.
4–15. 30.5 lb, 75°20′ with horizontal.
4–17. 2.94 in.
4–19. No answer.
4–21. (a) 4 ft; (b) 120 lb.
4–23. No answer.

Chapter 5

5–1. 6,000 dynes.
5–3. 428 lb.
5–5. 50,000 dynes.
5–7. 84,500 lb.
5–9. (a) 281 lb; (b) 2,081 lb.
5–11. (a) 30,000 dynes; (b) 275,000 dynes.
5–13. (a) 8 ft/sec^2; (b) 18 lb.
5–15. (a) 1.96 m/sec^2; (b) 23.5 nt.
5–17. (a) 150 lb; (b) 46.3 ft/sec.
5–19. 0.43.
5–21. 10 ft.
5–23. No answer.
5–25. 2.6×10^5 dynes.
5–27. 161 nt.
5–29. 15.8 sec.
5–31. (a) 20 lb; (b) 40 lb.
5–33. (a) 6.4 ft/sec^2; (b) 6.4 ft/sec^2; (c) 7.2 ft/sec^2.
5–35. 38.4 ft/sec.

Chapter 6

6–1. (a) 6π rad/sec; (b) 37.7 ft/sec; (c) 10,800°.
6–3. π rad/sec^2; (b) 1,100 rev.
6–5. No answer.
6–7. No answer.
6–9. (a) 3,125 dynes; (b) 20.8 cm/sec^2.
6–11. 2.7×10^{-3} m/sec; (b) 5.83×10^{24} kg.
6–13. (a) 6°; (b) 0.13.
6–15. (a) 4.95 rad/sec; (b) 1.22×10^5 dynes.
6–17. (a) 4,033 ft; (b) 1,260 lb.
6–19. (a) 2.97×10^4 m/sec; (b) 5.91×10^{-3} m/sec^2; (c) 3.53×10^{22} nt.
6–21. (a) 20 mi; (b) 2 mi
6–23. 1.97×10^{30} kg.
6–25. No answer.
6–27. 800 lb.

Chapter 7

7–1. (a) 600 ft lb; (b) 0.33.
7–3. (a) 1,120 ft lb; (b) 1,120 ft lb.
7–5. (a) 1,800 ft lb; (b) 1,200 ft lb; (c) 600 ft lb.
7–7. (a) 1,800 ft lb; (b) 960 ft lb; (c) 840 ft lb; (d) 0.4.
7–9. (a) 8.82×10^6 ergs; (b) 198 cm/sec; (c) 190 cm/sec.
7–11. 11,900 ft lb/sec.
7–13. (a) 1,875 lb; (b) 400 ft/sec; (c) 1,100 ft/sec.
7–15. No answer.
7–17. 6.7 lb.
7–19. 9×10^{16} ergs.
7–21. (a) -24; (b) 18, $-z$-direction; (c) 18, $+z$-direction.
7–23. 37.5 mi/hr.
7–25. (a) 0; (b) 30.25 ft lb; (c) no answer.

Chapter 8

8–1. (a) 15.6 slugs/ft^3; (b) 1.56 slugs/ft^3; (c) 0.8.
8–3. (a) 0.5 ft; (b) 17,500 lb; (c) 20,000 lb; (d) flat.
8–5. (a) 1.65 cm; (b) 1.55 cm.
8–7. 37.8 cm.
8–9. (a) 4.53×10^4 lb/ft^2; (b) 1.94×10^{10} lb.
8–11. (a) 10.4 lb/ft^2; (b) 2,130 lb/ft^2.
8–13. 76 cm.
8–15. (a) 4,830 tons; (b) 15,000 ft^3.
8–17. (a) 1,670 dynes; (b) 2,500 dynes; (c) 1,670 dynes; (d) 1,250 dynes.

Chapter 9

9–1. (a) 5.43 ft/sec; (b) 3.06 ft/sec.
9–3. (a) 27 ft/sec; (b) 141 ft^3/min.
9–5. 1.44×10^4 sec.
9–7. (a) 26.8 ft/sec; (b) 2.33 ft^3/sec.
9–9. 5,180 lb.
9–11. (a) 22.6 ft/sec; (b) 0.471 ft^3/sec; (c) 10 lb/in^2.
9–13. (a) 9.9 m/sec; (b) 14.1 m; (c) 17.1 m/sec, $\theta = -54°35'$.
9–15. 1.96×10^{-3} slug/ft sec.
9–17. (a) 0.436 ft^3/sec; (b) 10.3 lb/in^2.

Chapter 10

10–1. (a) 395 slug ft/sec; (b) 395 lb sec; (c) 20,000 lb; (d) 0.0198 sec.
10–3. (a) 150,000 dynes; (b) 0.96.
10–5. (a) $V_1 = 54$ cm/sec, $V_2 = 144$ cm/sec; (b) 0.64.
10–7. 31.2 ft/sec.
10–9. (a) 9,760 lb sec; (b) 9,760 lb sec.
10–11. No answer.
10–13. (a) 0.1 ft; (b) 0.001 ft.
10–15. $10e^2$; (b) $10e^4$, only if $e = 1$.
10–17. 69 lb.
10–19. (a) 650 lb sec; (b) 156 lb sec.
10–21. (a) 500 nt; (b) 625 joules; (c) 50 nt sec.

Chapter 11

11–1. 7×10^4 ft lb.
11–3. (a) 37.5 rad/sec^2; (b) 375 rad/sec.
11–5. No answer.
11–7. (a) 1.33 rad/sec^2; (b) 1.09×10^6 gm cm^2.
11–9. (a) 768 ft lb; (b) 768 ft lb; (c) 22.6 rad/sec.
11–11. (a) 7.68×10^4 gm cm^2; (b) 30.6 rad/sec^2; (c) 14.7×10^4 dynes; (d) 212 cm/sec.
11–13. 3.9 ft/sec.
11–15. 181 rad/sec.
11–17. 0.38 rad/sec.
11–19. Left.
11–21. Stops after second reversal.
11–23. No answer.
11–25. (a) 40 rad/sec^2; (b) 200 rad/sec; (c) 6×10^{14} m/sec^2; (d) 0.0195 kg m^2.
11–27. (a) 0; (b) 4.93×10^6 dyne cm; (c) 4.93×10^6 dyne cm.

Chapter 12

12–1. (a) 0.314 sec; (b) 1.2×10^6 dynes; (c) 2.4×10^3 cm/sec^2.
12–3. (a) 1.176×10^4 dynes/cm; (b) 0.448 sec.
12–5. 2.25 kg.

12–7. (a) 4.27×10^4 dynes; (b) 1.71 rad/sec^2; (c) 0.546 rad/sec; (d) 2.46×10^5 dynes; (e) 47.4 cm/sec, 85 cm/sec^2.
12–9. 977.3 cm/sec^2.
12–11. (a) 0.25 sec^{-1}; (b) -1.5 rad/sec; (c) 1.48 rad/sec^2; (d) 0.60 sec.
12–13. $\frac{1}{8}$ sec.
12–15. 1.58 sec.
12–17. 9.14 cm.
12–19. $2\pi(h\rho/\rho_0)^{1/2}$.

Chapter 13

13–1. (a) 10^8 dynes/cm^2; (b) 8×10^{-5}.
13–3. (a) 2.39×10^3 lb/in^2; (b) 1.82×10^{-4}; (c) 5.46×10^{-4} ft.
13–5. 7.13×10^{11} dynes/cm^2.
13–7. 1.9%.
13–9. (a) 3.6×10^6 dynes/cm^2; (b) 8.5×10^{-6}; (c) 8.5×10^{-6} rad.
13–11. (a) 1.27×10^4 lb/in^2. (b) 9.7×10^{-4} (brass), 7.3×10^{-4} (copper).
13–13. (a) 0°; (b) 45°.
13–15. -0.87 cm, -0.58 cm, -0.44 cm.
13–17. (a) 104 dynes; (b) 520 ergs.
13–19. 4,730 dynes/cm^2.
13–21. 1.5 cm.
13–23. No answer.

Chapter 14

14–1. 160°C.
14–3. 80.026 ft.
14–5. 4.43×10^9 dynes/cm^2.
14–7. 547.8 mm of Hg.
14–9. (a) 3.66×10^{-3} per °C; (b) 66.9°C.
14–11. No answer.
14–13. (a) 864 atmos; (b) 793 atmos.
14–15. 10.015 lb.
14–17. 1,740 lb/in^2.
14–19. (a) 3.7×10^{-4} per °C; (b) 6×10^{-4} per°C.

Chapter 15

15–1. 31.8°C.
15–3. 20.6°C.
15–5. 0.21°F.
15–7. $\Delta t = 6.5$°C.
15–9. 436 gm.
15–11. 0°C, 268 gm.
15–13. 1,760 watts.
15–15. (a) 89%; (b) 6,000 joules; (c) 1.43°C.

Chapter 16

16–1. 95.1 cm of Hg.
16–3. No answer.
16–5. 27 lb/in^2.
16–7. (a) 5.65×10^{-14} erg; (b) 7.72×10^{-14} erg.
16–9. 2.41×10^{19} molecules/cm^3.
16–11. (a) 348 cal; (b) 249 cal; (c) 22.4 liters.
16–13. (a) 362 cm^3; (b) 37°C.
16–15. 1.57×10^{10} ergs.
16–17. 103 cm^3.
16–19. 8.41 joules/kg °C.
16–21. 2 joules.
16–23. No answer.

Chapter 17

17–1. (a) 40 cal; (b) 500 cal; (c) 0.93.
17–3. (a) 1.013×10^6 dynes/cm^2; (b) 1.049×10^6 dynes/cm^2.
17–5. No answer.
17–7. (a) 46.9%; (b) 12.3%.

Chapter 18

18–1. 0.208 cal/cm sec °C.
18–3. 144,300 cal.
18–5. 15,100 Btu/hr.
18–7. 19.9 watts.
18–9. 49.93 watts.
18–11. (a) 72.7 °C; (b) 1.30 cal/sec.
18–13. (a) 15°F/in; (b) 180 Btu/hr ft^2; (c) 100°F.
18–15. 5,700°Abs.

Chapter 19

19–1. (a) 26.8%; (b) 268 cal; (c) 732 cal.
19–3. 46.9%.
19–5. (a) 750 cal; (b) 3,750 cal.
19–7. 33.3%.
19–9. 102°C.
19–11. 20%.
19–13. (a) −2.68 cal/°K; (b) +2.68 cal/°K; (c) 0; (d) 0.
19–15. (a) $+10^{-3}$ joule/°K; (b) 0; (c) -10^{-3} joule/°K; (d) 0; (e) 0.

Chapter 20

20–1. 2,000 cm/sec.
20–3. (a) 16 cm; (b) 16,000 cm/sec; (c) 1.02×10^7 dynes.
20–5. 0.17 sec.
20–7. No answer.
20–9. (a) $y = 5 \cos \frac{2\pi x}{100} \sin \frac{2\pi t}{0.01}$; (b) 9.5 cm.
20–11. 2°35′ N of E.
20–13. (a) 5 cm; (b) 1.54 cm.
20–15. 0.

Chapter 21

21–1. (a) 880 vib/sec; (b) 1.25 ft.
21–3. 22.9 vib/sec.
21–5. (a) 1,064 vib/sec; (b) 944 vib/sec.
21–7. (a) 997.3 vib/sec; (b) 1,002.7 vib/sec; (c) 5.4 beats/sec.
21–9. 4 beats/sec.
21–11. (a) 1.25 ft; (b) 880 vib/sec, 1,320 vib/sec.
21–13. (a) 2.16×10^8 dynes; (b) 1,040 vib/sec, 1,560 vib/sec.
21–15. (a) 2.46; (b) 24°.
21–17. 55 ft to 0.055 ft.

Chapter 22

22–1. 1.12×10^4 dynes.
22–3. 8.19×10^{-8} nt.
22–5. 0.304 dyne, $\theta = 205°15'$.
22–7. 9.0 gm.
22–9. 4.4×10^{-40}.

Chapter 23

23–1. $4 \times \mathbf{1}_r$ dynes/stcoul.
23–3. 0.352 dyne/stcoul.
23–5. $\mathbf{E} = (13.3\ \mathbf{1}_x + (0)\ \mathbf{1}_y)$ dynes/stcoul.
23–7. $\mathbf{E} = (1.48 \times 10^7\ \mathbf{1}_x - 2.04 \times 10^7\ \mathbf{1}_y)$ nt/coul.
23–9. 98 nt/coul, upward.
23–11. (a) $3.60 \times 10^5\ \mathbf{1}_n$ nt/coul; (b) 0.
23–13. (a) 0; (b) 0; (c) $2.25 \times 10^9\ \mathbf{1}_r$ nt/coul.

Chapter 24

24–1. (a) 6×10^4 dynes, in direction of **E**; (b) 2.4×10^5 ergs; (c) -2×10^4 stvolts.
24–3. 1.5×10^{-4} joules.
24–5. 33.3 stvolts.
24–7. (a) 6,600 volts; (b) 9,000 volts; (c) 0.012 joule.
24–9. $V_{r0} = \dfrac{\rho}{6\epsilon_0}(3a^2 - r_0^2)$.
24–11. 1 erg $= 6.25 \times 10^{11}$ ev.
24–13. 5.2×10^{-19} cal.
24–15. 1.02×10^7 m/sec.

Chapter 25

25–1. (a) 1.67×10^{-4} μfd; (b) 2.50×10^{-4} μfd.
25–3. No answer.
25–5. (a) 240 μcoul; (b) 1.44×10^{-2} joule.
25–7. (a) 159 stfd; (b) 1.99×10^4 stcoul; (c) 1.24×10^6 ergs.
25–9. $C = 4\pi\epsilon_0 \dfrac{r_1 r_2}{r_2 - r_1}$.
25–11. (a) 300 μcoul, 500 μcoul; (b) 8 μfd; (c) 0.04 joule.
25–13. (a) 17.1 μcoul, 12.9 μcoul; (b) 4.29 volts.
25–15. (a) 2.93 μfd; (b) 41.7 volts; (c) 166.7 μcoul.
25–17. 10 stcoul cm.
25–19. 0.334 dyne/stcoul, $\theta = 4°45'$.
25–21. $\mathscr{W}_V = \frac{1}{2}\kappa_e\epsilon_0 E^2 = \frac{1}{2}DE$.

Chapter 26

26–1. 3 amp.
26–3. 2.4×10^4 amp/m^2, $-x$-direction.
26–5. (a) 0.33 amp; (b) 360 ohms.
26–7. (a) 5.47×10^{-3} ohm; (b) 0.318 ohm.
26–9. 60.8°C.
26–11. 0.042 mho/m.
26–13. 6×10^6 ohms.
26–15. 1.46×10^7 ohms.

Chapter 27

27–1. (a) 18 ohms; (b) 4 amp, 2.67 amp.
27–3. (a) 1.33 amp; (b) 20 volts, 33.3 volts, 66.7 volts; (c) 160 watts.
27–5. 1 watt, 2 watts, 3 watts.
27–7. (a) 112 volts; (b) 70 ohms.
27–9. (a) 720 ohms; (b) 20 watts.
27–11. (a) 120 volts; (b) 0.2 ohm.
27–13. (a) 99 volts; (b) 297 watts; (c) 94%; (d) 82.5%.
27–15. 0.505 ohm, 0.051 ohm, 0.005 ohm; 950 ohms, 9,950 ohms, 99,950 ohms.
27–17. 1×10^{-4} amp.
27–19. R ohms.
27–21. (a) 0.635 amp; (b) -2.18 volts.
27–23. 2.56 ohms.

Chapter 28

28–1. 2.02 gm.
28–3. (a) 0.380 amp; (b) 1.52 volts.
28–5. (a) 0.207 amp; (b) 2.07 volts.
28–7. (a) $a = 0$, $b = 595 \times 10^{-3}$, $c = 1.45 \times 10^{-5}$, $\mathscr{E}$ in mv; (b) 158°C.

Chapter 29

29–1. (a) 36.5 dynes, $\theta = 218°15'$; (b) 44.4 dynes, $\theta = 0°$; (c) 794 dynes, $\theta = 104°45'$.
29–3. 1,640 dynes, repulsion.
29–5. 0.84 oersted, $\theta = 0°$.
29–7. (a) 8.75 nt/weber, $\theta = 90°$; (b) 4.30 nt/weber, $\theta = 236°50'$; (c) 321 nt/weber, $\theta = 104°35'$.
29–9. 760 cgs pole cm.
29–11. 900 nt/weber.
29–13. $\mathbf{H} = \dfrac{p}{\kappa_m r^2}\mathbf{1}_r$; $\mathbf{H} = \dfrac{p}{4\pi\mu_0\kappa_m r^2}\mathbf{1}_r$.

Chapter 30

30–1. (a) 125 amp/m; (b) 1.57 oersteds.
30–3. $25°15'$.
30–5. (a) 0.114 oersted; (b) 8.54 amp.
30–7. 40 dynes.
30–9. No answer.
30–11. 637 amp/m.
30–13. 105 dyne cm.

Chapter 31

31–1. 4.8×10^{-9} dyne.
31–3. 1.8×10^{-17} gm cm/sec.
31–5. $F_x = 6 \times 10^{-5}$ nt, $F_y = -4.5 \times 10^{-5}$ nt, $F_z = 1.6 \times 10^{-5}$ nt.
31–7. $F_x = 2.5$ nt, $F_y = 0$, $F_z = -1.5$ nt.
31–9. (a) 0.045 cm; (b) 3.52×10^3 turns/m.
31–11. 0.38 amp.
31–13. 9.6×10^{-10} nt m.
31–15. (a) 2.310×10^{-20} nt (repulsion); (b) 2.53×10^{-23} nt (attraction); (c) 2.307×10^{-20} nt (repulsion).
31–17. $R = W/BI$.

Chapter 32

32–1. 0.004 volt.
32–3. 1,770 volts.
32–5. 7.5 volts.
32–7. 2.26 volts.
32–9. 39.8 amp/m.
32–11. 0.05 henry.
32–13. 7.9×10^{-4} henry.
32–15. (a) 0.946 amp; (b) 2.396 amp; (c) 2.400 amp; (d) 0.002 sec.
32–17. No answer.
32–19. (a) 1.26×10^{-2} joule; (b) 251 joules/m^3.

Chapter 33

33–1. (a) 0.63 henry; (b) 0.34 henry; (c) 0.18 henry.
33–3. (a) 5.45×10^{-4} weber/m^2; (b) 12.6 cm.
33–5. (a) 8,000 amp/m; (b) 8.05 weber/m^2; (c) 8.04 weber/m^2.
33–7. 0.072 amp.
33–9. (a) $H = 0$; (b) $H = I/2\pi a$.

Chapter 34

34–1. (a) 12.1 ohms; (b) 49.5 ohms; (c) 2.22 amp; (d) 3.14 amp; (e) $14°10'$; (f) 106.6 volts; (g) 26.8 volts; (h) 237 watts.
34–3. No answer.

34–5. (a) $X_L = 678.6$ ohms, $X_C = 265$ ohms, $Z = 415.3$ ohms; (b) 0.265 amp; (c) $V_R = 10.6$ volts, $V_C = 70.2$ volts, $V_L = 180$ volts.
34–7. No answer.
34–9. (a) 55 to 1; (b) 0.273 amp.
34–11. 2.2×10^{-3} fd.
34–13. (a) 2.0 amp; (b) 0.2 henry; (c) 123.7 ohms.
34–15. (a) 1.5 amp; (b) 0.894; (c) 100 ohms.

Chapter 35

35–1. 50 ma.
35–3. (a) 215 ma; (b) 108 volts.
35–5. 5.25 ma.
35–7. No answer.
35–9. 300 m.

Chapter 36

36–1. 1.26×10^{-5} steradian.
36–3. (a) 3.38×10^{-12} erg; (b) 2.96×10^{17} quanta/sec.
36–5. (a) 260 lumens; (b) 1.53%; (c) 20.7 candelas; (d) 2,630 candelas/m^2.
36–7. (a) 0.92 watt; (b) 0.2 lumen/watt.
36–9. 195 candelas.
36–11. (a) 18.1 lumens/ft^2; (b) 7.35 lumens/ft^2; (c) 19.6 lumens/ft^2.
36–13. (a) 58 cents; (b) 13 cents.

Chapter 37

37–1. (a) 2.254×10^{10} cm/sec; (b) 2.245×10^{10} cm/sec.
37–3. (a) 1.68 cm; (b) 1.71 cm.
37–5. (a) 32°30′; (b) 44°45′; (c) 39°45′.
37–7. 3.0 cm^{-1}.
37–9. 1.5.
37–11. 3°15′.
37–13. No answer.

Chapter 38

38–1. 30 cm; real, inverted, enlarged.
38–3. (a) -6 cm; (b) 1.5; (c) upright, virtual, enlarged.
38–5. (a) -16.7 cm; (b) 25 cm^2.
38–7. (a) 53.3 cm; (b) 3.33 cm; (c) real, inverted, enlarged.
38–9. 1.7%
38–11. (a) 37.5 cm; (b) 35.2 cm.
38–13. (a) $s_1' = 20$ cm; (b) $s_2' = 3.33$ cm; (c) -0.67; (d) real, inverted, diminished.
38–15. (a) $s_2' = 6.78$ cm; (b) -0.261.
38–17. (a) 5 cm from arc; (b) 4.76×10^{-2} lumens/cm^2.

Chapter 39

39–1. 1 ft.
39–3. 3 diopters.
39–5. 50 cm.
39–7. (a) -66.7 cm; (b) 66.7 cm.
39–9. 0.23 in.
39–11. 6.15 cm, 73.85 cm.
39–13. 4.5 cm.

Chapter 40

40–1. (a) 0.103 cm; (b) 0.103 cm.
40–3. 66.3%.
40–5. No answer.

40–7. (a) 5.61×10^{-3} cm; (b) 20.8 lines/cm.
40–9. 0.118 cm. **40–11.** (a) 15°20′; (b) 31°50′.
40–13. 5,556 lines.

Chapter 41

41–1. (a) 58°45′; (b) 51°5′. **41–3.** $\sin\theta_C = \cot\theta_P$.
41–5. (a) N_0; (b) 1 : 0.64.
41–7. (a) 1.64×10^{-3} cm; (b) 3.28×10^{-3} cm; (c) retarded.
41–9. (a) 0.9; (b) 0.1; (c) 0.5. **41–11.** 9.38%.

Chapter 42

42–1. (a) 3.03×10^{-12} erg; (b) 1.89 ev.
42–3. 1.9 ev. **42–5.** 1.63×10^4 °K.
42–7. $\mathcal{E} = -\dfrac{2\pi^2 me^4 Z^2}{n^2 h^2}$.
42–9. (a) $\lambda = h^2 n/2\pi e^2 m$; (b) $n\lambda = 2\pi r_n$; (c) yes, they are equivalent statements of the same hypothesis.
42–11. 1.53×10^{-32} cm, particle-like behavior.
42–13. 5.46×10^{-13} cm.

Chapter 43

43–1. (a) 4°46′; (b) 9°32′; (c) 2.2%. **43–3.** 0.558 A.
43–5. K_α-0.574 A, K_β-0.485, K_γ-0.460 A, K_δ-0.449 A.
43–7. (a) Yes; (b) 0.124 A; (c) 9.92 Kev; (d) plate; (e) plate.
43–9. (a) 3.97×10^{-8} erg; (b) 1.92×10^{-9} erg; (c) 1.92×10^{-9} erg.
43–11. (a) 0.307 A; (b) 3°8′.
43–13. (a) 1.99×10^{-18} gm cm/sec; (b) 0.332 A.
43–15. (a) 0; (b) 2.4×10^{-61} cm; (c) 8.9×10^{-24} cm; (d) 8.9×10^{24} cm.

Chapter 44

44–1. 0.00785 = 7.31 Mev. **44–3.** 8.81 cm, 9.52 cm.
44–5. (a) 308 gausses; (b) 320 cm, 322 cm.
44–7. 190 Mev.
44–9. (a) 1.53×10^{11} cal; (b) 6.1×10^4 lb of coal.

Chapter 45

45–1. (a) 4.87×10^{-18} sec^{-1}; (b) 1.23×10^4 alpha particles/sec.
45–3. 3.32×10^{-7} curie. **45–5.** 6.96×10^{-4} curie.
45–7. (a) 0.00524 amu; (b) 0.08 Mev.

Chapter 46

46–1. $0.66L_0$. **46–3.** $0.995\,c$.
46–5. (a) 2.62×10^{10} cm/sec; (b) 0.545 Mev; (c) 0.661 Mev.
46–7. No answer.
46–9. (a) 19.7 Mev; (b) 39.4 Mev; (c) 78.8 Mev.
46–11. (a) ${}_{92}U^{238} + {}_0n^1 \to ({}_{92}U^{239}) \to {}_{56}Ba^{143} + {}_{36}Kr^{94} + 2{}_0n^1$; (b) 0.233 amu; (c) 217 Mev; (d) 217 Mev.