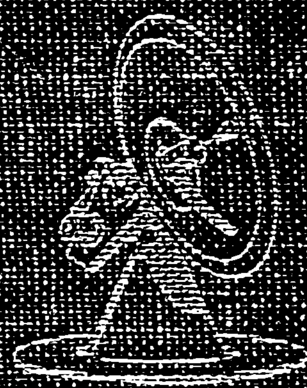
PHYSICS

PRINCIPLES AND APPLICATIONS

PHYSICS

PRINCIPLES AND APPLICATIONS



PHYSICS
PRINCIPLES AND APPLICATIONS

PHYSICS
Principles and Applications

PHYSICS

Principles and Applications

BY

HENRY MARGENAU
Professor of Natural Philosophy and Physics, Yale University

WILLIAM W. WATSON
Professor of Physics, Yale University

CAROL G. MONTGOMERY
Associate Professor of Physics, Yale University

FIRST EDITION

McGRAW-HILL BOOK COMPANY, INC.
NEW YORK TORONTO LONDON · 1949

PHYSICS

THE MAPLE PRESS COMPANY, YORK, PA.

PREFACE

This book has grown out of a collaboration of the authors in a sophomore course in physics at Yale University, a course which is aimed to equip engineering students and majors in the natural sciences with sufficient basic knowledge of physics to serve as a foundation for further work in the various exact sciences and technologies. We have been conscious of the need for a book that presents in a single volume an account of the subject which is at once rigorous, vital, and modern, and we have endeavored to fill that need.

The problem of how to use the calculus in an introductory physics course on the college level has given us serious concern over the whole period in which this course was taught. The students' knowledge of that discipline is still in the formative stage and amounts to a set of newly acquired skills rather than understanding. Yet physics requires the understanding as well as the skill. On the other hand, a use of the calculus is so clearly advantageous and so obviously desirable that its renunciation, in this day and age, is an unwarranted sacrifice. For the saying that calculus is the language of physics is true as well as trite, and continuity of training demands that a student should, even in his early work, be confronted with those methods of analysis which form the major tools of his later thinking. Altogether too many students have to confess, at the end of their senior year, that their introductory physics course was nearly useless because it did not acquaint them with the "elegant" methods of the calculus. And the fact that the physics course provides opportune applications for a student's fresh mathematical acquisitions is not to be overlooked.

The fact remains, however, that few students can "take it" when the calculus is employed as freely as these considerations suggest. Hence, in composing this book, we have adopted a device which is neither new nor ingenious: We have allowed the student to catch his breath by treating subjects not requiring the use of this form of mathematics in the first four chapters and then exposing him to it increasingly, but in easy stages and with full explanation. Toward the end of the book the student is expected to take elementary integrations in his stride. Our treatment is suited to the level of attainment of readers who have had an introduction to the calculus and are taking a more solid course concurrently.

Students often complain that a course like the one for which this book may serve as a text is very difficult, and they invariably, though somewhat confusedly, blame the "emphasis on mathematics" for their dissatisfaction. In our experience this diagnosis has often been in error: What has created difficulty is not the mathematics but the volume of novelties that is being pushed out to the students. The wise teacher, when using a book like the present, will select his topics with care and limit his selection with pedagogical foresight.

For that reason, we have put in the book rather a plenum of material for selection. It is our guess that coverage of all items presented will require three full semesters' work. In our own course, which meets four times a week (exclusive of laboratory exercises) for two semesters, we have succeeded in covering about three-fourths of the subject matter without incurring accusations of having idled. This was done by omitting many of the starred sections. To make possible this selective procedure without prescribing what ought to be left to every teacher's judgment, we have divided the book into *starred* and *unstarred* sections. The unstarred sections, which can be covered in leisurely fashion in a two-semester course meeting three hours per week, form a basic and coherent matrix of work and are understandable by themselves. In only a few places will it be found necessary to take for granted a result derived in a starred section. In a similar way some of the problems at the end of each chapter are distinguished by stars. Unstarred problems go with the unstarred material.

Worked examples are employed very freely everywhere in the text. They represent a simple pedagogic device for concrete understanding, a device which the authors have found very useful in their teaching.

Our outlook, we believe, is fairly modern. In electromagnetism the exposition recognizes that the forces between charges, at rest and in motion, are fundamental, and hence magnetic poles take on minor significance in this book.

We have shunned the sensational for its own sake. Rather have we endeavored to broaden the treatment and to make contact with the students' general studies by inserting historical references, by explaining the etymology of important technical terms, and by an occasional comment in a philosophic vein.

The proper treatment of units in physics is a subject for controversy, often heated, and we have tried to adopt a moderate viewpoint. In the first chapters, the familiar British units are most often employed, but as the subject is developed emphasis is shifted to the metric systems in both the cgs and the mks forms. In the study of electricity, where three or four metric systems are in common use today, the attempt has been made to minimize troubles by employing three constants that have different values and dimensions in the different systems. All the fundamental equations are written, therefore, in terms that are independent of the units except for these constant factors. Although this device may not lessen the difficulties, it is hoped that it will help the student to understand more clearly the distinctions between arbitrary conventions and physical principles.

THE AUTHORS

NEW HAVEN, CONN.
January, 1949

CONTENTS

PREFACE . . . v

1. The Methods of Physics . . . 1
2. Introduction to Mechanics; Forces . . . 10
3. Statics . . . 27
4. Elasticity . . . 50
5. Description of Motions; Kinematics . . . 61
6. Force and Acceleration; Dynamics . . . 79
7. Special Motions in a Plane . . . 93
8. Work, Momentum, Conservation Principles . . . 105
9. Dynamics of a Rigid Body . . . 132
10. Dynamics of a Rigid Body—*Continued* . . . 148
11. Motions under Inverse-square Laws of Force . . . 159
12. Oscillations . . . 173
13. Hydrostatics . . . 190
14. Hydrodynamics . . . 203
15. Intermolecular Forces . . . 212
16. Temperature and Thermal Expansion . . . 222
17. Calorimetry . . . 233
18. Change of State . . . 246
19. Properties of Gases and Vapors . . . 258
20. Kinetic Theory . . . 275
21. The Laws of Thermodynamics . . . 290
22. Transfer of Heat . . . 303
23. Meteorology . . . 312
24. Electric Charges and Coulomb's Law . . . 323
25. The Electric Field and Potential . . . 333
26. Dielectrics and Capacitance . . . 353
27. Electric Currents . . . 366
28. Chemical and Thermal Electromotive Forces . . . 378
29. Magnetic Forces on Currents. Electrical Instruments . . . 390
30. The Magnetic Field Produced by Currents . . . 408
31. Magnetic Fields and Poles . . . 419
32. Induced Electromotive Force . . . 429
33. The Magnetic Properties of Matter . . . 444
34. Alternating Currents . . . 455
35. Electronics . . . 471
36. Electronics—*Continued* . . . 489
37. Wave Motion . . . 500
38. Production and Reception of Sound . . . 523

39. The Nature of Light 551
40. Reflection and Refraction at Plane Surfaces 566
41. Reflection and Refraction at Spherical Surfaces. 581
42. Lenses. 593
43. Optical Instruments. 606
44. Photometry 619
45. Spectra and Color . . . 627
46. Interference 639
47. Diffraction. 651
48. Radiant Energy . . . 671
49. Atomic Structure. . . . 686
50. Nuclear Physics 708

APPENDIXES
1. Units 729
2. Values of Important General Physical Constants 731
3. Conversion Table of Electrical Units 732
4. Greek Alphabet 733
5. Useful Mathematics. 734
6. Natural Trigonometric Functions. . 735
7. Logarithms . 736

INDEX 739

INDEX OF TABLES 759

CHAPTER 1

THE METHODS OF PHYSICS

Physics is an *exact science.* As such it uses methods of procedure that set it apart from many other fields of interest, notably the "humanistic" subjects, and it is well to note these differences of method at the very beginning of our study. A clear view of *what* the physicist aims to do and *how* he does it is important for an understanding of physics. Most of this book is devoted to the question of "what"; this chapter is briefly concerned with a few aspects of the "how."

★1.1. Definitions. In physics, symbols are used with the utmost care. The symbols may be words, or algebraic letters and signs that usually represent words, or numbers or diagrams. The artifacts of letters and mathematical signs are chosen to promote clarity of thought; diagrams and pictures are used to ensure complete visualization. Even words are employed in a manner typical of an exact science: they are given a specificity of meaning that poets would regard as monotonous. Thus words frequently take on a narrow significance not necessarily found in ordinary discourse. When the physicist speaks of force, work, energy, he refers to matters far more definite than are implied by a historian's use of the same words. He is forced to such narrowness by the difficulty of his task, and he accepts the limitations upon his discourse gladly because of the successes they have brought. To ensure the careful use of words the scientist must *define* his words and symbols with great precision. In physics, therefore, proper understanding and proper use of definitions are an important preliminary to progress. This explains why much emphasis is laid upon definitions in this book. Definitions are usually given in the form of explanatory sentences; but they may also appear in the form of mathematical equations provided that the symbols are defined in words. Sometimes the last-named method is simpler. For instance, if we wish to define density, it is not too much trouble to say: "Density is mass per unit volume." But to define kinetic energy as "one-half the product of the mass of an object by the square of its velocity" tends to be a mouthful. It is preferable to say it is $\frac{1}{2}mv^2$ after we have defined the symbols m and v.

★1.2. Use of Mathematics. While most subjects taught at the college level involve the use of some form of logic, physics relies heavily on the most nearly perfect form of reasoning, on mathematics. Far from being ashamed of diluting his science with a seemingly foreign formalism, the physicist takes pride in the circumstance that his subject permits him the use of so lofty a discipline, for it betokens a degree of refinement in the growth of his science that workers in other sciences covet. But mathematics is never used for its own sake. It is always a tool employed either to shorten the labor of experimentation or to predict new facts on the basis of already established knowledge. This book will contain numerous proofs. These were introduced, not to make the treatment of the subject "highbrow," but to show what is actually going on in it. It is hoped that the reader will perceive, in every instance of the use of mathematics, the purpose of clarity or economy of thought for which it was designed. The demonstrations given are not always the best available; they are nearly always compromises between simplicity and rigor.

1.3. Measurements and Units. Physics is a science of measurement. To be sure, it contains a great deal of reasoning, but this reasoning never departs very far from the solid ground of measured magnitudes. Accurate measuring is an art to which this book is only an imperfect guide. Some concrete appreciation of it is taught in the laboratory accompanying this course. An understanding of what is involved in the measuring process, aside from technique, however, can be prepared here.

It is important to realize that almost every measured result contains two things, a *number* (sometimes called a *numeric*) and a *unit*. The number alone has no significance unless the unit is also stated. A change in the unit requires a corresponding change in the numeric. Such changes will often engage our attention and will at once be made clear by examples. Suppose we measure a certain length and find it to be 9 feet. In this case it is apparent that we may also say 3 yards or 108 inches. But it is well to formulate a rule for making such conversions, for in less familiar instances simple intuition may not work. The rule is this: *Treat all units as algebraic symbols, and express the symbols by other equivalents.* The numbers appearing in this process are then combined to give the new numeric. Thus

$$9 \text{ ft} = 9(12 \text{ in.}) = 108 \text{ in.} = 9(\tfrac{1}{3} \text{ yd}) = 3 \text{ yd}$$

or, to choose a slightly more difficult example,

$$50.0 \text{ miles/hr} = 50.0(5{,}280 \text{ ft})/3{,}600 \text{ sec} = 73.3 \text{ ft/sec}$$

It will later be seen that units can also be canceled like algebraic symbols. A certain feature, which might be called the "counterplay of unit and numeric," is also apparent in these examples: when the unit increases by a certain factor, the number decreases by the same factor, or vice versa; a larger unit requires a smaller number.

While nature presents the scientist with certain obvious units such as the day or the year or the foot, most units are man-made and employed by convention. They must always be defined with great care and precision. In fact the student should always expect, when he is introduced to a new scientific subject, the following sequence of events: first the definition of *quantities* (*e.g.*, energy, power, etc.) which are important in the field under study; second a definition of *units* in terms of which these quantities are measured; finally a discussion of *laws* and *regularities* which prevail among the quantities and through which they can often be measured or calculated.

1.4. Knowledge Taken for Granted. It seems well to state what knowledge on the part of the student will be taken for granted in this book. First of all, familiarity with the ideas of *distance* and *time* is assumed. These are in fact taken as primitive concepts, and no attempt to derive them will be made. Such attempts would fall into the domain of philosophy. Distance covered per unit of time is a *speed.* While *mass* and *force* are quantities to be defined more precisely later on, a general acquaintance with them will be presupposed. Finally the reader will be expected to know the simpler units in which these quantities are customarily measured. There are certain everyday measuring devices, such as the *yardstick*, the *clock*, and the *balance*, the use and action of which are assumed to be understood. These three are in a sense the basic and universal instruments of physics; all others can be analyzed into modifications and combinations of them. The quantities they measure, *length*, *time*, and *mass*, are sometimes called *basic* or *primitive*, for out of them, too, all other quantities can be constructed in a manner discussed in Section 6.7.

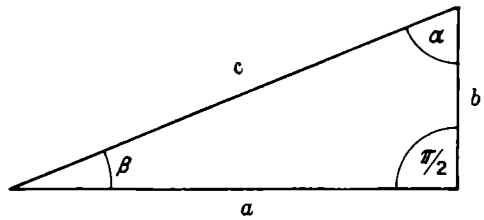


FIG. 1.1. Illustration of trigonometry.

Finally a certain amount of mathematical skill should have been acquired by the student. Aside from arithmetic, geometry, and the elements of algebra, he will know the notation and the use of trigonometry. He is reminded of the definitions[1]

$$\begin{aligned} \sin \alpha &= \cos \beta = a/c \\ \sin \beta &= \cos \alpha = b/c \\ \tan \alpha &= \cot \beta = a/b \end{aligned}$$

which hold for any right triangle (cf. Fig. 1.1); and of the following relations, which hold for any angles θ, θ_1, and θ_2:

$$\tan \theta = \sin \theta/\cos \theta \qquad \cot \theta = 1/\tan \theta$$

$$\sec \theta = 1/\cos \theta \qquad \csc \theta = 1/\sin \theta$$

$$\sin^2 \theta + \cos^2 \theta = 1 \qquad \text{(theorem of Pythagoras)}$$

$$\sec^2 \theta = 1 + \tan^2 \theta$$

$$\sin (\theta_1 + \theta_2) = \sin \theta_1 \cos \theta_2 + \cos \theta_1 \sin \theta_2 \tag{1.1}$$

$$\cos (\theta_1 + \theta_2) = \cos \theta_1 \cos \theta_2 - \sin \theta_1 \sin \theta_2 \tag{1.2}$$

[1] Greek letters will sometimes be employed because they are customary or because

From the last two we get

$$\sin 2\theta = 2 \sin \theta \cos \theta \tag{1.3}$$

and

$$\cos 2\theta = \cos^2 \theta - \sin^2 \theta \tag{1.4}$$

by putting

$$\theta_1 = \theta_2$$

We also recall the following relations, valid for *any* triangle such as the one in Fig. 1.2:

$$\frac{\sin \alpha}{a} = \frac{\sin \beta}{b} = \frac{\sin \gamma}{c} \tag{1.5}$$

$$c^2 = a^2 + b^2 - 2ab \cos \gamma \tag{1.6}$$

$$c^2 = a^2 + b^2 + 2ab \cos \gamma' \tag{1.6'}$$

Angles, while usually measured in degrees, are more naturally expressed in *radian* measure. The definition of the radian is based on the observation that the ratio of any arc to its corresponding radius, such as s_1/r_1 or

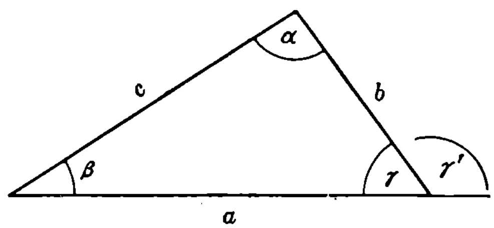


FIG. 1.2. Illustration of trigonometry.

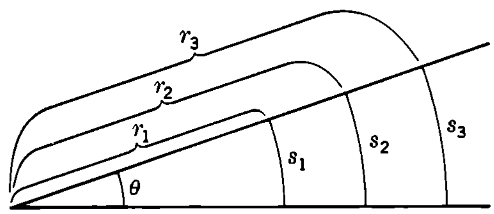


FIG. 1.3. Radian measure.

s_2/r_2 or s_3/r_3 in Fig. 1.3, is a constant as long as θ is constant, and hence a convenient measure of θ. This ratio is the size of θ in radians. Clearly, $90° = \pi/2$ radians, $360° = 2\pi$ radians, and 1 radian $= 360/2\pi$ degrees. In the expression of an angle in radians the unit (which is after all only a ratio!) is often omitted.

A bundle of rays going out from a point O (cf. Fig. 1.4) is said to subtend a *solid* angle, Ω, at O. To obtain a measure for Ω, let us suppose that, at a distance r from O, the solid angle has an opening—no matter how irregular—of area A. Now construct a sphere of radius r about O. The area cut out from this spherical surface by the rays forming Ω will be the projection of A upon the spherical surface. We call the area of this projection A'. Then the magnitude of Ω is said to be

$$\Omega = \frac{A'}{r^2} \text{ steradians}$$

One *steradian* is a solid angle whose opening at a distance of 1 cm from

---

the Roman alphabet is exhausted. Their symbols, together with their names and their modern equivalents, are given in the Appendix.

its origin is 1 cm^2. Since the full sphere has an area of $4\pi r^2$, the solid angle which it subtends at the center is $4\pi r^2/r^2 = 4\pi$ steradians.

Suppose now that Ω is very small, of magnitude $d\Omega$. The area A in Fig. 1.4 will then also be small and will be dA. In that case the projection of dA upon the sphere, dA', will be $dA \cos \theta$, if θ is the angle between the *normal* to dA (cf. the figure) and the radius vector drawn from O. Hence

$$d\Omega = \frac{dA \cos \theta}{r^2}$$

This formula will be used in Sec. 25.4 and elsewhere.

Calculus will be used with increasing frequency as the development in the book proceeds. At the beginning, however, its use will be restricted to very elementary operations of differentiation and integration. As far as possible, mathematical devices will be explained before they are used; but it is supposed that the student has a nodding acquaintance with the calculus and is pursuing it with greater attention as the work in this book goes forward.

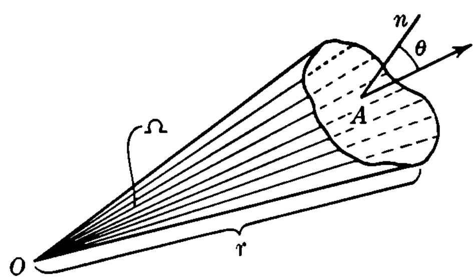


FIG. 1.4. A solid angle.

1.5. Special Systems of Elementary Units. The variety of units used in science and engineering is great, and it is important to introduce some order into their complexity at once. There are two large *systems* of units in common use, one called the *British* system, the other the *metric* system. The basic units in each are three, as already noted: the unit of length, that of time, and that of mass. The basic units in the British system are the *foot*, the *pound* (mass), and the *second*. Hence a common abbreviation for this system, and one that we shall use henceforth, is the fps system. The metric system is used in two different forms, the *centimeter-gram-second* (cgs) and the *meter-kilogram-second* (mks) varieties.

Table 1.1. *Basic Units in fps, cgs, and mks Systems*

System	Quantity		
	Length	Mass (or force)	Time
fps	Foot (ft)	Pound (lb)	Second (sec)
cgs	Centimeter (cm)	Gram (gm)	Second (sec)
mks	Meter (m)	Kilogram (kg)	Second (sec)

All three systems are given in Table 1.1. In another interpretation, favored by engineers, the pound, the gram, and the kilogram represent *forces*, not masses. This matter will be carefully discussed in Secs. 2.3, 6.2, and 6.3. For the present the student may regard these terms as representing either masses or forces.

Observe that the unit of time is the second in all systems. The convenience of the cgs and mks systems arises chiefly from the ease with which smaller or larger units can be formed from the basic ones. The units differ by factors that are powers of 10. The rule for formation of desired units is simple if the reader will remember a few Greek and Latin prefixes, which are listed in Table 1.2.

Table 1.2. *Derived Metric Units*

1/1,000,000	1/1,000	1/100	1/10	Unit	10	100	1,000	1,000,000
micro-	*milli-*	*centi-*	deci-	Meter, gram, second, liter, etc.	deka-	hekto-	*kilo-*	mega-

It will be seen that, while the centimeter is the basic unit of length in the cgs system, the linguistic unit happens to be the meter. From it, multiple or submultiple units are derived in accordance with Table 1.2. The prefixes in italics are the most common. Any of them may be employed with other basic units; for instance, a milliliter is a thousandth of a liter, a microsecond is 10^{-6} second, and so forth. The mícrometer (not to be confused with the micrómeter, an instrument for measuring small lengths) is usually called the *micron*.

One further advantage of the metric systems is the existence of a simple relation between the basic unit of length and that of mass. A cube of side length 1 cm (*i.e.*, a volume of 1 cubic centimeter, abbreviated 1 cc or 1 cm^3) filled with water (at 39°F to be exact) has a mass of 1 gram. One cubic meter (m^3) of water has a mass of 1,000,000 gm, or 1,000 kg, or 1 metric ton.

One pound equals 453.59 *gm.* On the continent of Europe a slightly larger pound, called the "metric pound," is used; it equals 500 gm, that is, ½ kg. But this will not be employed in the present text. Finally we remark that the meter was originally chosen to be 1/10,000,000 of the distance between the earth's equator and the pole, but it is now defined as the length between two scratches on a standard meter bar. A complete conversion table between metric and British units is given in Appendix 1.

***1.6. Numerical Computations, Accuracy.** In physics, great emphasis is laid upon numerical computations. In fact all algebraic formulas are derived only for the purpose of using them in computation; they serve no mathematical purpose in themselves. A few elementary points have to be borne in mind when formulas are put to numerical use. To illustrate them, we consider a geometrical example, the formula for the area of an ellipse,

$$A = \pi ab$$

where a and b are the semimajor and semiminor axes, respectively.

The subsequent remarks may at first seem trivial, but they are important in respect to the more complex formulas of physics that occur throughout this book. First we observe that some of the symbols, such as a and b, have units as well as magnitudes, while π is merely a number. At the beginning, then, it is necessary to distinguish the purely numerical symbols from those which, in addition to a numeric, require units. The latter are said to possess "physical dimensions," a matter to be discussed in Sec. 6.7.

Next, definite units have to be chosen for a and b, and here it is important to use the *same* units. Let us choose feet. Both number and unit must be introduced into the formula, so that it reads, for example,

$$A = \pi \times 2 \text{ ft} \times 3 \text{ ft}$$

thus $A = 6\pi \text{ ft}^2 = 18.8496 \cdots \text{ft}^2$. The units of A are given automatically by this procedure. Had we introduced a in feet and b in inches, a mixed unit for area that is of obscure significance would have appeared. The lesson we have learned is this: In using a formula, all symbols denoting quantities of the same kind ("having the same physical dimensions") should be expressed in the same unit. We could, of course, have chosen inches in the present example. The answer would then be

$$A = \pi \times 24 \text{ in.} \times 36 \text{ in.} = 2{,}714.34 \cdots \text{in.}^2$$

Now the question arises as to how many decimal places should be used in stating the answer. Obviously, since π is a never-ending decimal fraction, the result could be stated to any number of figures. Brief reflection tells us that this would be absurd unless we knew a and b with unlimited accuracy. In general, a and b will not be the round numbers we have chosen. Suppose they were measured with an ordinary foot rule, which can be read accurately to within $\frac{1}{32}$ in., or 0.03 in. approximately. That is to say, the true value of a may lie anywhere between 23.97 in. and 24.03 in., that of b between 35.97 in. and 36.03 in. To put it another way, we are not sure of the fourth figure in the numerics of a and b (notice that this would also be true if we had expressed a and b in feet!); hence we cannot be sure of the fourth figure in A. A little thought will convince the student that there is no sense, under these circumstances, in using more than four figures in π.

The number of digits in any result of which one can be sure in view of the accuracy of the data (in this case a and b) is the *number of significant figures.* When a scientist states a result to a great number of significant figures, he claims great accuracy. This should always be remembered in giving numerical answers. Further, the use of a greater number of digits in any numerical constant such as π, or $\frac{1}{3}$ in its decimal expansion, than corresponds to the number of significant figures of the data is pointless and hence unscientific. For the same reason, one should never employ six-place logarithms when computing data that have only three significant figures. If the data are not all known to the same accuracy, the number of significant figures of the least accurate datum dominates the entire process.

Worked Examples.[1] *a.* The distance from an observer at which lightning strikes is determined by measuring the time interval between lightning flash and thunderclap with a stop watch that can be read with an accuracy of 0.2 sec. The interval is found to be 11.6 sec. The speed of sound is 1,127.5 ft/sec (at a temperature of 20¼°C, which we shall assume). The distance is

$$1{,}127.5 \text{ ft/sec} \times 11.6 \text{ sec} = 13{,}079.00 \text{ ft.}$$

However, since the third significant figure in 11.6 is already uncertain, this result should properly be rounded off to 13,100 ft. In fact the judicious student will never use the accurate figure 1,127.5 ft/sec in conjunction with the crude 11.6 sec but will round it off at once to 1,130 ft/sec. He will then obtain the answer 13,108 ft, which should be reduced to 13,100 ft.

On the other hand this reduction must not be overdone. It is true, of course, that an answer of three apparently significant figures can be obtained by multiplying 11.6 by 1,100 ft/sec. This, however, leads to answers that are not correct to three figures, as the multiplication will show. Hence it is never safe to use fewer significant figures in *any* factor than the number of significant figures in the one *least accurately known.*

b. Normal atmospheric pressure is 14.7 lb/in^2. Convert this into gm/cm^2. Solution:

$$14.7 \frac{\text{lb}}{\text{in.}^2} = 14.7 \frac{(454 \text{ gm})}{(2.54 \text{ cm})^2} = \frac{14.7 \times 454}{(2.54)^2} \frac{\text{gm}}{\text{cm}^2} = 1{,}035 \frac{\text{gm}}{\text{cm}^2}$$

PROBLEMS

1. Sound travels through the atmosphere with a velocity of 340 m/sec, light with a velocity of 3×10^8 m/sec. Convert both these into ft/sec and into miles/hr.

2. The density of water is 1 gm/cm^3. What is it in lb/ft^3?

3. If a band were tightly stretched around the earth's equator (regarded here as a perfect circle), its length would be 40,000 km. Suppose the band were lengthened by 10 ft, still forming a circle with its center at the earth's center. What would be the distance from the earth's surface to the band?

[1] The student should do the worked examples himself and check his work step by step with that of the book.

4. The height of a tower is to be found by triangulation (cf. Fig. 1.5). The base a is 300 ft, $\theta_1 = 35°$, $\theta_2 = 45°$. What is h?

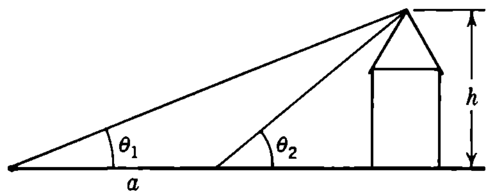


FIG. 1.5. Problem 4.

5. Express $\sin 3\theta$ and $\cos 3\theta$ in terms of θ.

6. Convert the following angles into radian measure: 10°, 45°, 90°, 180°, 360°.

7. Make graphs of $\sin\theta$ and $\cos\theta$ vs. θ. Show that

$$\sin(\theta + 2\pi) = \sin\theta \qquad \cos(\theta + 2\pi) = \cos\theta$$
$$\sin(\theta + \pi) = \sin(\theta - \pi) = -\sin\theta$$
$$\cos(\theta + \pi) = \sin(\theta - \pi) = -\cos\theta$$
$$\sin\left(\theta + \frac{\pi}{2}\right) = -\sin\left(\theta - \frac{\pi}{2}\right) = \cos\theta$$
$$\cos\left(\theta + \frac{\pi}{2}\right) = -\cos\left(\theta - \frac{\pi}{2}\right) = -\sin\theta$$

8. How many cubic millimeters are there in a cubic meter?

★9. A rectangular box was measured with a meter stick and found to have a height of 45.4 cm, a width of 37.8 cm, and a length of 107.9 cm. State its volume, rounding off the answer to the proper number of significant figures.

CHAPTER 2

INTRODUCTION TO MECHANICS; FORCES

★2.1. Brief History of the Science of Motion. Throughout the ages motion, though one of the most familiar spectacles of nature, has presented to the curious mind some very puzzling features. The philosophers of ancient Greece were greatly concerned about the difference between the state of rest and that of motion. To them it seemed that, for a thing to *exist*, its properties must be permanent. But moving bodies have no permanence of position, and from this dilemma some of them, notably *Parmenides*, drew the conclusion that moving bodies do not exist. Motion was explained as an illusion.

Others in ancient Greece, among them the philosopher *Anaxagoras* (500 B.C.?), regarded motion as real but felt called upon to distinguish clearly two states of being—rest and motion. They speculated extensively about the origin of motion; in fact they invented a special stuff, or substance, that would account for motion. Later this view was rejected, and *Leucippus*, the inventor of the idea of atoms, regarded motion as an eternal property of all atoms, thus denying in effect the state of rest.

Out of this speculative welter there crystallized a theory, chiefly promoted by the great philosopher and scientist *Aristotle* (384 B.C.), that dominated the thinking of the Middle Ages. It was an unsophisticated theory of motion, no longer acceptable today but nevertheless interesting and appealing. Against the background of this theory the progress of physics since the Renaissance must be judged. Aristotle was preoccupied with the question of why heavy bodies fall toward the earth instead of rising to the heavens like a flame. He found the answer to this paradox by an appeal to the belief, then current, in the existence of a fiery sphere in the heavens and of the earthy realm below ground. The solution was this: The flame rises because it seeks the realm to which it belongs, its natural place; the stone falls because it, too, seeks its natural place. Thus the motion of bodies "going home," or seeking their natural places, seemed to be accounted for. It was called *natural motion*.

But a stone can also be projected upward! To be sure, this requires an external agency like an arm or a slingshot; hence it is in a sense "unnatural" motion. Aristotle called it *violent* motion. And in this connection he introduced a famous postulate, which was upheld until the days of Galileo almost 2,000 years later. The postulate was this: *Violent* motion requires the application of *force; natural* motion proceeds *without* force. It is not difficult to see the homely reasonableness of this assumption.

Whether we regard it as correct depends on our acceptance of this particular meaning of the term force. The Middle Ages accepted it. Hence to the medieval mind the falling stone required no force, and it is unjust for us to accuse our pre-Galilean scientific forebears of error in this respect. However, the Aristotelean postulate and the definition of force that it implies are not very fortunate and fruitful and have therefore been abandoned.

It was discovered by Galileo (1564–1642) and proved convincingly by Newton (1642–1727) that the whole science of motion becomes at once simple and universally valid if *force is taken to mean any agency which produces*, or may produce, an *acceleration* in physical objects. The subject of acceleration will be studied in considerable detail later. In a preliminary way we define it here as *rate of change in velocity*. An object at rest or moving at constant velocity has no acceleration. Thus, according to this new conception a falling body must experience a force because it is accelerated.

2.2. Newton's Laws of Motion. Among the truly great discoveries of mankind are Newton's three laws of motion. Though simple in wording and in conception, they contain almost the whole of the science of motion, the entire intricate formalism of modern mechanics. We state them here in their historical form and recommend that the student commit them to memory and endeavor to understand them fully. Much more will be said about them in due course. Their full meaning will not be made clear in the present chapter, which is devoted to a study of a very limited aspect of them. In Newton's famous work "Principia mathematica" the laws are stated in the following form (translated from the Latin):

1. Every body continues in its state of rest or of uniform motion in a straight line, except in so far as it is compelled by forces to change that state.

2. Change of motion (acceleration) is proportional to the force and takes place in the direction of the straight line in which the force acts.

3. To every action there is always an equal and opposite reaction; or the mutual actions of any two bodies are always equal, and oppositely directed, along the same straight line.

The meaning of the first law is clear; it is only to be noted that uniform motion is motion with *constant velocity*. The second law is qualitatively understandable; however, to give it quantitative and precise meaning, we must carefully state how (1) forces and (2) acceleration are to be measured. These two items will be postponed, the first to the next section, the second to a later chapter. The third law is more difficult to comprehend and will be the subject of much further comment. We elucidate here merely by pointing out that *action and reaction are names for forces acting on two different bodies which tend to accelerate each other.*

For the present, our interest will be in only one consequence of the first and second laws: When the force acting on a body is zero, it is not accelerated and is said to be in equilibrium. The engineer often wants to know the conditions under which bodies, such as parts of machinery, bridges, and other structures, are and will remain in equilibrium. The part of physics dealing with this question is called "statics." But before we turn our attention to this, we must learn more about forces.

2.3. Forces. There are several ways of defining the meaning of force, all of which may be shown to be equivalent. The simplest and most useful to the engineer is this: *A force is a push or a pull.* Everyone has an intuitive appreciation of force through his experience of muscular exertion.

The present definition may become inadequate and require revision under some circumstances. For instance, it is true that the sun exerts a force on the earth, a force that is transmitted through a vacuum, and it is a little hard to visualize how ths sun can exert a pull across empty space. But these complications will not concern us now. Later we shall see how it is possible also to define force in terms of acceleration with the use of Newton's second law.

If a force is a push or a pull, how is it to be measured? Clearly, our muscular sense is too crude. We note, however, that a given force has a specifiable and under proper conditions an invariable effect on an elastic body, such as a spring. It is possible, therefore, to use a spring balance as an instrument for measuring forces. Imagine, then, a standard spring balance, a certain mark upon which designates the unit of force, the *pound*. By comparison with it, other balances can be calibrated. There are other units of force than the pound, but their consideration will be postponed.

Unfortunately there is an ambiguity in physical usage, for the pound is also a unit of *mass*. This should be interpreted, not as meaning that mass and force are identical physical quantities, but as a regrettable verbal coincidence. It arises from the fact that a *mass* of one pound is pulled toward the center of the earth by gravity with a *force* of one pound called its "weight." Hence *weight is force and must be distinguished from mass*. In the stratosphere the weight-force on a pound mass is less than one pound, as a delicate spring balance would show.

The pound (lb) is the British unit of force. It alone will be used in this chapter and the next. Two thousand pounds (2,000 lb) is called a *ton*.

2.4. Reckoning with Forces. Forces do not satisfy the laws of ordinary arithmetic. One object plus another object will always make two objects, but one force plus another force may well be no force whatever. It is because of this possibility of *compensation* that the reckoning with forces has to be carefully scrutinized.

To begin with, it is necessary, in every situation involving forces, to distinguish between the forces exerted *on* an object and those exerted *by* an object. In Fig. 2.1, for example, two bodies are connected by a spring. Aside from gravity, which is ignored, there are altogether four forces. F_1 is a force exerted *by* the spring *on* body A; F_2 is exerted *by* A *on* the spring; F_3 is exerted *by* the spring *on* body B; F_4 is exerted *by* B *on* the spring. Thus A and B are subject to one force each, while the spring is acted on by two forces.

We now consider all the forces *on* a body, such as the one in Fig. 2.2. To specify them it is clearly necessary to state the *magnitude* (in pounds) and the *direction* of each. Because forces have both magnitude and direction, they require different treatment from quantities that have only magnitude, like length, time, and mass. With regard to the latter it is easy to see what is meant by adding them. The sum of 2 lb and 3 lb is simply 5 lb; it is the result of putting together 2 lb and 3 lb. The sum of two forces is also the result of putting them together, but in this process we must respect directions as well as magnitudes. Thus to add forces requires a new procedure, which will now be described. But it should always be borne in mind that we add forces acting on the same body, never forces acting on different bodies.

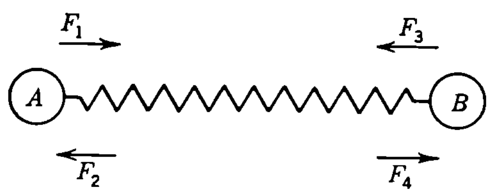


FIG. 2.1. Forces exerted *on* and *by* bodies.

2.5. Addition of Two Forces. Suppose that a body is subject to two forces, one of magnitude 2 lb acting from left to right, and another of magnitude 3 lb acting from right to left. Their effect is the same as one force of magnitude 1 lb acting from right to left, as a little consideration or a simple experiment with spring balances would show. These two forces, when acting simultaneously, produce a *resultant force*, which happens to be equal in magnitude to their difference. If both forces act in the same direction, the magnitude of the resultant would be their sum.

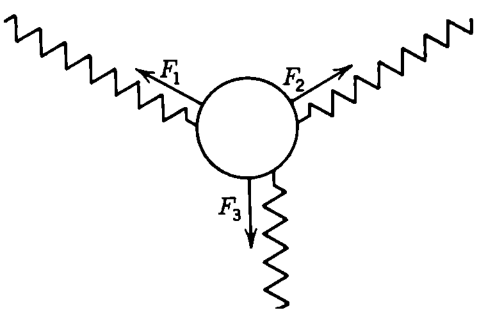


FIG. 2.2. Forces have both magnitude and direction.

The situation is more complicated when the two forces do not act in the same direction. In Fig. 2.3*a* the heavy lines represent two forces on a body (not shown in the diagram). If a spring balance were inserted, it would show that this body is urged in the direction of the dotted arrow by a force of 5 lb. The latter force is called the *resultant*, or *sum*, of the other two. It is not the arithmetical sum but can be found by the following rule (cf. Fig. 2.3*b*):

Represent each force by an *arrow* the direction of which is that of the force. Its length is taken to be proportional to the magnitude of the force so that a given length, perhaps an *inch* or a centimeter, represents one pound, the same for each force. In Fig. 2.3*b* the two representative arrows are denoted by F_1 and F_2. To find the resultant of F_1 and F_2, place the end of F_2 in coincidence with the head of F_1. Then draw an arrow

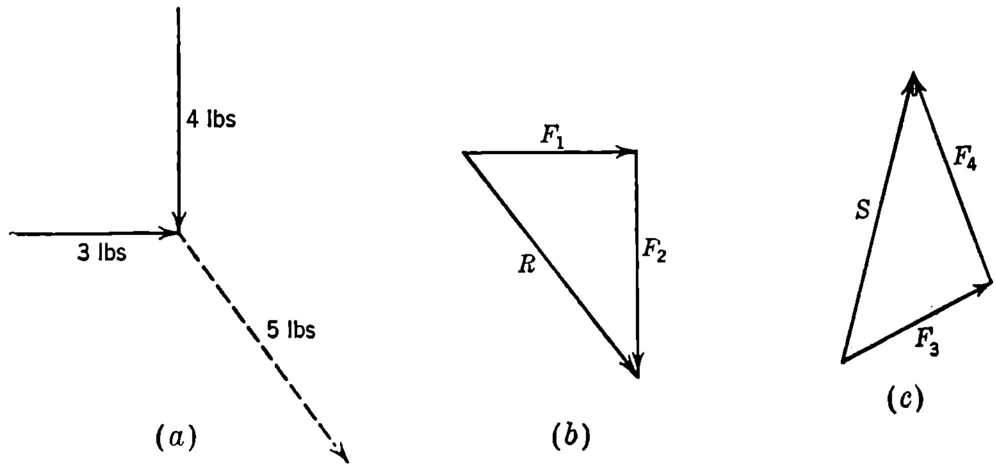


Fig. 2.3. Addition of forces.

from the end of F_1 to the head of F_2. This arrow is the resultant, labeled R in the diagram. This rule is general; in Fig. 2.3*c*, S is the resultant of F_3 and F_4. In the laboratory this rule can be verified with spring balances in all instances. It can also be proved theoretically, but this will not be done at present (see Sec. 2.12).

For brevity we shall also call the resultant the *sum* of the two other forces, and we shall refer to the rule as the *rule for adding forces*. Illustrations of it are numerous and can easily be supplied by the reader. If, for instance, two people were pulling on a body in the manner of F_3 and F_4 in Fig. 2.3*c*, their effect would be the same as if a single person pulled in the direction of S with a force corresponding to that magnitude.

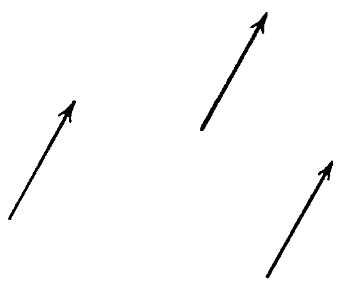


Fig. 2.4. Equal forces.

It is well to become thoroughly familiar with the representation of a force as an arrow and to think of it always in this symbolism. Two arrows represent equal forces if they have the same magnitude and direction, no matter where they are drawn. For example, the three forces shown in Fig. 2.4 are equal.

2.6. Addition of Many Forces. The rule for adding forces can be generalized to apply to cases where more than two forces act on a body.

Four are shown in Fig. 2.5*a*, corresponding perhaps to different pulls by guide ropes on a tree to be felled. To find the one force that is equivalent to these four, that is, to find their resultant, or sum, we might first add F_1 and F_2, then add F_3 to the resultant of these, then add F_4 to this second resultant. If the student will carry out this procedure, he will find it to be equivalent to the simpler process illustrated in Fig. 2.5*b*, where the forces are added, tail to head, without drawing each resultant. The sum of all four is then simply the arrow drawn from the tail of the first to the head of the last force. For obvious reasons this procedure for obtaining the sum of a number of forces is called the *polygon method:* the resultant

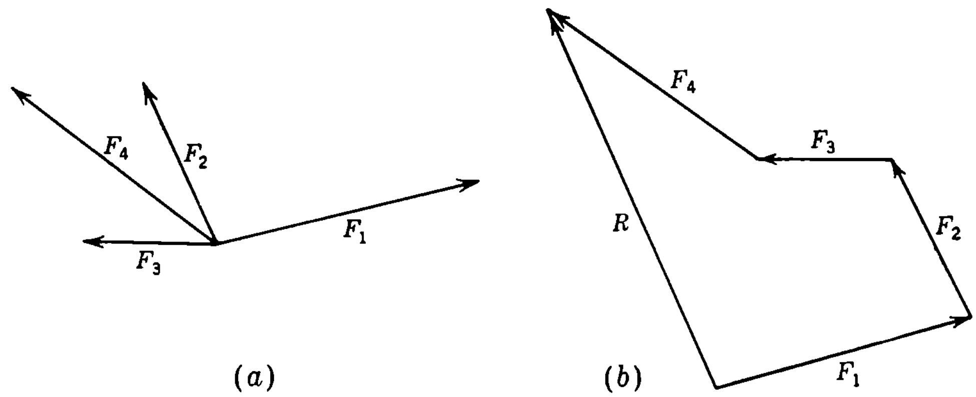


FIG. 2.5. Adding four forces.

(sum) is the closing side of the polygon made up of all the forces, with the arrow properly drawn.

In ordinary algebra we write

$$a + b = c \tag{2.1}$$

in order to denote that c is the sum of the ordinary quantities a and b. Thus we might also designate the addition of the forces $\mathbf{F}_1$ and $\mathbf{F}_2$ in accordance with Fig. 2.3*b* by

$$\mathbf{F}_1 + \mathbf{F}_2 = \mathbf{R} \tag{2.2}$$

or that according to Fig. 2.5 by

$$\mathbf{F}_1 + \mathbf{F}_2 + \mathbf{F}_3 + \mathbf{F}_4 = \mathbf{R} \tag{2.3}$$

The plus signs in Eqs. (2.2) and (2.3) have the meaning given to them by the polygon rule, a meaning different from that of the plus sign in Eq. (2.1). But confusion can be avoided if, while retaining the symbol +, we use **boldface** type for forces as in Eqs. (2.2) and (2.3), or in general for all quantities that are added by the polygon method. This convention will be followed throughout this text.

In ordinary algebra the order of the summands is immaterial in the computation of a sum. That is,

$$a + b = b + a$$

Is it also true that $\mathbf{F}_1 + \mathbf{F}_2 = \mathbf{F}_2 + \mathbf{F}_1$? The latter sum is constructed in Fig. 2.6. Comparison with Fig. 2.3*b* will show that it is equal to $\mathbf{F}_1 + \mathbf{F}_2$. The student will easily convince himself that the forces in Fig. 2.5, when added in any order, will always yield the same sum **R**. Hence the order in which forces are added is immaterial.

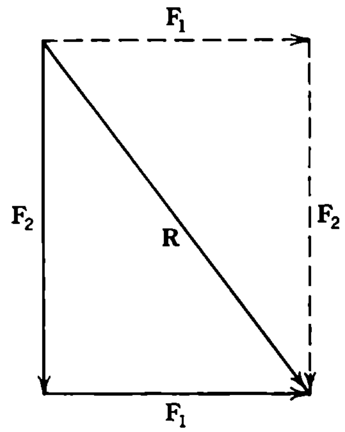


FIG. 2.6. $\mathbf{F}_1 + \mathbf{F}_2 = \mathbf{F}_2 + \mathbf{F}_1$

The rule here given is not the only one by which addition of forces may be performed. In Fig. 2.7*a* we are employing our rule; in Fig. 2.7*b* we lay off $\mathbf{F}_1$ and $\mathbf{F}_2$, not head upon tail, but *tail* upon *tail*. If now we complete the parallelogram and draw the diagonal, this diagonal is identical with **R**. Hence the parallelogram rule: To add two forces, put their tails together, complete the parallelogram, and draw the diagonal from the coincident tails outward. This diagonal is the resultant. The parallelogram rule is sometimes the more convenient to apply, but it cannot be generalized so easily to many forces as can the other.

2.7. Subtraction and Resolution of Forces. Suppose a mass m (Fig. 2.8*a*) is to be held in place by two forces $\mathbf{F}_1$ and $\mathbf{F}_2$, exerted by ropes tied to the mass. We know $\mathbf{F}_1$, and we also know that the total force needed to hold the mass in place is equal to its weight, say 1,000 lb and must act upward (**R** in the diagram). We wish to find $\mathbf{F}_2$. The solution is easy. Since

$$\mathbf{F}_1 + \mathbf{F}_2 = \mathbf{R}$$
$$\mathbf{F}_2 = \mathbf{R} - \mathbf{F}_1$$

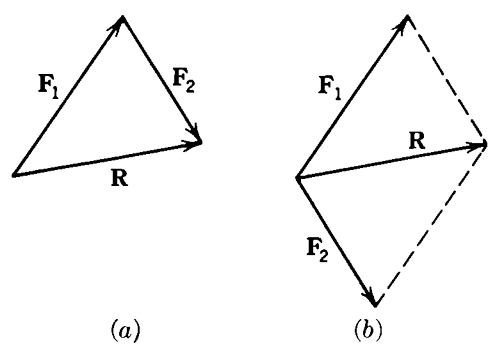


FIG. 2.7. Parallelogram rule for adding forces.

$\mathbf{F}_2$ is the difference of **R** and $\mathbf{F}_1$. But the difference of two quantities is always the *sum with the sign of the second quantity, in this case* $\mathbf{F}_1$, *reversed.* Now obviously $-\mathbf{F}_1$ is $\mathbf{F}_1$ with the arrow reversed. Hence to get $\mathbf{F}_2$ we add **R** and $\mathbf{F}_1$ but reverse the arrow on $\mathbf{F}_1$. This is done in Fig. 2.8*b*.

The next chapter will present problems in which a force is given and it is necessary to find two or more forces which, when added together, will have the given force as resultant. The rules for solving such problems are now at hand, though not the terminology. Two forces that add up to a given force are said to be its *components*, and the process of finding two components of a force is called *resolution*. Thus, in Fig. 2.3*b*, $\mathbf{F}_1$ and $\mathbf{F}_2$ are components of **R**; in 2.3*c*, $\mathbf{F}_3$ and $\mathbf{F}_4$ are components of **S**.

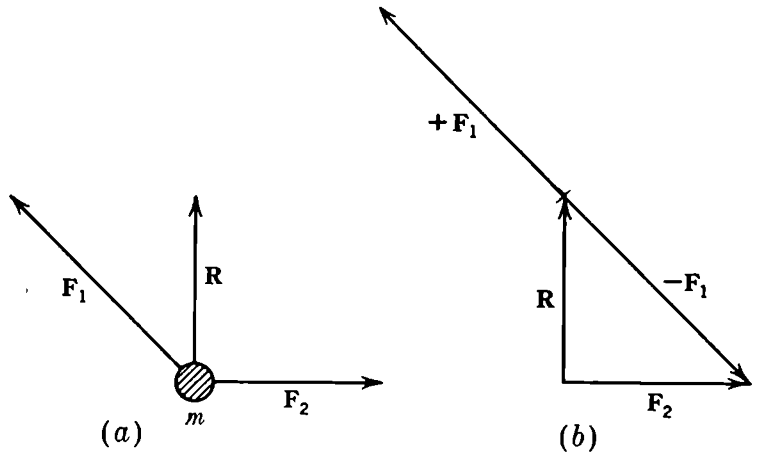


FIG. 2.8. Difference of two forces.

A force can be resolved into components in an infinite variety of ways. Most important, however, is its resolution into components in two given directions, usually along the X and Y axes. The parallelogram rule serves to bring about such resolution easily. Figure 2.9 shows that $\mathbf{F} = \mathbf{F}_x + \mathbf{F}_y$. The forces $\mathbf{F}_x$ and $\mathbf{F}_y$ are called *rectangular components* of **F**. They are found by drawing **F** from the origin of the coordinate system and then dropping perpendiculars from the head of **F** upon the X and Y axes. The notation just used will be continued; that is, $\mathbf{F}_x$ will *always* denote the component of force along the X axis, and so forth.

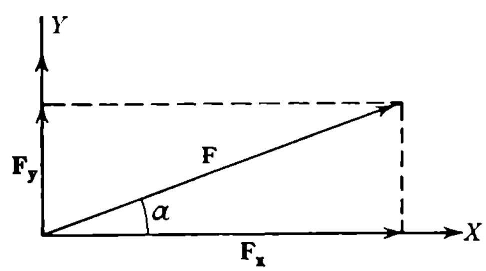


FIG. 2.9. Rectangular components of a force.

Practical applications requiring resolution of forces are very numerous and will be discussed in the next chapter. To mention a few: If we want to know whether or not a body will slide down an incline, we must determine whether the force of friction is smaller or greater than the component of gravity along the incline; the force of the wind upon a sail "resolves" itself into one component along and one component perpendicular to the surface of the sail; drag and lift of an airplane are rectangular components of the force of the air upon the plane.

2.8. Addition, Subtraction, and Resolution of Forces by Trigonometry. Addition of forces, as explained so far, is essentially a graphical procedure. Often, however, the use of trigonometry recommends itself for simplicity

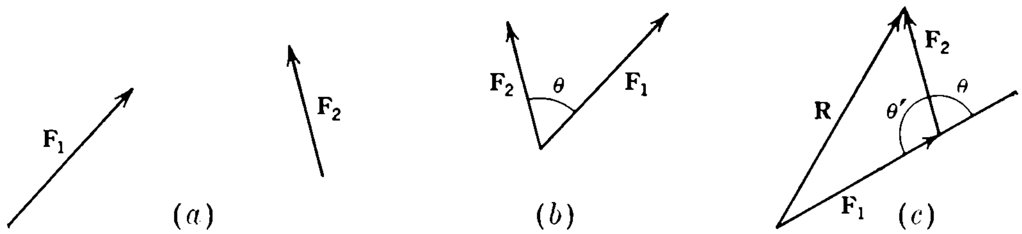


Fig. 2.10. Addition of forces by trigonometry.

and accuracy. If an accurate value of the *magnitude* of the resultant of two forces is wanted, the following method is suitable:

Let θ be the angle between the positive directions of $\mathbf{F}_1$ and $\mathbf{F}_2$, as shown in Fig. 2.10. Then, from (Fig. 2.10*c*), and because of Eq. (1.6),[1]

$$\begin{aligned} R^2 &= F_1^2 + F_2^2 - 2F_1F_2\cos\theta' \\ &= F_1^2 + F_2^2 + 2F_1F_2\cos\theta \end{aligned} \tag{2.4}$$

This is a most useful relation for obtaining the magnitude of $\mathbf{R}$; note, however, that it gives no information as to its direction.

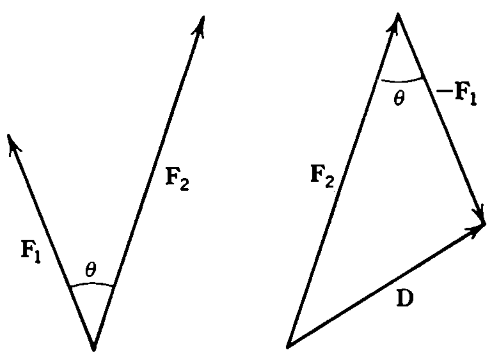


Fig. 2.11. Subtraction of forces by trigonometry.

Inspection of Fig. 2.11 will show that the difference

$$\mathbf{D} = \mathbf{F}_2 - \mathbf{F}_1$$

is given as to magnitude by the square root of

$$D^2 = F_1^2 + F_2^2 - 2F_1F_2\cos\theta \tag{2.5}$$

Resolution of a force along X and Y may also be performed trigonometrically. From Fig. 2.9 it is clear that

$$F_x = F\cos\alpha \tag{2.6}$$
$$F_y = F\sin\alpha \tag{2.7}$$

provided α is taken to be the angle between F and X.

[1] Note that the symbol R stands for the magnitude of the force $\mathbf{R}$, F_1 for the magnitude of $\mathbf{F}_1$, etc.

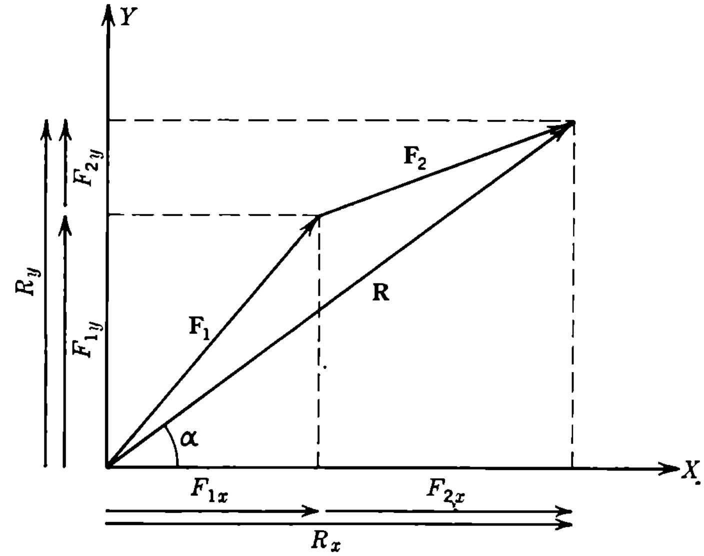


FIG. 2.12.

Figure 2.12 shows that, if

$$\mathbf{R} = \mathbf{F}_1 + \mathbf{F}_2 \tag{2.8}$$

then

$$R_x = F_{1x} + F_{2x} \tag{2.9}$$

$$R_y = F_{1y} + F_{2y} \tag{2.10}$$

An equation in boldface type, like Eq. (2.8), is therefore really equivalent to the *two* equations (2.9) and (2.10).

If a force **F** does not lie in the XY plane, it has not two but three components, F_x, F_y, F_z. These are indicated in Fig. 2.13. Here P is the point where the perpendicular from the head of **F** meets the XZ plane; the arrow **F′** is called the *projection* of **F** upon the XZ plane, F_x and F_z being its rectangular components. In this more general case the resultant of two forces not in the XY plane is also given by

$$\mathbf{R} = \mathbf{F}_1 + \mathbf{F}_2 \tag{2.11}$$

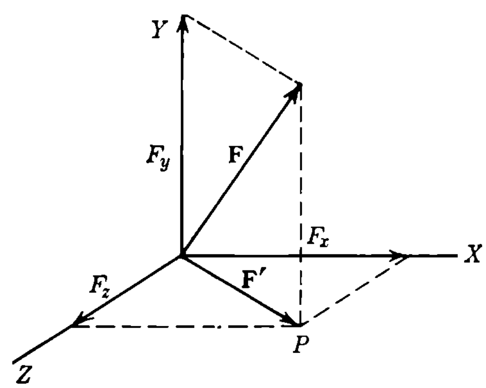


FIG. 2.13. Force components in three dimensions.

but now this equation is equivalent to, or an abbreviation of, the following three:

$$\left.\begin{aligned} R_x &= F_{1x} + F_{2x} \\ R_y &= F_{1y} + F_{2y} \\ R_z &= F_{1z} + F_{2z} \end{aligned}\right\} \tag{2.12}$$

We shall finally consider a problem in which we have to add two forces that are given not in magnitude and direction but in their three components. In other words, we know F_{1x}, F_{1y}, F_{1z} individually, as well as F_{2x}, F_{2y}, F_{2z}; we wish to find **R,** the resultant, both with regard to magnitude and direction. First we observe, using Fig. 2.13 and the Pythagorean theorem, that

$$F^2 = F_y^2 + F'^2$$

while $$F'^2 = F_x^2 + F_y^2.$$

Hence $$F^2 = F_x^2 + F_y^2 + F_z^2$$

This relation holds, of course, for any force and may be applied to **R** appearing in Eq. (2.11). Thus

$$R^2 = R_x^2 + R_y^2 + R_z^2$$

But the components of **R** are given by Eqs. (2.12). Therefore

$$R^2 = (F_{1x} + F_{2x})^2 + (F_{1y} + F_{2y})^2 + (F_{1z} + F_{2z})^2 \qquad (2.13)$$

If $\mathbf{F}_1$ and $\mathbf{F}_2$ lie in the XY plane, we may simply put F_{1z} and F_{2z} equal to zero in this formula.

To find the direction of **R** is difficult in the general case (three dimensions) but very easy when Z components are absent. For then we see, this time from Fig. 2.12, that $\tan \alpha = R_y/R_x$; or if we use Eqs. (2.9) and (2.10),

$$\tan \alpha = \frac{F_{1y} + F_{2y}}{F_{1x} + F_{2x}} \qquad (2.14)$$

The angle α determines the direction of **R**; its value may be found in tables when the quantities on the right are known.

2.9. Worked Example. An airplane momentarily coasting in horizontal flight is subject to two forces, the thrust of the air $\mathbf{F}_1$ and its weight $\mathbf{F}_2$. The thrust of the air has two components, a lift of 8 tons and a drag of 5 tons. The weight is 10 tons. Find magnitude and direction of the resultant force on the airplane.

We have $F_{1x} = -5$ tons $\quad F_{1y} = 8$ tons

$F_{2x} = 0 \quad F_{2y} = -10$ tons

Hence, by Eq. (2.13), $R = \sqrt{(-5 \text{ tons})^2 + (-2 \text{ tons})^2} = \sqrt{29}$ tons This is the magnitude. The direction is obtained from Eq. (2.14),

$$\tan \alpha = \frac{-2 \text{ tons}}{-5 \text{ tons}} = \frac{2}{5} \qquad \alpha = 22° = 0.38 \text{ radian} \quad \text{or}$$

$$202° = 3.52 \text{ radians}$$

Both answers are possible, and from this work we cannot tell whether α lies in the first or the third quadrant.

The example can also be worked graphically, as is shown in Fig. 2.14. We first add $\mathbf{F}_{1x}$ and $\mathbf{F}_{1y}$, then the downward force $\mathbf{F}_{2y}$, obtaining $\mathbf{R}$. It now turns out that only the *latter* of the two answers for α is correct.

If the forces here considered are the only forces acting on the airplane, then it follows from Newton's second law that the airplane, though at present moving horizontally, is starting downward with an acceleration in the direction of $\mathbf{R}$.

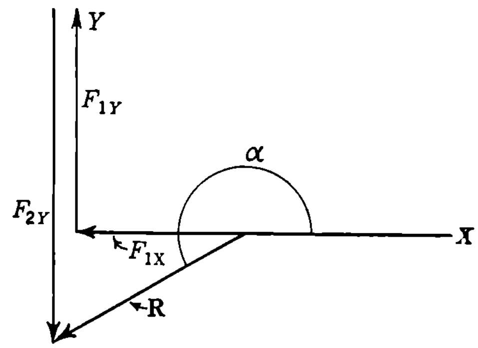


FIG. 2.14.

2.10. Composition of Many Forces; Summation Convention. Assume now that many forces are to be compounded. Thus we write, instead of Eq. (2.11),

$$\mathbf{R} = \mathbf{F}_1 + \mathbf{F}_2 + \mathbf{F}_3 + \cdots + \mathbf{F}_n \tag{2.15}$$

implying that there are n forces, n being some integer. If now a diagram like Fig. 2.12 is drawn for n forces (the student should do this for $n = 4$), it will be seen that

$$R_x = F_{1x} + F_{2x} + F_{3x} + \cdots + F_{nx} \tag{2.16}$$
$$R_y = F_{1y} + F_{2y} + F_{3y} + \cdots + F_{ny} \tag{2.17}$$

provided that all forces lie in the XY plane. The case where they extend in all three dimensions will not be treated.

As before, it follows from the Pythagorean theorem that

$$R^2 = R_x^2 + R_y^2$$

whence, in view of Eqs. (2.16) and (2.17),

$$R^2 = (F_{1x} + F_{2x} + \cdots + F_{nx})^2 + (F_{1y} + F_{2y} + \cdots + F_{ny})^2 \tag{2.18}$$

To find the direction of the resultant $\mathbf{R}$ the equation

$$\tan \alpha = \frac{F_{1y} + F_{2y} + \cdots + F_{ny}}{F_{1x} + F_{2x} + \cdots + F_{nx}} \tag{2.19}$$

must be used.

In problems like this the notation we have used is cumbersome, and we wish to introduce to the student a shorthand way of writing these results. Mathematicians and physicists use the symbol Σ (pronounced sigma, Greek letter for capital S) to represent a *sum*. They write, for example, instead of $a_1 + a_2 + a_3 + \cdots + a_{10}$ simply $\Sigma\, a$ and read it "Sum over all a." But this convention is not very accurate, for it does not tell us how many summands a there are. To remedy this defect one puts a *subscript* i on a and writes, for example,

$$\sum_{i=1}^{10} a_i = a_1 + a_2 + \cdots + a_{10}$$

The left side of this identity is read, "Sum over a_i as i goes from 1 to 10." In general, then, $\sum_{i=1}^{n}$ is to be understood as an order to write down many times the expression which follows it, giving the index i which appears in that expression successively all integer values from 1 to n, and then adding all these expressions. Ordinarily, however, when there is no danger of misunderstanding, or when one does not wish to specify what n is, the limits 1 and n are omitted and one simply writes $\sum_i$ instead of $\sum_{i=1}^{n}$. Using this convention, the formulas of the present section may be written as follows:

$$\mathbf{R} = \sum_i \mathbf{F}_i \tag{2.15}$$

$$R_x = \sum_i F_{ix} \tag{2.16}$$

$$R_y = \sum_i F_{iy} \tag{2.17}$$

$$R^2 = (\sum_i F_{ix})^2 + (\sum_i F_{iy})^2 \tag{2.18}$$

$$\tan \alpha = \frac{\sum_i F_{iy}}{\sum_i F_{ix}} \tag{2.19}$$

We wish to stress here that Σ is not a new and strange mathematical operation but merely a compact way of expressing a familiar sum. Its convenience, apparent already in the formulas above, is even more clearly seen in the example

$$\sum_{i=1}^{10} i = 1 + 2 + 3 + 4 + 5 + 6 + 7 + 8 + 9 + 10 = 55$$

This notation will occasionally be employed in this book.

2.11. Scalar and Vector Quantities. We have learned the rules for adding (or "compounding"), subtracting, and resolving forces acting on a body. These rules are different from those for ordinary addition and subtraction, for forces have magnitude and direction. Actually the methods discussed are part of a very general kind of mathematics known as the *vector calculus*, for they are applicable not only to forces but to all physical quantities having magnitude and direction. These are quite numerous.

To make the matter more definite, physical quantities that are completely specified by their magnitude are called *scalars* (from the Latin *scala,* meaning ladder, a symbol for increasing or decreasing magnitudes reminiscent of a musical scale). Examples of scalar quantities are time, mass, volume, energy. Quantities having magnitude and direction are called *vectors* (from the Latin *vehi,* to ride or to be displaced). A displacement is in fact the simplest kind of a vector; others are velocity, force, acceleration, momentum. All these will be studied in due time. We now see that what has been said about forces is actually true for all vectors. Force was chosen here for illustration because it is perhaps the most familiar vector.

All vectors are symbolized by arrows. In this book, vectors will be printed in boldface type when attention is to be drawn to their vector character. An equation like

$$\mathbf{A} = \mathbf{B}$$

is really equivalent to three equations

$$A_x = B_x, \qquad A_y = B_y, \qquad A_z = B_z$$

and may be regarded as a shorthand way of writing them.

Vector addition may be performed either by the polygon rule or by the parallelogram rule, which we here repeat:

Polygon Rule: The sum (or resultant) of vectors **A, B, C,** . . . is found by placing the tail of **B** upon the head of **A,** then the tail of **C** upon the head of **B,** etc. Finally the polygon is closed by drawing a line from the tail of **A** to the head of the last vector. This line, with the arrowhead at the head of the last vector, is the sum (or resultant).

Parallelogram Rule (good only for adding two vectors): The sum of **A** and **B** is found by placing the tail of **B** upon the tail of **A,** completing the parallelogram, and drawing the diagonal passing through the tails of **A** and **B.** This diagonal, with its arrow pointing away from the tails of **A** and **B,** represents the sum.

***2.12. Theoretical Proof of Rule for Vector Addition.**—Now, with the larger view that the foregoing section has given, we pause and wonder why so strange a device as the polygon rule should work. The reason can be seen best from an examination of the vector called *displacement,* which is exactly what the word implies. If a body is subjected to a displacement of magnitude 3 miles and direction north, it is carried from its present position to one 3 miles north of it. If it is subject to *two* displacements, first the one just mentioned, then another one of magnitude 4 miles and direction east, it will be found a distance

$$\sqrt{(3 \text{ miles})^2 + (4 \text{ miles})^2} = 5 \text{ miles northeast}$$

of its original position, and this position is clearly found by the polygon (in this case, triangle) rule. Several displacements, carried out one after another, will land the body precisely at the place predicted by the polygon rule, and the sum total of all the displacements is *equivalent* to a single displacement that takes the body directly from its original to its final place. This is the sense of the statement that the vector sum of the individual displacements *equals* the displacement given by the polygon rule.

A little reflection will show that it makes no difference whether the two displacements 3 miles north and 4 miles east are carried out successively or simultaneously or in the reverse order; the resultant displacement is always the same.

One can see, therefore, why the polygon rule holds for displacements. But why should it work for *forces*? This can be shown in three more steps, which we shall merely indicate. One goes from displacement to velocity, which, being displacement per unit time, must obey the same rules as displacement. Next, acceleration, being change in velocity per unit time (cf. Sec. 5.2), must obey the same rules as velocity. Finally force, being proportional to acceleration by virtue of Newton's second law, must obey these rules also. This completes the logic of the vector calculus.

PROBLEMS

1. Find the resultant of the two forces in Fig. 2.15 both graphically (polygon rule) and trigonometrically (using formulas). Compare the results.

2. Find the resultant of three forces, all acting in the same plane, F_1 acting north and of magnitude 6 tons, F_2 acting due southwest and of magnitude 10 tons, F_3 acting due southeast and of magnitude 3 tons. What are the components of this resultant in an easterly and a northerly direction? (It is suggested that answers be obtained both graphically and trigonometrically.)

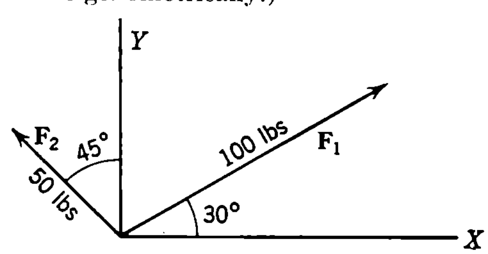


FIG. 2.15. Problem 1.

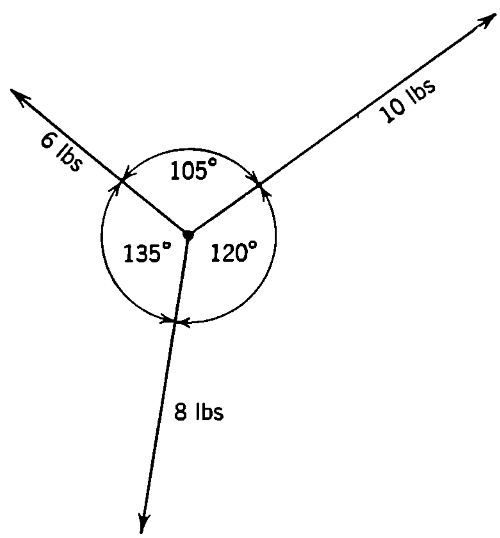


FIG. 2.16. Problem 3.

3. The knot in Fig. 2.16 is subjected to the three forces shown. In what direction will it start to move?

4. The pulley in Fig. 2.17 is fastened in such a way that it can sustain a weight (downward force) of 1 ton. Both sides of the belt B make an angle of 30° with the horizontal. What force, or "tension," along the belt can the pulley stand?

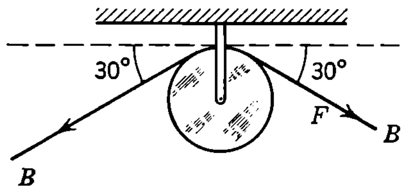


FIG. 2.17. Problem 4.

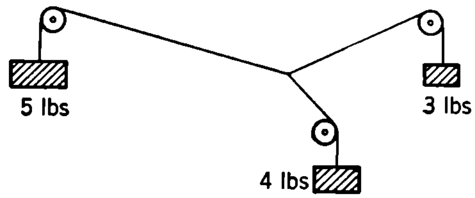


FIG. 2.18. Problem 5.

5. Three weights are arranged over fixed pulleys as shown in Fig. 2.18. Find the angles between the strings at the knot such that the resultant of the forces on the knot be zero.

6. Find the resultant force on the hexagon of Fig. 2.19, all individual forces being of equal magnitude.

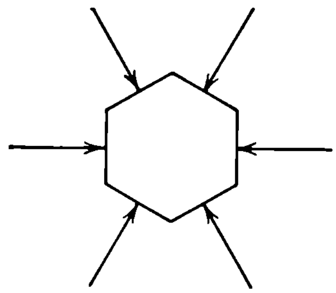


FIG. 2.19. Problem 6.

7. Prove that, if a triangle has three forces acting on it, each perpendicular to a side and of magnitude proportional to the length of that side, the sum of these forces is zero.

8. Prove that, if a polygon is subjected to as many forces as it has sides and each force acts perpendicularly to a side and has the magnitude of that side, the sum of these forces is zero. (This is the situation with respect to pressure forces produced by water upon submerged objects whose horizontal cross section is the polygon.)

9. A rocket in level flight experiences two forces, one directed downward and of magnitude 2 tons (its weight), the other in opposition to its motion and of magnitude 8 tons (air resistance). Find the resultant force on the rocket.

10. A body weighing 10 lb rests on an inclined plane as shown in Fig. 2.20. Find the components of its weight *along* the plane and *perpendicular* to the plane. (HINT: Resolve the weight, which acts vertically downward, along these two directions.)

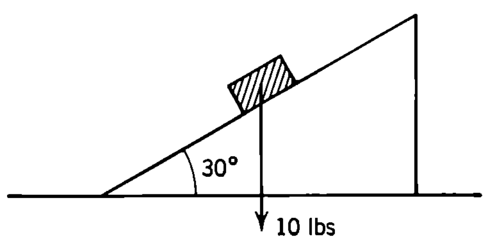


FIG. 2.20. Problem 10.

11. Using Eqs. (2.18) and (2.19), find the magnitude of the resultant, as well as the angle it makes with the X axis, of the following forces:

	Magnitude	Direction with X
F_1	2 lb	15°
F_2	3 lb	110°
F_3	6 lb	90°
F_4	5 lb	−60°
F_5	9 lb	160°
F_6	8 lb	−90°

12. The wind exerts a force of 100 lb perpendicularly against the sail of the boat in Fig. 2.21. Only its component along the line of the keel is effective in propelling the boat. What is this component?

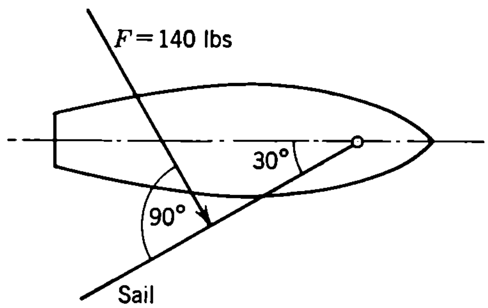


Fig. 2.21. Problem 12.

13. What is the vector sum of all the displacements made by the tip of a second hand as it moves from 12 to 6 on a dial?

14. An airplane is subject to two velocities, its own velocity relative to the air of magnitude 250 miles/hr and directed north, and the wind velocity, which is 50 ft/sec southwest. Find the velocity of the airplane relative to the ground.

15. A rain drop weighs 0.01 gm. As it reaches the earth it falls with constant speed (no acceleration!). Air resistance on the drop is proportional to its speed; it is in fact $(10^{-6}v)$ gm, if v is expressed in cm/sec. Using the information of Chap. 2, find the terminal speed of the drop.

16. The boat drawn in Fig. 2.22 is to move to the right. The wind blows as indicated. How must the sail be set so that the component of the wind force perpendicular to the sail will drive the boat in the forward direction? Taking the angle between wind and keel to be 30°, find the best setting of the sail (assuming the sail to be a plane, which is far from true).

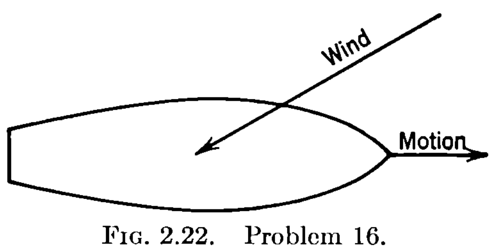


Fig. 2.22. Problem 16.

CHAPTER 3
STATICS

3.1. Equilibrium of a Particle; Concurrent Forces. The subject of *statics* treats of bodies in equilibrium; the word (Greek *stenai*, to stand still) implies that the bodies in question are not moving. This, however, is not meant to be literally true; for the earth itself is moving, and bodies at rest relative to it are certainly in equilibrium in the simple sense of the word. A more adequate definition of equilibrium is *absence of acceleration*. Strictly speaking, therefore, a body moving with a uniform velocity, even relative to the earth's surface, is in equilibrium; our interest, however, will be confined chiefly to cases of rest upon the earth's surface.

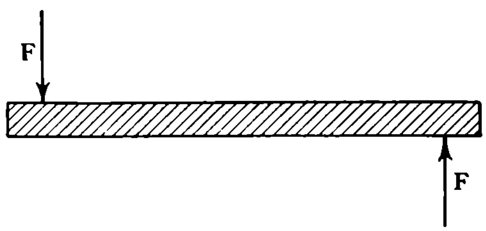


FIG. 3.1. Equal and opposite forces do not necessarily produce equilibrium.

If equilibrium means absence of acceleration, then Newton's second law at once informs us that a body in equilibrium must be subject to no force or, more precisely, to no *resultant force*. It may well have forces acting on it, but their vector sum must be zero. This is often called the condition for equilibrium of forces.

There is, however, an important fact that needs to be considered. Will a body always be in equilibrium when the sum of all forces on it is zero? The answer is no, as reference to Fig. 3.1 clearly shows. The rod will rotate even though the sum of the two equal and opposite forces is zero. We see that the condition for equilibrium of forces, which we now write in the form

$$\Sigma \mathbf{F} = 0 \tag{3.1}$$

does not ensure equilibrium with respect to rotation.

Let us, then, exclude the possibility of rotation for the present. This limits our attention to objects which do not rotate or of which the rotation does not concern us. For convenience we shall here think of a point or a very small particle. If this particle is to be in equilibrium, Eq. (3.1) must be satisfied and we understand by the quantities **F** all forces acting *on* the particle. These forces will necessarily be *concurrent*, *i.e.*, they all are applied at the point.

From what has been learned in the previous chapter it is apparent that Eq. (3.1) can also be written in terms of components,

$$\left.\begin{aligned} \Sigma F_x &= 0 \\ \Sigma F_y &= 0 \\ \Sigma F_z &= 0 \end{aligned}\right\} \qquad (3.2)$$

Equations (3.1) and (3.2) are equivalent statements of what we shall call the *force condition* for equilibrium, or the condition for equilibrium of a particle (as distinguished from an extended body). In applications it is sometimes more convenient to use Eq. (3.1); sometimes Eqs. (3.2) are preferable.

★3.2. Examples of Concurrent Forces. A weight is suspended as shown in Fig. 3.2(*a*), and it is desired to know the tension T in the ropes. Since the point P is in equilibrium, the force condition must hold when reference is made to this point as the "particle." Equation (3.1) is best applied graphically, as in Fig. 3.2(*b*). We draw the known downward arrow representing 100 lb. Then we must

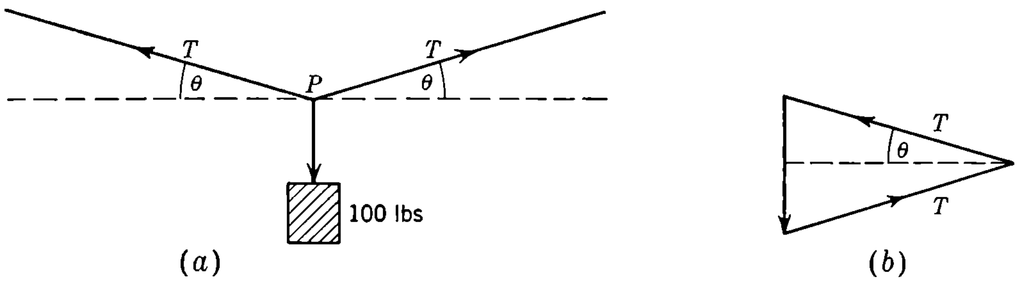


FIG. 3.2. Equilibrium of concurrent forces.

complete the triangle of forces by laying off the two tensions T assumed to be equal, at the angles that they make in Fig. 3.2(*a*), a procedure that determines the length of these arrows. It may be seen from Fig. 3.2(*b*) that $T \sin\theta = \frac{1}{2} \times 100\,\text{lb}$, whence $T = 50\,\text{lb}/\sin\theta$.

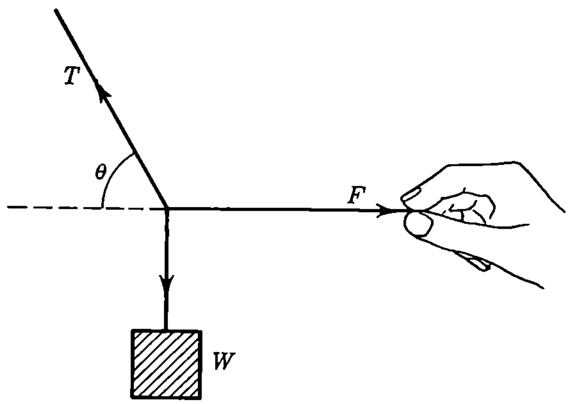
One may also solve this problem by using Eqs. (3.2). The three forces involved have the following components (we take X for the horizontal, Y for the vertical axis):

Force	X component	Y component
1	$-T\cos\theta$	$T\sin\theta$
2	$T\cos\theta$	$T\sin\theta$
3	0	-100 lb

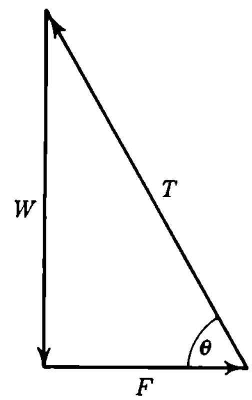
There are no Z components in this example. On adding the X components we obtain zero; hence the first of Eqs. (3.2) is automatically true and gives us no new information. The second yields $2T\sin\theta = 100$ lb, which is the previous result.

Notice that the smaller the angle θ the greater will be the force T. In fact T would be infinite if the ropes were actually horizontal. Here is the reason why telephone wires must not be too taut, since otherwise the collection of any weight on them (*e.g.*, ice) would cause an extremely large tension and consequent breaking.

As another example, consider the situation depicted in Fig. 3.3(*a*). Here both **T** and **F** can be determined when the weight **W** and the angle θ are known. The use of Eq. (3.1) is illustrated in Fig. 3.3(*b*). We lay down the vector **W**, which is known both as to magnitude and direction. Then the triangle is completed by drawing **F** and **T** in their proper directions; closure of the triangle determines the lengths of both **F** and **T**.



(*a*)



(*b*)

FIG. 3.3. Equilibrium of concurrent forces.

The trigonometric method, based on Eqs. (3.2), leads to the following tabulation of force components:

	X component	Y component
T	$-T\cos\theta$	$T\sin\theta$
F	F	0
W	0	$-W$

and to the equations

$$-T\cos\theta + F = 0$$
$$T\sin\theta - W = 0$$

When solved they give $T = W/\sin\theta$, $F = W\cot\theta$. These results can also be obtained from Fig. 3.3(*b*) directly.

As a third example we consider the forces acting on a particle that is at rest on a rough inclined plane. Since the particle is in equilibrium and we know it to be subject to the gravitational force **W**, the plane must be pushing upward on it with a force $-$**W** (cf. Fig. 3.4). (Note that $-$**W** is a force along the positive Y axis; this is because we have followed convention in calling the weight force, which is always directed downward, +**W**.) The force $-$**W** is called *the reaction of the plane* against the particle. Later we shall find it convenient to resolve this reaction in two perpendicular directions, along the plane and at right angles to the plane. Their magnitudes are $W \sin \theta$ and $W \cos \theta$, respectively (cf. Fig. 3.4). If friction is not strong enough to supply a force $W \sin \theta$ along the plane, the particle cannot be in equilibrium and will slide down the plane with an acceleration.

In the preceding examples attention has been restricted to *three* forces acting in the *same* plane (*coplanar* forces). More general cases can always be solved by the same methods. Note that, when the forces are not coplanar, all three equations (3.2) become significant. But before we treat this general situation it is well to enlarge our outlook in another respect: we wish to be able to deal with extended bodies for which, as we have seen, the force condition alone is not sufficient.

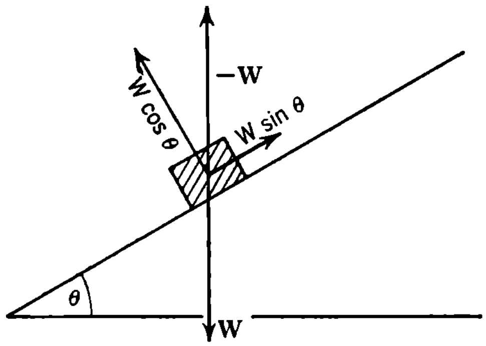


FIG. 3.4. Body at rest on an incline.

3.3. Torques, or Moments of Force. Figure 3.1 showed how two equal and opposite forces may produce, not equilibrium, but a tendency toward rotation. It now becomes necessary to introduce a *measure* for this tendency. Suppose you wish to close an open door, which obviously is a matter of producing rotation of an extended body. A given force applied at the handle is more effective than an equal force applied near the hinge. Hence the distance from the axis of rotation to the point at which the force acts is an important quantity in this connection. Furthermore a force applied at right angles is more effective than one acting at some other angle with the plane of the door.

These simple facts of everyday experience may be summed up in this way: Using the word *torque* (Latin *torquere,* to twist or rotate) or its synonym, *moment of force,* for the measure of the effectiveness of a force in producing rotation, we define: *A torque is the product of a force times the perpendicular distance between its line of application and the axis of rotation.* Thus, in Fig. 3.5, the torque due to the force **F** about an axis through the point A (and perpendicular to the plane of the paper) is

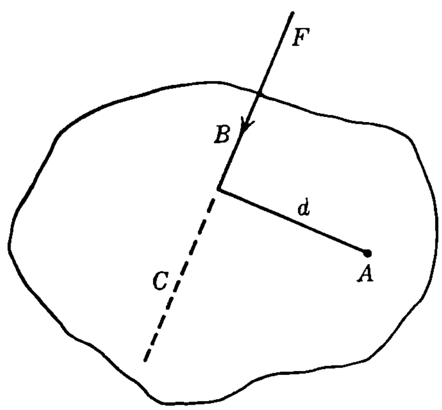


FIG. 3.5. Torque equals $F \cdot d$.

$$L = Fd \qquad (3.3)$$

We shall always use the symbol L for torques. A torque refers to an axis of rotation; there is no sense in speaking of the torque due to a force without reference to an axis. The distance d is sometimes called the lever arm of **F**. Notice that the *point* of application is immaterial so long as it lies on the dotted line of Fig. 3.5. If **F** were applied at C instead of B, it would produce the same torque about A.

It is sometimes useful to express the definition of L in other, equivalent ways. If we denote by r the distance (not perpendicular) between the axis and the point B at which **F** is applied (cf. Fig. 3.6), then

$$L = Fd = Fr \sin \theta \qquad (3.4)$$

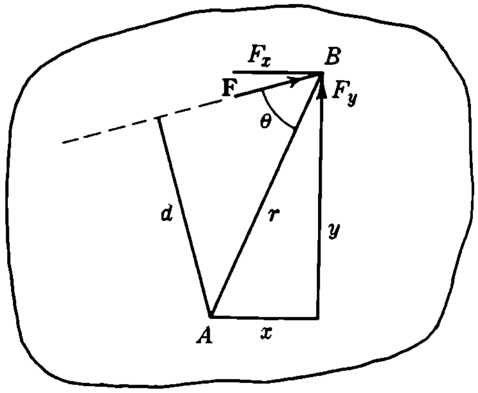


FIG. 3.6. Torque $= Fr \sin \theta = xF_y - yF_x$.

since $d = r \sin \theta$, θ being the angle swept out clockwise from r to F. There is, however, one important difference between the torque represented in Fig. 3.5 and that in Fig. 3.6; the former produces a *counterclockwise*, the latter a *clockwise* rotation. To indicate this difference we shall speak of counterclockwise torques as *positive* and of clockwise torques as *negative* torques. With this convention, we should therefore affix a negative sign to the right-hand side of Eq. (3.4).

Let us return to Fig. 3.6. The force **F** is equivalent to its two rectangular components F_x and F_y. The torque due to **F** must therefore be the same as that due to the two forces F_x and F_y acting about A. But if we take the origin of a coordinate system at A, then the rectangular components of r become x and y, these two torques are, respectively, xF_y and $-yF_x$, and we find

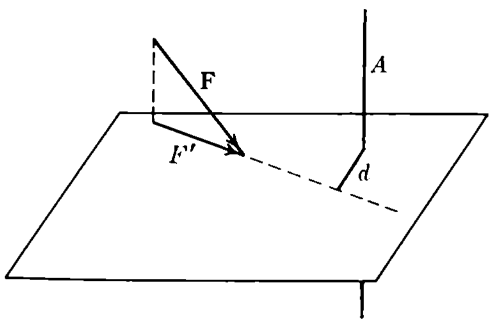


FIG. 3.7. General definition of torque: $L = F'd$.

$$L = xF_y - yF_x \qquad (3.5)$$

This result will be useful later.

Thus far we have not treated the most general case, for we have assumed that the force acts in a plane at right angles to the axis of rotation. Other cases will not interest us in this book. Nonetheless we shall give here the most general definition (cf. Fig. 3.7). If **F** does not lie in a plane perpendicular to the axis A, construct the perpendicular plane. Then draw the component of **F** in this plane, calling it $\mathbf{F}'$, and drop a perpendicular d upon its line of action. The torque is given by $F'd$.

From the definition of a torque it follows that its *units* are the pound foot, or the pound inch, or the ton foot, or the gram centimeter.

3.4. Equilibrium of an Extended Body; Noncurrent Forces. The conditions for equilibrium of an extended body are now easily stated. If the resultant of all forces were not zero, the body would have an acceleration; if the sum of all torques about *any* axis were not zero, there would be rotation about that axis. Hence the full conditions for equilibrium are

$$\left.\begin{aligned}\Sigma\,\mathbf{F} &= 0\\ \Sigma\,\mathbf{L} &= 0\end{aligned}\right\}\qquad(3.6)$$

The second equation has been written as if $\mathbf{L}$ were a vector, which is in fact true, as we shall see later. But what matters here is that this second equation must be satisfied for *any* axis we care to choose. If it is true for one, it may be shown to be true for any other parallel axis. Hence in applying the second equation, which we call the *torque condition for equilibrium*, we may select the axis that is most convenient.

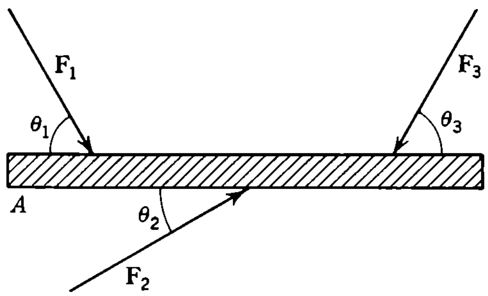


FIG. 3.8. Torques on a bar.

In the following we shall always be concerned with *coplanar* forces. Coordinate axes may therefore be chosen in the plane of the forces. Labeling them X and Y one can write Eqs. (3.6) in component form,

$$\left.\begin{aligned}\Sigma\,F_x &= 0\\ \Sigma\,F_y &= 0\\ \Sigma\,L_z &= 0\end{aligned}\right\}\qquad(3.7)$$

provided that we mean by L_z a torque due to a given force *about* the Z axis, which is taken at right angles to the plane of the forces. An example will clarify the meaning of these equations

Figure 3.8 shows three forces applied to a bar. Choosing a line perpendicular to the plane of the paper and passing through A as axis, and denoting the distances of the points of application along the bar from A by x_1, x_2, x_3, respectively, we get the components listed in the following table:

	F_x	F_y	L_z
F_1	$F_1 \cos\theta_1$	$-F_1 \sin\theta_1$	$-F_1x_1 \sin\theta_1$
F_2	$F_2 \cos\theta_2$	$F_2 \sin\theta_2$	$F_2x_2 \sin\theta_2$
F_3	$-F_3 \cos\theta_3$	$-F_3 \sin\theta_3$	$-F_3x_3 \sin\theta_3$

Equations (3.7) read

$$F_1 \cos \theta_1 + F_2 \cos \theta_2 - F_3 \cos \theta_3 = 0$$
$$-F_1 \sin \theta_1 + F_2 \sin \theta_2 - F_3 \sin \theta_3 = 0$$
$$-F_1 x_1 \sin \theta_1 + F_2 x_2 \sin \theta_2 - F_3 x_3 \sin \theta_3 = 0$$

Unless these are satisfied, the bar cannot be in equilibrium. The reader should note a purely mathematical fact not without practical importance. In the case of coplanar forces (regardless of how many forces) equilibrium implies three equations, no more. Since three equations can determine at most three unknowns, we can never determine the values of more than three quantities in an equilibrium problem. We shall observe an interesting consequence of this fact in our applications.

***3.5. Resultant and Equilibrant of Nonconcurrent Forces.** The resultant of a number of forces was seen to be a single force that is in all respects equivalent to

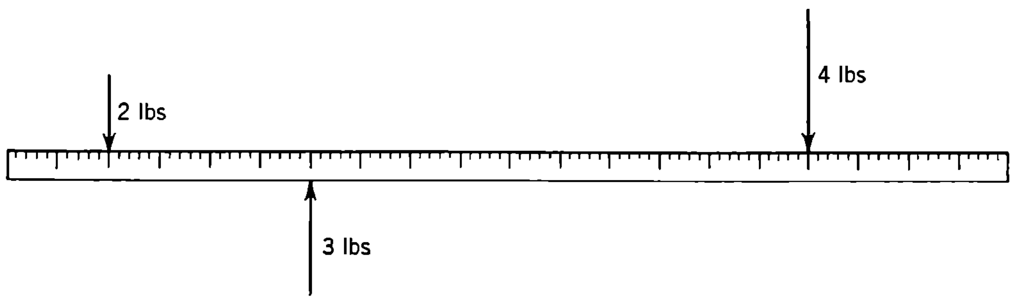


FIG. 3.9. Resultant of nonconcurrent forces.

them. It is the vector sum of the individual forces. This is true whether the individual forces act upon a small particle or upon an extended object. But in the case of an extended object it matters where the resultant force is applied! In other words, the resultant of a number of nonconcurrent forces is still the vector sum of these forces, but further consideration is needed in order to determine where the resultant is to be applied.

Clearly, if the resultant shall be equivalent to the individual forces, it must produce the same torque as do these forces. The distance from the axis of rotation is to be chosen so that this will be true. How this is done will be shown through an example. But first we emphasize this definition: The resultant of a number of nonconcurrent forces is a single force equal in magnitude and direction to their vector sum and producing the same torque as these forces. Thus the resultant of a set of nonconcurrent forces has a definite location in space as well as a definite magnitude and direction.

The negative of the resultant is a force that, when applied in the presence of the individual forces, would hold them in equilibrium. It is therefore called their *equilibrant.*

The three forces applied to a meter stick at distances 10 cm, 30 cm, and 80 cm, respectively, which are drawn in Fig. 3.9, have a resultant of magnitude 3 lb, directed downward. To be a true resultant it must be applied at a point, say x

cm from the left end, so that it produces the same torque as the three forces. Their total torque, however, is

$$-2 \text{ lb} \times 10 \text{ cm} + 3 \text{ lb} \times 30 \text{ cm} - 4 \text{ lb} \times 80 \text{ cm} = -250 \text{ lb cm}$$

Hence $\quad 3 \text{ lb} \times x = 250 \text{ lb cm} \quad$ and $\quad x = 83.33 \text{ cm}$

The *equilibrant* is therefore a force of 3 lb applied in an *upward* direction 83.33 cm from the left end. If it were added to the forces drawn, the bar would be in equilibrium.

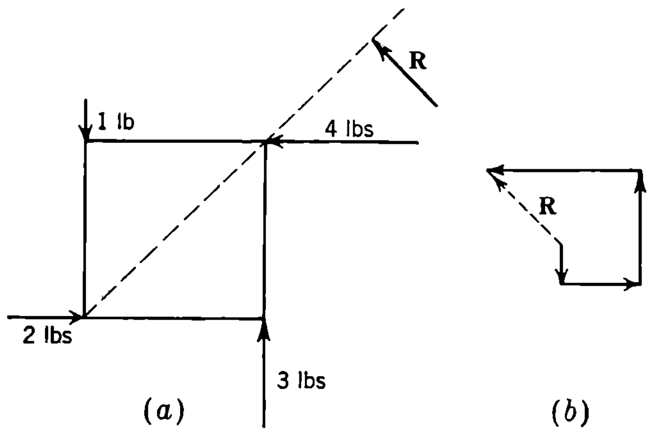


FIG. 3.10. Resultant of nonconcurrent forces.

Next we treat a slightly more complicated example. Suppose four forces are applied at the corners of a square of side length 20 in., as in Fig. 3.10*a*. From the polygon rule the student will find that their resultant is a force directed northwest and of magnitude $\sqrt{8}$ lb (cf. Fig. 3.10*b*). But where must it be applied? To find out we calculate torques about some axis, let us say through the southwest corner. The forces produce the following torques: 0×1, 0×2, 20×3, and 20×4 lb in., a total of 140 lb in. If the resultant is to produce the same torque about this axis, it must be applied a perpendicular distance x from the southwest corner, such that

$$\sqrt{8} \text{ lb} \times x = 140 \text{ lb in.}$$

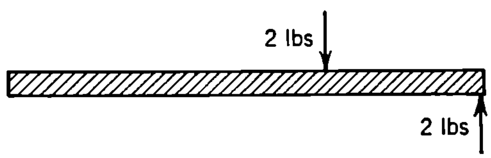


FIG. 3.11. The couple.

Hence $x = 49.6$ in. This resultant, marked **R,** is roughly indicated in Fig. 3.10*a*. The student should show that, if torques were calculated about any other axis, the resultant would still act at the same place.

***3.6. The Couple.** There are some curious exceptions to the rules just discussed. Certain arrangements of forces on extended bodies do not have a resultant at all! To see how this can be we look at Fig. 3.11, where two equal and opposite forces are shown acting on a bar at distances 2 ft and 3 ft from the left end, respectively. Their resultant has a zero magnitude; yet the torque, when computed about the left end, is +2 lb ft. If a resultant of zero magnitude were to produce this torque, it would have to be applied an infinite distance from the axis—which is, of course, absurd. Hence we must conclude that two equal and opposite forces which are not in the same line possess neither resultant nor equilibrant. They are called a *couple.* To produce equilibrium in a case like that of Fig. 3.11 it is necessary to introduce *two* other forces, *i.e.*, another couple.

A couple has another queer property; it has the same torque about all axes perpendicular to its plane. The student will convince himself of this fact by computing torques about other axes—for instance, the one through the right end of the bar in Fig. 3.11. This torque, which is independent of the axis, is sometimes called the *strength* of the couple. It equals the magnitude of *one* of the forces, multiplied by the perpendicular distance between them.

Every arrangement of nonconcurrent forces whose vector sum is zero is *equivalent* to a couple. For instance, four *equal* forces **F**, applied in the manner of Fig. 3.10 to the corners of a square of side length l, produce a torque of $2lF$ about any perpendicular axis and are equivalent to a single couple of that strength.

When parallel forces act upon a body, all of them in the same direction and sense, they can never constitute a couple. Such forces will therefore always have a resultant and an equilibrant.

3.7. Center of Gravity. The force of gravity on every small part of a rigid body is directed vertically downward. The resultant of all these

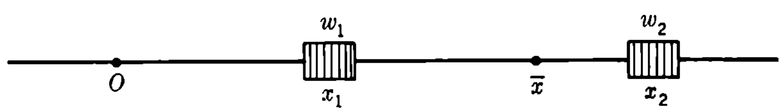


Fig. 3.12.

vertical forces is called the "weight" of the body. In accordance with what has been said, the weight must act through a definite point of the body, and this point is called its *center of gravity.* For the purpose of discussion we assume a weightless bar (cf. Fig. 3.12) to be carrying two small masses, one of weight w_1, the other of weight w_2. Let their distances from a fixed point O on the bar be x_1 and x_2. Their resultant, or the weight of the system of two masses, is $W = w_1 + w_2$. To yield the same torque, W must be applied at $\bar{x}$, such that

$$W\bar{x} = w_1x_1 + w_2x_2$$

Hence

$$\bar{x} = \frac{w_1x_1 + w_2x_2}{w_1 + w_2} \tag{3.8}$$

The point between the two masses that lies a distance $\bar{x}$ from the fixed point is the *center of gravity of the two masses.*

If the fixed point, *i.e.*, the origin from which distances are measured, is itself taken at the center of gravity, then $\bar{x} = 0$ and Eq. (3.8) indicates that

$$w_1x_1 = -w_2x_2$$

or

$$\frac{-x_1}{x_2} = \frac{w_1}{w_2}$$

Note that x_1 is now a negative quantity. Hence the center of gravity of two (point) masses lies at a point that divides the distance between the masses in the inverse ratio of their weights. This rule can often be applied without formal calculation.

The center of gravity of more than two distinct masses may be found by the same method, used repeatedly. If there are four masses, find the center of gravity between mass 1 and 2 and replace these two masses by a single one of weight $w_1 + w_2$. Next find the center of gravity between this fictitious mass and of mass 3. Place a mass of weight $w_1 + w_2 + w_3$ at this new center of gravity and compute the center of gravity between it and mass 4. This is the center of gravity of the four masses.

Let us return to the case where weights are arranged linearly, as at the beginning of this section; but assume there are n weights, not two. Let their positions, as measured from some fixed origin on the line, be $x_1 \ldots x_n$. Their resultant will be a weight $W = w_1 + w_2 + \cdots + w_n$. Equality of torques requires

$$W\bar{x} = w_1x_1 + w_2x_2 + \cdots + w_nx_n$$

and the center of gravity $\bar{x}$ becomes

$$\bar{x} = \frac{\sum_i w_ix_i}{\sum_i w_i} \tag{3.9}$$

if we use the summation convention explained in Sec. 2.10. This equation is a generalization of Eq. (3.8); it represents the center of gravity as the *weighted mean* of the distances of all masses from the origin.

Although we shall not be concerned in this chapter with the calculation of anything but linear-weight distributions, we remark here for the sake of completeness that three-dimensional arrangements of weights with coordinates (x_i,y_i,z_i) yield a center of gravity whose coordinates are $\bar{x}$, $\bar{y}$, $\bar{z}$, and these are given by formulas like Eq. (3.9), *viz.:*

$$\bar{x} = \frac{\sum_i w_ix_i}{\sum_i w_i} \qquad \bar{y} = \frac{\sum_i w_iy_i}{\sum_i w_i} \qquad \bar{z} = \frac{\sum_i w_iz_i}{\sum_i w_i}$$

The center of gravity of a uniform rod lies at its center. A rectangular slab may be regarded as composed of uniform rods, and by applying the rules already discussed one finds that its center of gravity lies at its center. This result is found to be generally true for uniform bodies; the center of gravity of a uniform body is at its geometric center.

For bodies of irregular shape this center is hard to calculate. But here experiment sometimes helps. If it is desired, for example, to determine the center of gravity of an irregular plane lamina, one proceeds as follows: Suspend it from some arbitrary point A (cf. Fig. 3.13), and draw a plumb line through A. Next suspend it from some other point B, and draw a plumb line through B. The center of gravity lies at the intersection of these lines.

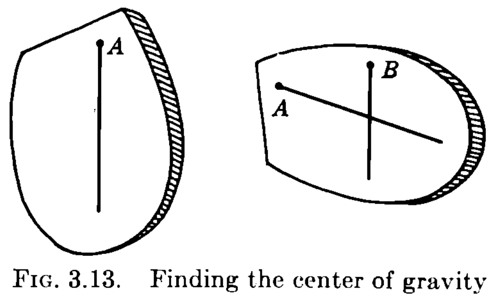


FIG. 3.13. Finding the center of gravity of an irregular lamina.

***3.8. Equilibrium of Extended Bodies. Examples.** Before continuing the student is asked to read again Sec. 3.4, the contents of which will now be used. Let us study the forces on a horizontal bar, pivoted at one end and supported by a string, as in Fig. 3.14. As a first step it is a good thing to *isolate* the body *on* which the forces are presumed to be acting. In the figure this is done by drawing an envelope around the bar. A moment's reflection will show that there are altogether three forces acting on the bar, its weight **w**, at the center of gravity, which is the center of the bar; the tension **T**; and the "reaction" **R** of the hinge against the bar.

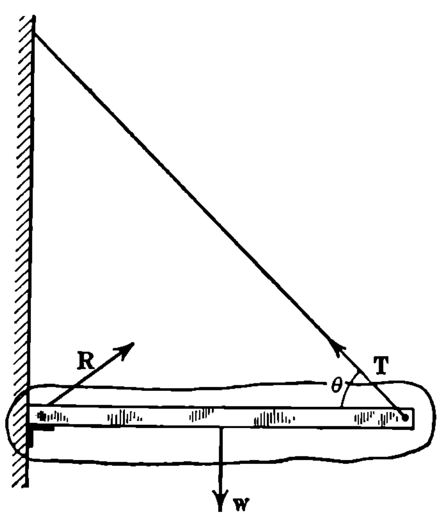


FIG. 3.14. Forces on a horizontal bar.

If we know **w**, **R** and **T** may be computed from the conditions of equilibrium [Eqs. (3.7)]. The vector **R** really represents the two unknowns R_x and R_y, for we do not know its direction. Computing force components and taking moments about the left end of the bar, we obtain

$$R_x - T\cos\theta = 0$$
$$R_y - w + T\sin\theta = 0$$
$$-w(l/2) + Tl\sin\theta = 0$$

if l is the length of the bar. From the last of these,

$$T = w/(2\sin\theta)$$

Hence, from the first two,

$$R_x = (w/2)\cot\theta \qquad R_y = w/2 \qquad R = \sqrt{R_x^2 + R_y^2} = w/(2\sin\theta) = T$$

In the present example these results could have been obtained more simply. Whenever *three coplanar* but nonparallel forces hold an extended body in equilibrium, *they must intersect in a point.* For if they intersected in three points, and we computed moments about any one of the points of intersection, one force would always produce a torque about an axis through that point and equilibrium could not result. For more than three forces a similar theorem does not hold.

On applying this fact to the previous example, it is clear that **R** and **T** must intersect vertically above the middle of the bar. Hence $R = T$, the three forces form an isosceles triangle, and all the results just obtained are seen to follow.

If the bar carried an additional weight W at its end, Eqs. (3.7) read

$$R_x - T\cos\theta = 0$$
$$R_y - w - W + T\sin\theta = 0$$
$$-w(l/2) - Wl + Tl\sin\theta = 0$$

Their solution is

$$T = \frac{W + (w/2)}{\sin\theta} \qquad R_x = [W + (w/2)]\cot\theta \qquad R_y = w/2$$

In the absence of W the angle α between the bar and **R** was equal to θ, since, in that case, $\tan\alpha = R_y/R_x = \tan\theta$. This is no longer true, for now

$$\tan\alpha = \frac{R_y}{R_x} = \frac{w}{2W + w}\tan\theta$$

and α is smaller than θ.

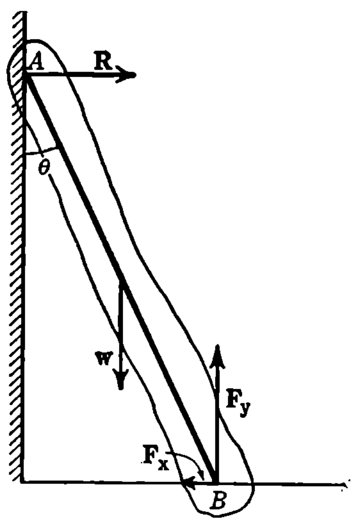


FIG. 3.15. Ladder problem.

As another example we treat the problem of a ladder leaning against a wall (cf. Fig. 3.15). The wall will be considered as *perfectly smooth*, which is another way of saying that it can exert no force tangential to its surface. Hence the reaction at A is at right angles to the wall. We start again by isolating the ladder (envelope), and then consider forces *on* it. In addition to **R** there is the weight, which, if the ladder is uniform, will act at its center. At B there will be a force **F**, which we at once resolve into components F_x and F_y. The equations of equilibrium are

$$R - F_x = 0 \qquad F_y - w = 0$$
$$-Rl\cos\theta + w(l/2)\sin\theta = 0$$

and their solution is

$$F_y = w \qquad F_x = R = (w/2)\tan\theta$$

These results can also be obtained by noting that the forces **R**, **w**, and **F** must intersect in a common point. Observe that **F** is not necessarily directed along the ladder.

If the wall is not smooth, the force **R** may extend in any direction. Instead of the previous three unknowns R_x, F_x, F_y, we then have four unknowns, namely, R_x, R_y, F_x, F_y. But we have only three equations at our disposal! At this point one is tempted to say that the problem cannot be done. Nevertheless, even in this peculiar situation the mathematics is trying to tell us something. *Three* equations can be satisfied by *four* variables in a great variety of ways; the problem is indeterminate. This means that, when the wall is rough, the ladder can be in equilibrium for any one of many choices of F_y and R_y, even for a given θ.

***3.10. The Beam Balance.** A beam balance consists of a rigid beam with pointer, pivoted at O (cf. Fig. 3.16), and two pans hinged at A and B. When equal weights are placed on the pans and the pans are of equal weight themselves, the beam is horizontal and the pointer is at some zero mark on the scale. Now

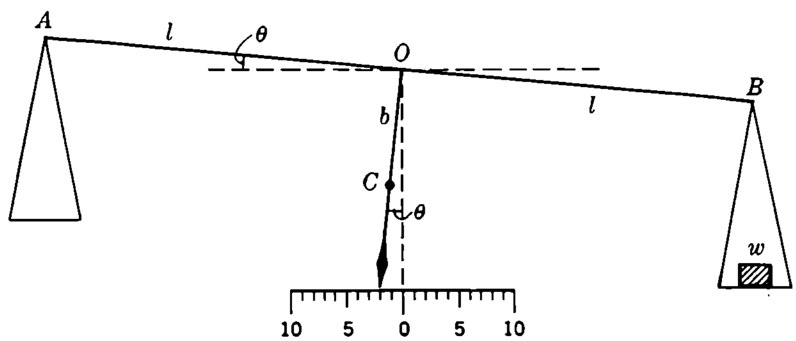


FIG. 3.16. The beam balance.

let a small weight w be placed on the right pan, so that a final deflection θ results. Assume the center of gravity of the rigid beam-pointer system to lie at C, and denote the distance OC by b, the weight of the rigid system by W. The weight of each pan is P. If we take moments about an axis through O when the deflection is θ as indicated in the diagram, the following torque equation is found:

$$Pl \cos \theta + Wb \sin \theta - (P + w)l \cos \theta = 0$$

since the system is in equilibrium. From this we see that

$$\tan \theta = \frac{w}{W} \frac{l}{b}$$

Now $\tan \theta$ is a measure of the number of divisions on the scale corresponding to the deflection θ. (Indeed the distance along the scale equals $\tan \theta$ times the distance between O and the scale.) The quotient $\tan \theta / w$ is proportional to the deflection *per unit weight w*. This expression is called the *sensitivity* of the balance; it is

$$S = \frac{\tan \theta}{w} = \frac{l}{Wb} \tag{3.10}$$

Equation (3.10) tells us how a very sensitive balance may be obtained. The balance arm l must be made as large as is feasible, the weight of the balance system, W, and the distance b must be very small. To minimize W and yet have l large requires delicate construction and design. The distance b is adjustable on some balances by means of a movable weight on the axis of the pointer, a device that allows S to be varied in accordance with the task to be performed. Because of the possibility of deformation of the beam structure, which may change the quantity b appreciably, the sensitivity is not always independent of the load.

If b were zero, the sensitivity would be infinite but the balance useless, for the slightest weight w would tip it completely.

A balance with two exactly equal arms is an ideal that is not attainable. For sensitive weighing, the difference in the lengths of the arms must be taken into account. This is done by the method of *double weighing*. The unknown weight W is placed first on the left pan, where it appears to be balanced by a weight W_1 on the right. It is then put on the right pan, where it is balanced by W_2 on the left. W_1 and W_2 will be nearly equal. How are they related to the true weight W? If the lengths of the arms are assumed to be l_1 and l_2, respectively, the torque equation in the first act of weighing is

$$l_1 W = l_2 W_1$$

and in the second

$$l_2 W = l_1 W_2$$

Multiply the two equations together to obtain

$$l_1 l_2 W^2 = l_1 l_2 W_1 W_2$$

or

$$W = \sqrt{W_1 W_2}$$

In words, the true weight is the *geometrical mean* of the two apparent weights. But if two quantities, like W_1 and W_2, differ by a very small amount, the geometrical mean is very nearly the same as the arithmetical mean. (The student should convince himself of this fact by numerical examples. See also Prob. 24.) Hence for practical purposes one may usually replace the last formula by

$$W = \tfrac{1}{2}(W_1 + W_2) \tag{3.11}$$

3.11. Machines. In designing machinery the engineer is everywhere confronted with the necessity of changing one force into another. The force of the steam on a piston in a cylinder must be converted into a force on the wheels of a locomotive, a force different in magnitude and in direction. The force of the revolving propeller must be changed into a lift and a propulsive force on an airplane. Even in the simplest circumstances problems of this sort are met; one wants to lift a rock by applying a smaller force on another object (*e.g.*, lever); one may wish to raise a heavy load by applying a downward force (pulley) because that is more convenient.

It is clear that the general subject of machines is an important and difficult one; it does not belong primarily to the province of physics although it involves nothing but physical and mathematical principles.

Our aim is therefore merely to arouse the student's interest in this field and to present a basic though elementary survey of it.

The *function* of a machine is always one of the following three simple tasks or a combination of them: to increase the magnitude of a force (lever), to decrease the magnitude of a force (lever), or to change its direction (single pulley). As to *composition*, every machine consists of combinations of the following elementary types of machine: (1) the lever; (2) the pulley; (3) the wheel and axle; (4) the inclined plane. These fundamental forms may appear in different guises but can always be recognized on close inspection.

Because machines transform forces, it is useful to have a measure of their ability to do so. This is called the *mechanical advantage*. To define it we recognize that, however a machine may be constructed in detail, it has one force applied to it, and it applies another. The latter is called the output force F_o, the former the input force F_i (see Fig. 3.17). The mechanical advantage is defined as their ratio

$$M = \frac{F_o}{F_i} \tag{3.12}$$

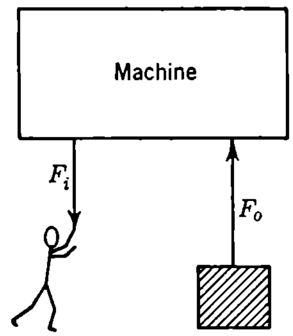


FIG. 3.17. Mechanical advantages equals F_0/F_i.

If it is smaller than 1, the machine *reduces* the applied force; if larger than 1 it *increases* the applied force. Many machines are reversible; that is to say, one can use either side as the input or output side. In that case, if one mode of use produces a mechanical advantage M, the other will yield $1/M$ (provided that we disregard friction; see below).

It is possible to calculate the mechanical advantage of every machine by the principles of equilibrium; but before we proceed to this task a little must be said about a very peculiar, though very common, kind of force that influences the performance of all machines, *viz.*, friction.

3.12. The Strange Force of Friction. When a weight rests on a table and the hand pushes against it slightly, the weight does not move; it is still in equilibrium. The force of the hand is being counteracted by an equal and opposite force on the weight. This can only be exerted by the table and must be directed horizontally. It is this *force of friction* that prevents the object from being accelerated by the force of the hand.

Now let the hand push a little harder: still the weight remains at rest. The force of friction has the remarkable quality of adjusting its magnitude exactly to that of the force tending to produce motion. But this is not true indefinitely. If the hand pushes with sufficient strength, the weight *will* move, but with an acceleration corresponding to a force smaller than that applied. We describe this occurrence by saying that up to a certain limit, the so-called "maximum force of friction," friction is exactly equal and opposite to the moving force. Only when the latter exceeds that limit does motion take place, and even then the net force acting is the difference between the force applied and the maximum frictional force.

Friction possesses adaptability not only with respect to magnitude but also with respect to direction. For if the hand pushes the other way, friction, too, turns around and opposes it. *In all instances,* however, *it acts parallel to the surface.*

A moving body is also retarded by friction. Experiment shows that the retarding force on sliding objects is in general slightly smaller than the maximum force of friction on a stationary one. Hence it is necessary to make a distinction between the latter, which is called *static friction,* and the former, known as *kinetic friction* (Greek *kinein,* to move).

Friction, both in its static and in its kinetic aspect, arises from a variety of causes, not all of which are easy to understand. Surfaces, when viewed under a microscope, are far from flat, and the motion of one surface over another requires the overcoming of interlocking obstacles, the lifting of the moving body, and perhaps abrasion of irregular projections. Friction between metals may even involve local melting of small portions in the surface. In addition to all this, there are intermolecular forces attracting one surface to the other, and these must also be considered in some instances. Fluids are used to reduce friction, and the science of lubrication, a very young one, is a fascinating but complicated subject. In general a lubricant spreads itself over the surface, smooths its irregularities, and causes one film of lubricant to slide over another film, thereby reducing friction.

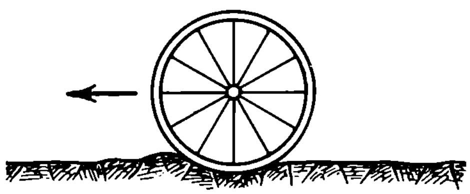


FIG. 3.18. Rolling friction.

Friction is often a most beneficent agency, without which few machines could work. No amount of engineering could construct a wheeled vehicle if the friction between tire and road were not present. Needless to say, the force of friction is different for different pairs of surfaces.

The rolling of a wheel is also impeded by a force, but for different reasons. No surface is completely rigid, and a wheel will make a slight impression on it. As the wheel moves forward, it has to overcome continually a slight hump in front of it (see Fig. 3.18), and the effect is the same as if it were rolling uphill. When a pneumatic tire wheels along, the rubber is constantly deformed, an action that is tantamount to the process depicted in the figure and also gives rise to *rolling friction.*

Later, in Sec. 6.6, we shall see how the force of friction can be measured quantitatively.

★3.13. Simple Machines and Their Mechanical Advantages. *a. The Lever.* Everybody knows the lever from his youthful experiences with the seesaw. If a lever is used as a machine to lift a load, as in Fig. 3.19, it has a mechanical advantage that is easily calculated from the laws of equilibrium. When F_i just balances the load F_o and moments are taken about the fulcrum, we find

$$F_o l_1 \cos\theta - F_i l_2 \cos\theta = 0$$

or

$$M = \frac{F_o}{F_i} = \frac{l_2}{l_1} \tag{3.13}$$

There are many ways of using levers, and levers are sometimes divided into classes. In every instance, however, M can be computed by the method here explained.

A word of caution must now be inserted. In deriving Eq. (3.13) we have neglected friction altogether. This force, which opposes motion and therefore effectively decreases F_o, makes the actual M smaller, as can be seen from Eq. (3.13). In mechanics it is often necessary to distinguish the *actual* mechanical advantage M_a from the *theoretical* one, which is given by Eq. (3.13). There is no simple way of calculating M_a; it must be determined by measuring the actual F_o and the actual F_i and then forming their ratio. But we should remember that M_a is always smaller than M. This is true also when the lever is used the other way around so that its M is smaller than 1. Henceforth we shall always be concerned with the theoretical M.

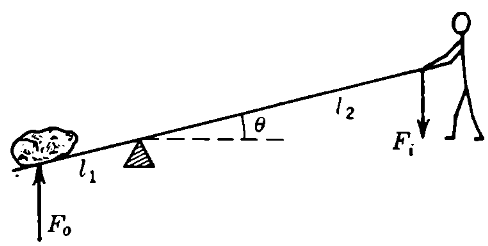


FIG. 3.19. The lever.

b. The Pulley. A single pulley, as depicted in Fig. 3.20*a*, has a mechanical advantage 1 and serves merely to change the direction of a force. Because of friction in the pulley axle the tension in the rope on the left of the pulley will not in general be equal to that on the right; since, however, we are ignoring friction in our present study, we assume, and shall assume in what follows, that *tension is transmitted undiminished from one side of a pulley to the other.* This will be true only when there is no friction in the pulley bearing and when the rope is perfectly flexible; it is usually a good approximation to reality.

A *block and tackle* is a pulley system arranged as shown in Fig. 3.20*b* and used to lift a weight. The input force F_i is transmitted to all parts of the rope, and by counting the rope strands supporting the weight we find that $F_o = 4F_i$; hence $M = 4$. For a machine of this type, M is equal to the number of parallel ropes supporting the load.

Nature, in building the animal organism, makes use of levers very profusely but abstains entirely from employing free wheels and pulleys, which are so prevalent in man-made machinery.

c. Wheel and Axle. The wheel and axle are shown in Fig. 3.21; the student will have no difficulty in proving, by taking torques about the axis of the system,

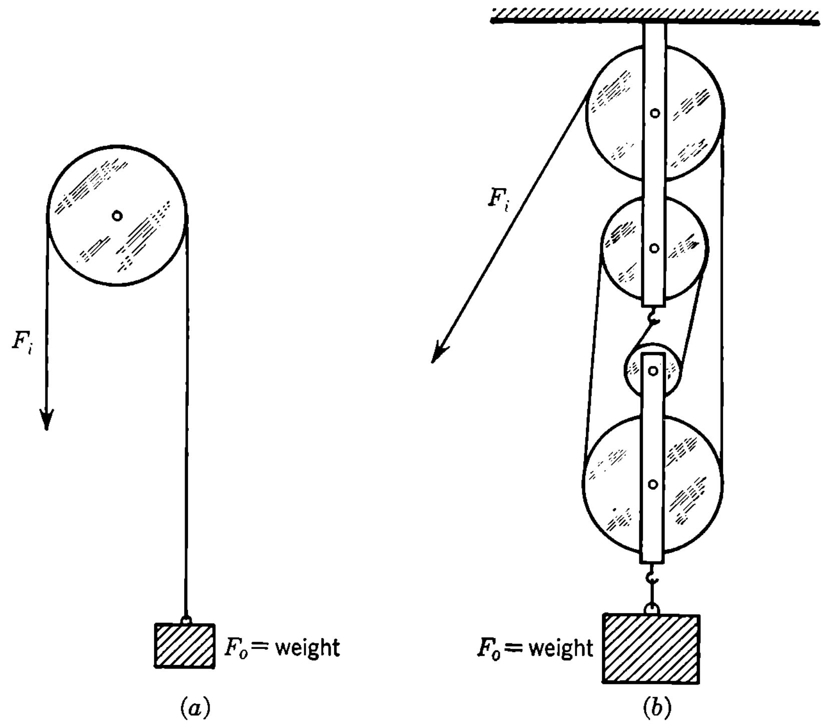


FIG. 3.20. (*a*) Single pulley. (*b*) Block and tackle.

that $M = r_2/r_1$. An interesting modification of the wheel and axle is the *differential pulley*, explained in Fig. 3.22. The smaller pulley is rigidly attached to the large one, and both turn about O. By pulling down on the rope at the extreme right (F_i) the weight can be lifted. To calculate the mechanical advantage, we note that the system is in equilibrium and consider forces and torques on the top pulleys of radii r_1 (smaller) and r_2. The sum of the torques about an axis through O is

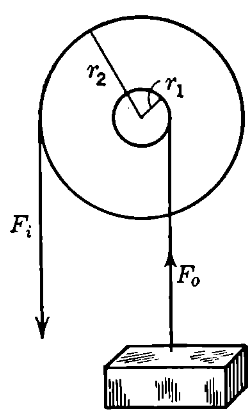


FIG. 3.21. Wheel and axle.

$$\frac{F_o}{2} r_2 - \frac{F_o}{2} r_1 - F_i r_2 = 0$$

whence

$$F_o = \frac{2r_2}{r_2 - r_1} F_i$$

and

$$M = \frac{2r_2}{r_2 - r_1}$$

As a rule the pulleys in Fig. 3.22 are toothed wheels, and the rope is replaced by a chain to avoid slipping when heavy loads are to be lifted.

d. Inclined Plane. Truck drivers know the advantage of using an inclined plane for loading a truck. They wish to lift a weight $W = F_o$ but manage to do so by applying a force F_i along the incline (Fig. 3.23). Now F_i is the *component* of **W** along the plane, and we have previously (see Fig. 3.4) seen this to be $W \sin \theta$. Hence

$$M = \frac{1}{\sin \theta} = \frac{l}{h}$$

the length of the incline divided by the height to which the weight is to be carried. In splitting wood the incline is used in the form of a *wedge*.

A *screw* is an inclined plane in disguise, for it may be regarded as an incline wrapped around a cylinder, the thread being the slanting side. To compute its mechanical advantage one needs to know how the screw is used, for it is rarely turned directly by hand. In the case of the *screw jack* (Fig. 3.24) the vertical force F_o causes a force $F_o \sin \theta$ parallel to the thread, and $\sin \theta \cong p/2\pi r$ if[1] we denote the pitch of the screw (distance between successive threads) by p and its radius by r. The screw is thus subject to a torque $F_o \sin \theta \, r = F_o p/2\pi$,

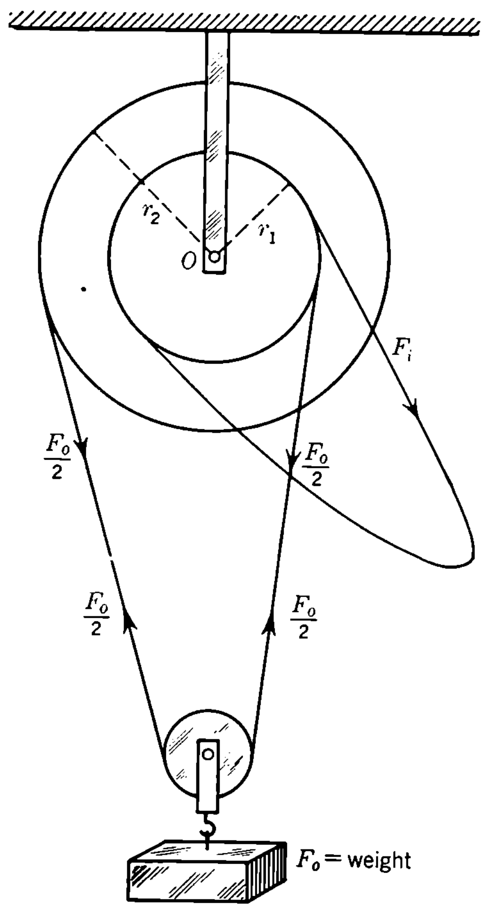


FIG. 3.22. Differential pulley.

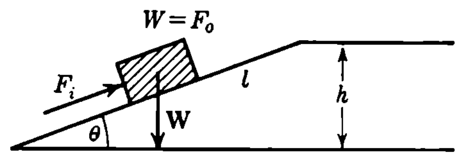


FIG. 3.23. Inclined plane.

and this must be equal and opposite to the torque $F_i l$ supplied by the hand. On equating the two we find for the ratio F_o/F_i the value

$$M = 2\pi l/p$$

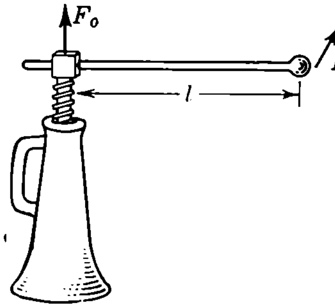


FIG. 3.24. Screw jack.

We wish to remind the reader that, in all examples treated here, M represents the *theoretical* mechanical advantage. The ratio M must not be confused with the efficiency of the machine, a term that will be explained in Sec. 8.10. Suffice it to say here that the efficiency which is defined in Sec. 8.10 may be shown to be the ratio of the *actual* mechanical advantage to the *theoretical* mechanical advantage,

$$\text{Eff} = \frac{M_a}{M}$$

This is never greater than 1.

[1] The symbol $\cong$ is used throughout this book to denote approximate equality.

PROBLEMS

1. A hammock extends between two trees, 8 ft apart. The length of hammock and ropes (cf. Fig. 3.25) is 10 ft. A person weighing 150 lb sits at the center of the hammock. What tension will he produce in the ropes?

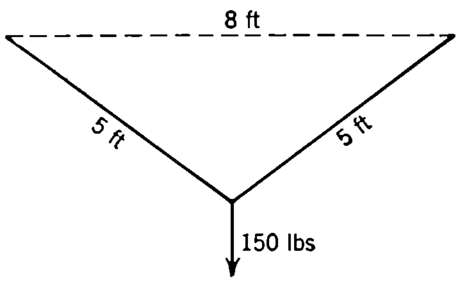


FIG. 3.25. Problem 1.

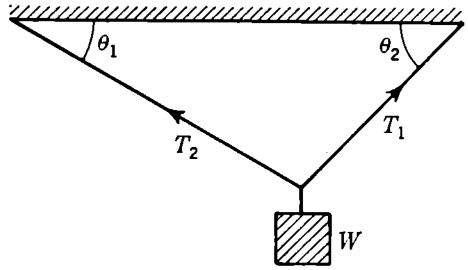


FIG. 3.26. Problem 2.

2. A weight is suspended as shown in Fig. 3.26. Show that the tensions in the two strings are

$$\frac{W \cos \theta_2}{\sin (\theta_1 + \theta_2)} \quad \text{and} \quad \frac{W \cos \theta_1}{\sin (\theta_1 + \theta_2)}$$

respectively. If $W = 500$ lb, $\theta_1 = 45°$, and $\theta_2 = 60°$, compute the two tensions.

3. A boy of 80 lb, sitting in a swing, is pulled sidewise until the rope supporting him makes an angle of 20° with the vertical. Find the tension in the rope and the force pulling sidewise.

4. A bracket supports a weight of 1,000 lb (cf. Fig. 3.27). Find the compressional force **C** along the diagonal strut.

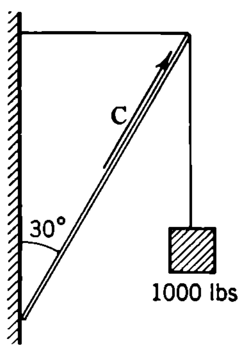


FIG. 3.27. Problem 4.

5. A truss (cf. Fig. 3.28) supports a weight W. Find the thrust F along each beam. (Note similarity to hammock problem!)

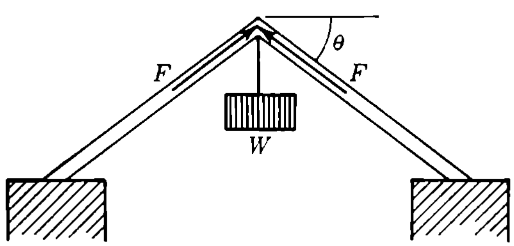


FIG. 3.28. Problem 5.

6. To pull a car out of a ditch requires a horizontal force of 1 ton (cf. Fig. 3.29). To supply this a rope is tied to a tree, and a force **F** is exerted as shown. How much force **F** is needed to move the car?

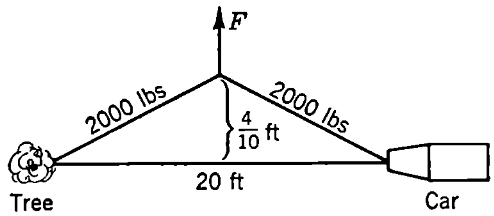


FIG. 3.29. Problem 6.

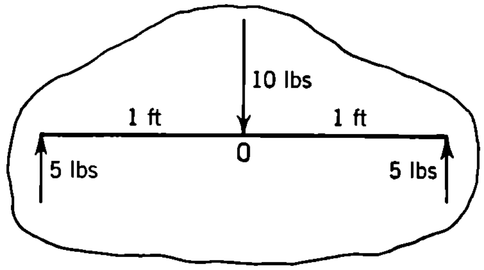


FIG. 3.30. Problem 7.

7. Figure 3.30 shows a set of forces that have a zero total torque about the point O. Select any other axis perpendicular to the plane of the paper, and show that the sum of all torques about this axis, too, is zero.

***8.** Calculate the resultant and the equilibrant of the following vertical forces, all acting on a horizontal bar at the given distances from its left end:

+1 lb	−2 lb	+4 lb	−6 lb	+8 lb
10 in.	15 in.	20 in.	30 in.	45 in.

9. Forces of 1, 2, 3, 4, 5, 6 lb act at the corners of a regular hexagon, each parallel to an edge in the manner of Fig. 3.10. The side of the hexagon is 1 ft. Calculate the resultant of these forces. Where does it act?

***10.** Show that six equal forces **F**, acting at the corners of a regular hexagon in the manner of Prob. 9, are equivalent to a couple of strength $3\sqrt{3}\,lF$, l being the length of one side.

11. Find the center of gravity of three equal masses placed at the corners of an equilateral triangle. Does the answer depend on the position of the triangle?

12. The corners of a weightless square carry successively weights of magnitude 1, 2, 3, and 4 lb. Find the center of gravity. Does it matter whether the square is vertical or horizontal?

13. Find the center of gravity of the three masses arranged as in Fig. 3.31.

14. The eight corners of a cube of side length 4 ft carry the masses shown in the figure. Find their center of gravity.

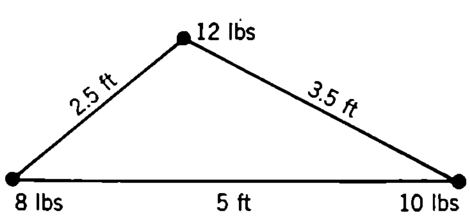


FIG. 3.31. Problem 13.

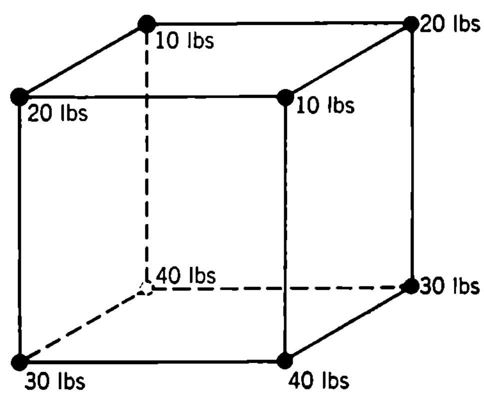


FIG. 3.32. Problem 14.

15. Find the center of gravity of a flat piece of metal having the shape given in Fig. 3.33.

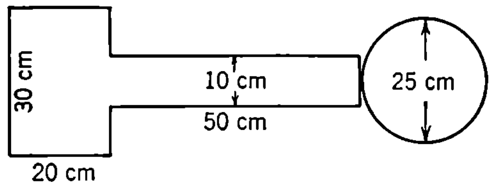


FIG. 3.33. Problem 15.

16. Find the forces in the ladder problem of Sec. 3.9 by utilizing the fact that the three forces must pass through a common point.

17. A ladder weighing 100 lb rests against a smooth wall, making an angle of 75° with the ground. A man of weight 180 lb stands two-thirds of the way up the ladder, which has a length of 20 ft. Find the forces exerted by the wall and the ground upon the ladder. Does the length of the ladder matter?

18. Calculate the tension in the cable and the horizontal and vertical forces on the hinge of the derrick in Fig. 3.34. (NOTE: Consider forces on the boom.)

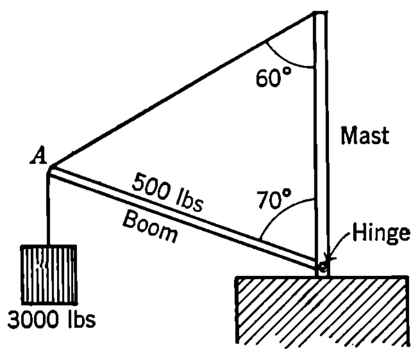


FIG. 3.34. Problem 18.

19. By considering the equilibrium of the point A in Fig. 3.34 determine the thrust of the boom on point A.

20. A telephone pole of height 30 ft at the end of a line supports eight wires, each having a tension of 30 lb. At the pole these wires are inclined at 4° with the horizontal. On the opposite side a guy wire is attached to the top of the pole and fastened to the ground at a point 20 ft from its base. Assuming the pole to be under no horizontal tension, find the tension in the guy.

21. An automobile of 130 in. wheel base registers a weight of 1,800 lb under the front wheels, a weight of 1,500 lb under the rear wheels. Find the distance of its center of gravity from the front axle.

22. A meter stick weighing 200 gm supports four weights—1 kg at 10 cm, 3 kg at 30 cm, 5 kg at 60 cm, 2 kg at 90 cm. Where must it be supported in order to balance?

★23. A balance has a sensitivity of 0.05 per mg. Through what angle will it deflect under a load of 10 mg?

★24. An object weighs 0.346 gm on the left and 0.351 gm on the right pan of a sensitive balance. What is its true weight computed as the geometric mean? As the arithmetic mean? What is the ratio of the lengths of the balance arms?

★25. Using the (binomial) theorem that

$$(1 + x)^{1/2} = (1 + x/2) \text{ for } x \ll 1$$

prove that the geometric mean of two nearly equal quantities equals their arithmetic mean.

★26. A balance has a beam of length 40 cm and a weight 50 gm. It is desired to attach to it a uniform pointer of length 30 cm. The balance is to produce a deflection of 5° per mg or better. How much may the pointer weigh?

27. A man and a boy are carrying a 200-lb weight on a uniform 10-ft rod weighing 20 lb, each holding one end of the rod. Where must the load be placed if the man is to carry twice the load of the boy?

★**28.** Select three simple machines from the human anatomy, and consider their mechanical advantages.

29. A smooth circular cylinder rests, its axis horizontal, between two smooth planes that make angles of 70° and 20°, respectively, with the horizontal. Calculate the force exerted by each plane.

30. A smooth rectangular board of width 1 ft rests between two smooth walls as indicated in Fig. 3.35, the walls being perpendicular to the plane of the paper. Find the inclination of the board in equilibrium. If the board weighs 100 lb, find the forces exerted on the board by the walls.

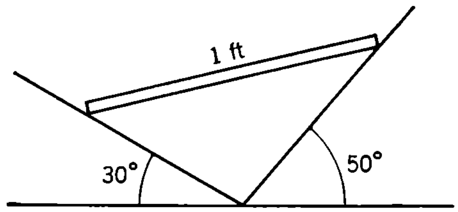


FIG. 3.35. Problem 30.

31. Pulleys are arranged as in Fig. 3.36. (*a*) Assuming the pulleys to be weightless, find the unknown weight w and the tension in all ropes. (*b*) Obtain the answers when each of the small pulleys weighs 20 lb, the large one 40 lb.

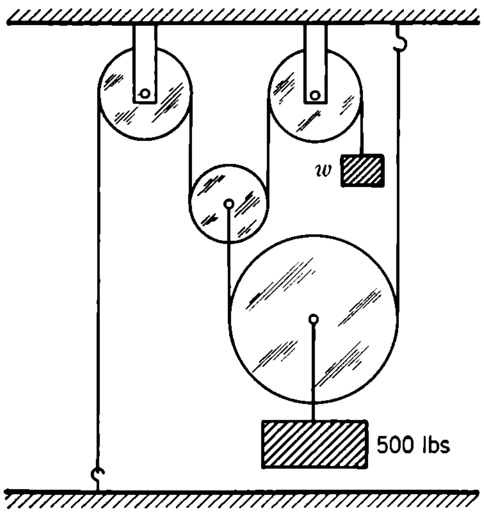


FIG. 3.36. Problem 31.

★**32.** A jackscrew with 4 threads to the inch is turned by a lever 1 ft long, to which a force of 100 lb is applied at right angles. What weight can it raise? (Neglect friction.)

★**33.** How much force must be exerted on a 20° wedge if it is to separate two surfaces held together by a force of 500 lb?

★**34.** A nut that requires a breaking force of 9 lb is placed 1 in. from the fulcrum of a 9-in. nutcracker. What force must be applied to its ends to crack the nut?

35. Two children weighing 120 lb and 90 lb wish to seesaw with a man who weighs 170 lb. The seesaw has its fulcrum at the center, and its beam is 15 ft long. If the man sits 1 ft from one end, find a way of disposing the weight of the children correctly.

CHAPTER 4 · ELASTICITY

4.1. Hooke's Law. In the preceding discussion we often encountered forces that were called *tensions* and *compressions*. Our interest, however, has been confined to the effect of these forces on *other* bodies, not on the body that is itself carrying the tension or compression. The present chapter deals with the changes caused by these forces in a body sustaining them. First, however, we must form a clear idea of their nature.

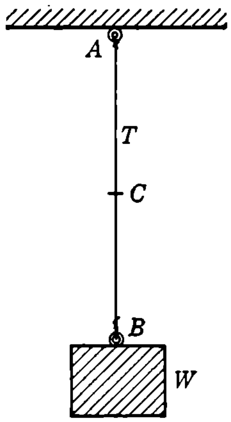


FIG. 4.1. Tension.

The wire in Fig. 4.1 is said to be under tension T, equal to the weight W. This means, first of all, that an external force, *viz.*, W, acts on the wire at B. It also means that the wire pulls with a force W on the ceiling at A. But as far as the wire itself is concerned, there is a force W across every cross section of it; if a cut be imagined at the point C or at any other point, the lower part of the wire would pull on the upper, and the upper part on the lower, with a force equal to W.

If a weight W rests on a vertical bar, the bar is said to be under compression, of magnitude W. Analysis of this situation yields the same results as the tension above, except that all pulls have become pushes.

The effect of a tension or a compression on a body is of course a change in its physical dimensions, and it is our aim to see what correlation there is between the force and the change in dimensions. Bodies which undergo a fairly regular deformation under the application of a tension or compression and which in addition return to their undeformed state when the force is removed are said to be *elastic*. But the term elastic is not one of high scientific precision and is not much used in quantitative work. It tends to be replaced by terms like "resilience," "tensile strength," and "stiffness," which convey much clearer physical ideas.

The connection between a tension and the longitudinal deformation that accompanies it was first formulated by Robert Hooke (1635–1703), a contemporary and antagonist of Newton. For fear of having his discovery "stolen" by unscrupulous scientists, he published it at the end of one of his lectures in the form of an anagram,

ce iii nosss tt uv

Later he divulged the proper order of the letters to be

"ut tensio sic vis"

(as the extension, so is the force). In more modern terms we state *Hooke's law* as follows: *Elongation is proportional to tension.* In this form the law refers to longitudinal deformations; it implies that a weight of 3 lb will stretch a wire three times as much as a weight of 1 lb, a fact that can easily be verified by experiment. Hooke's law is also applicable to compressions and indeed to other less familiar kinds of forces to be discussed presently. It is valid for all so-called "elastic" bodies *if the deformation is small*, but the range of validity differs for different substances (cf. Sec. 4.5).

A steel spring is an object that obeys Hooke's law particularly well, and even for large elongations. Since it is often used as a sort of model of an elastic body, a special terminology is employed in connection with it.

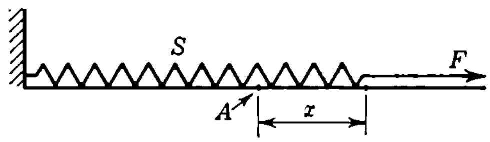


FIG. 4.2. Hooke's law.

The spring S (Fig. 4.2) is assumed to lie on a smooth horizontal surface and to be fastened to a rigid body at the left. The equilibrium position of its right end is A. Under a force F this end moves a distance x to the right. In accordance with Hooke's law, $x \propto F$, and this may be written

$$F = kx \tag{4.1}$$

The constant k in this equation is called the *stiffness constant* of the spring; it represents force per unit extension.

When the deformation is not a mere elongation, it has to be considered more carefully, which we now proceed to do.

4.2. Strain and Stress. In everyday speech, strain and stress are often used interchangeably; in physics a *strain is a measure of deformation; a stress measures the effectiveness of a force in producing deformation.*

An extended body can be deformed in many ways; a ball can be squeezed into almost any imaginable shape. But even in the most complicated sort of deformation, each small portion (of infinitesimal volume) of the body undergoes only two basic kinds of change: it may alter (1) its volume, (2) its shape, or both. Let us study these two *elementary* kinds of deformation in more detail.

1. For simplicity we consider a small cube. If all faces of it are subjected to equal forces acting perpendicularly upon the faces, as would be the case if the cube were submerged in a liquid, it will undergo a volume change only, as indicated in Fig. 4.3. If the original volume was V and

the final volume is $V - \Delta V$, the change in volume ΔV is, of course, a measure of the deformation. But it is advantageous to use the *ratio* $\Delta V/V$ for this purpose. We therefore define as the strain for this type of deformation, called the "volume strain," the dimensionless quantity

$$\text{Volume strain} = \frac{\Delta V}{V}$$

2. Next let the cube be deformed in such a way that its volume will remain unaltered but its shape will change. Clearly, this cannot be done by applying equal forces perpendicularly to two of its faces; it requires the action of a *couple.* The simplest case is depicted in Fig. 4.4, where two opposite sides are subjected to equal and opposite *tangential* forces. The cube is there drawn in section. A deformation of this type is called a *shear,* and it is measured by tan θ, θ being the angle of the shear. If θ is small, as will always be supposed, tan θ is very nearly equal to θ, and the shear, or shearing strain, is simply θ, *measured in radians.*

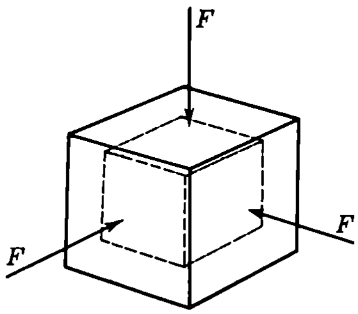


Fig. 4.3. Volume strain.

3. Shears and volume strains may be combined in many ways to give a more complicated form of strain. Among these there is one that, because of its frequent occurrence, may be singled out for special consideration. It is called a *stretch,* or longitudinal strain, and occurs as the result of a simple tension. When a weight is hung from a wire, the wire changes its shape (though it still remains a cylinder) as well as its volume (cf. Sec. 4.8). Figure 4.5 shows what happens. We define longitudinal strain as $\Delta l/l$, presupposing always, and in contrast with the figure, that $l \gg \Delta l$, that is, the stretch is small.

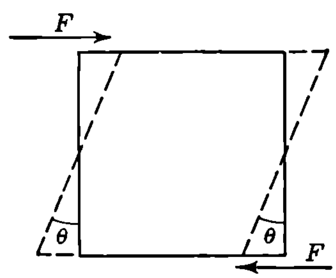


Fig. 4.4. Shear.

Having carefully explained the meaning of three types of strain, we now turn our attention to the corresponding stresses. *Stress is always defined as force per unit area;* it is therefore not simply a force. The following definitions are used by physicists and engineers.

In case (1), Fig. 4.3, the *volume stress* is taken to be $F_\perp/A$, where A is the area of one face of the cube and we have added the symbol $\perp$ as a subscript to F in order to remind the reader that F is perpendicular to the face A. In case (2), Fig. 4.4, the *shearing stress* is taken to be $F_\parallel/A$; A is again the area of one face, and $\parallel$ is to indicate that F is parallel to the face whose area is A. Finally, in case (3), Fig. 4.5, the *longitudinal stress* is defined as F/A, where now F is the total weight W applied to the wire and A is the cross section of the wire. In these definitions it is assumed that the forces F are distributed uniformly over the corresponding surfaces A, a condition that is usually met in practice.

A volume stress $F_\perp/A$ is also called a *pressure*. It is this kind of stress that is exerted by fluids upon surfaces immersed in them.

As to units, we note that all strains are ratios, or pure numbers; all stresses are clearly measured in $\frac{\text{lb}}{\text{in.}^2}$, or $\frac{\text{tons}}{\text{in.}^2}$, or $\frac{\text{gm}}{\text{cm}^2}$.

Ordinarily we think of a stress as a condition *inside* a body that has been deformed. Our definitions seem to depart from this idea, since they involve external forces producing deformations. The fact is, however, that the deformed body opposes the external forces with equal internal forces, so that the definition of stress here given may equally well be said to describe the internal stress condition of the body.

4.3. Elastic Moduli. In a somewhat generalized and more refined form, Hooke's law states that, for all types of small deformations, *stress is proportional to strain.* The ratio of stress to strain is therefore a constant for any given elastic object; furthermore, if the definitions of the preceding section are used, this constant is, within wide limits, independent of the size and shape of the object and dependent only on the characteristics of the substance of which it consists. This quotient, stress/strain, is called a *modulus.* Moduli have been given special names,

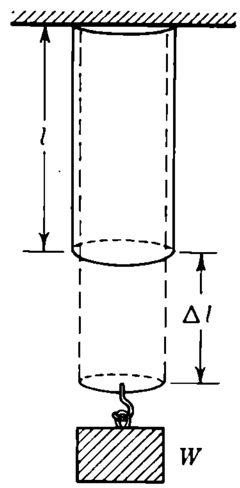


FIG. 4.5. Longitudinal strain or stretch equals $\Delta l/l$.

(1) $$\textit{Bulk modulus} = M = \frac{F_\perp/A}{\Delta V/V} = \frac{F_\perp V}{A\Delta V} \quad (4.2)$$

It indicates the force needed to compress the substance in question. Its reciprocal, $1/M$, is called the *compressibility.*

(2) $$\textit{Shear modulus} = N = \frac{F_\parallel/A}{\theta} = \frac{F_\parallel}{A\theta} \quad (4.3)$$

(3) $$\textit{Young's modulus} = Y = \frac{F/A}{\Delta l/l} = \frac{Fl}{A\Delta l} \quad (4.4)$$

Every modulus has the units of force/area, strain being a pure number. Table 4.1 gives some representative values of all three moduli for a number of substances. They are only approximate, for they vary greatly with composition in the case of alloys, and with temperature in general.

For fluids, *i.e.*, liquids and gases, it is more customary to list the compressibility (that is, $1/M$), expressed as "relative volume contraction per atmosphere," the atmosphere being a unit of stress, or pressure, equal to about 14.7 lb/in.^{2} (cf. Chap. 13). But we have listed M for comparison. Note, first of all, that the shear modulus as well as Young's modulus is zero for all liquids and gases. This is fairly obvious, for a fluid cannot sustain a shear; it takes practically no force whatever to deform a cube of fluid as in Fig. 4.4.

Furthermore, liquids are seen to have smaller bulk moduli, hence greater compressibilities, than solids, and the bulk modulus of gases is about a million times smaller than that of solids.

The question is sometimes asked whether a stress is the cause of a strain, or vice versa. A moment's reflection will convince the reader that, when a stress is produced in an elastic body, the strain will invariably follow. But it is also true that when a strain is produced, a stress will be set up at once. It is useless to look upon one as the cause and the other as the effect; stress and strain always accompany each other; they form an inseparable pair and should be recognized as such.

Table 4.1. *Elastic Moduli*
In lb/in.^{2}

	Bulk modulus M	Shear modulus N	Young's modulus Y
Aluminum	10×10^6	3.6×10^6	10×10^6
Brass	9×10^6	5.4×10^6	13×10^6
Copper	18×10^6	6×10^6	14×10^6
Gold	23×10^6	4.4×10^6	12×10^6
Iron:			
Drawn	22×10^6	12×10^6	29×10^6
Cast	13×10^6	7.5×10^6	17×10^6
Steel	28×10^6	11×10^6	30×10^6
Tin	7×10^6	3×10^6	7×10^6
Lead	1.2×10^6	1×10^6	2.5×10^6
Benzene	1.5×10^5	0	0
Carbon tetrachloride	1.5×10^5	0	0
Water	3×10^5	0	0
Mercury	4×10^6	0	0
Air (at atmospheric pressure)	14.7	0	0

***4.4. Worked Examples.** The theory of the preceding paragraph permits numerous applications.

a. How much bigger is a brass ball of volume 1 liter (1,000 cm^3) in a vacuum than in air? The air exerts a stress (or pressure) of about 15 lb/in.² Let us call the change in volume ΔV. Equation (4.2) then tells us that

$$\frac{\Delta V}{V} = \frac{F_{\perp}/A}{M} = \frac{15 \text{ lb/in.}^2}{9 \times 10^6 \text{ lb/in.}^2} = 1.67 \times 10^{-6}$$

where the value of M is taken from Table 4.1. The actual change in volume ΔV is therefore 0.00167 cm^3, or 1.67 mm^3, a small though not insignificant amount.

In a similar way, the student may show that a liter of benzene would expand 60 times as much.

b. A force of 1 ton is applied parallel to the upper surface, of area 2 ft^2, of a steel block ½ ft high. What lateral displacement will ensue? From Eq. (4.3),

$$\theta = \frac{F_{\parallel}/A}{N} = \frac{2{,}000 \text{ lb}/288 \text{ in.}^2}{11 \times 10^6 \text{ lb/in.}^2} = 6.31 \times 10^{-7} \text{ radian}$$

If the lower surface of the block is stationary, the upper moves through $6.31 \times 10^{-7} \times 6$ in. = 0.0000038 in.

c. How much weight is needed to elongate a 50-mil copper wire of length 6 ft by ½ in.? From Eq. (4.4),

$$F = Y\frac{\Delta l}{l}A = 14 \times 10^6 \text{ lb/in.}^2 \times \tfrac{1}{2} \text{ in.}/72 \text{ in.} \times \pi(0.025)^2 \text{in.}^2 = 191 \text{ lb}$$

provided that we remember that a mil = 0.001 in. (diameter).

Finally we calculate the stiffness constant of a helical spring that, under a weight of 3 lb, is elongated 6 in. Clearly, from Eq. (4.1), $k = \frac{1}{2}$ lb/in. Note that the units of k are different from those for the moduli.

4.5. Elastic Limit, Tensile Strength. It has been emphasized throughout this chapter that Hooke's law holds for small strains only. If the stress is increased indefinitely, an ordinary solid shows a behavior represented by the stress-strain diagram of Fig. 4.6. Up to strains corresponding to the point A the graph is linear, indicating validity of Hooke's law. Beyond that point curvature sets in, the modulus is no longer constant, and the material stretches more than it does in the initial stage. If the strain is carried to a point past A and the stress is then released, the material will not return to its original undeformed state; it has acquired a *permanent strain*, or *set*. The process has not been reversible. On the other hand, if the stress is further increased, the specimen will break at some strain B. Strain A is called the "elastic limit," strain B the "breaking point" of the material under test. For an ordinary metal under tensile stress, A corresponds to about 0.0005, B to

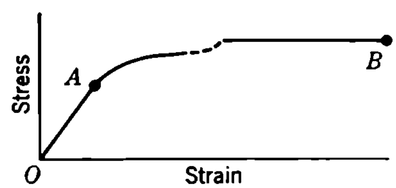


FIG. 4.6. Stress-strain diagram.

0.3. The inelastic range AB is seen to be very much larger on the strain axis than the elastic range OA. By *tensile strength* is usually meant the stress at the point B.

It is interesting to consider the physical mechanism that is responsible for the behavior exhibited in Fig. 4.6. Every solid consists of molecules, and these are regularly spaced, each occupying a point in definite relation to the others. For small strains the distances between all molecules are slightly altered, but forces between them draw them back into their original positions. Beyond the point A, large displacements occur among the molecules, and some of them, indeed, exchange places. Release of tension is then no longer able to restore the original condition, certain regions of the substance having been permanently deformed. At the

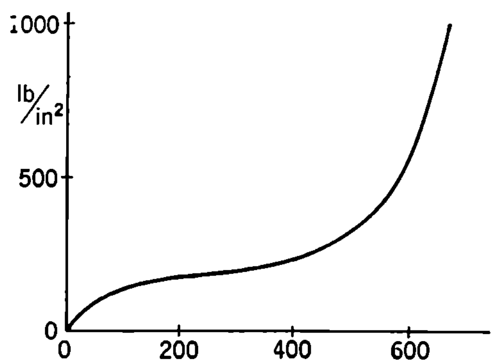


FIG. 4.7. Stress-strain diagram for natural rubber.

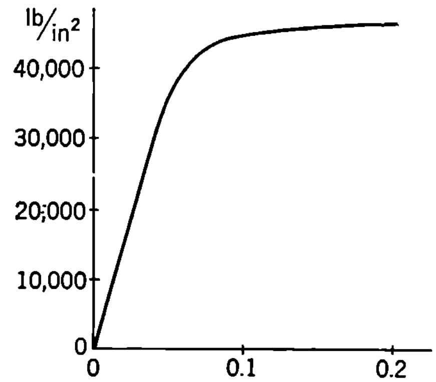


FIG. 4.8. Stress-strain diagram for iron wire.

point B, enough molecules have been severed from their neighbors so that the bonds between the remaining ones do not suffice to hold the substance together.

★4.6. Rubberlike Elasticity. Substances called rubbers exhibit properties that differ markedly from those so far described. Figures 4.7 and 4.8 show stress-strain diagrams for natural rubber and for an iron wire under tensile stress. First the reader should observe the vast differences in scale, indicating, of course, that rubber stretches very easily. But it is also clear that Hooke's law does not describe the strain of rubber in a significant way. Different rubbers such as neoprene and butyl have widely different stress-strain diagrams.

Another phenomenon, called *hysteresis* (Greek *hysterein,* to remain behind), is observed as a rule for rubberlike substances. The extension curve and the retraction curve do not always coincide. A typical example is shown in Fig. 4.9. It is due to the generation of heat in extension, which causes retraction to take place at a different temperature.

The physical processes which account for rubberlike behavior are quite different from those to which attention has been drawn in connection with ordinary solids. The molecules of rubber are long, flexible threads, which can be bent and twisted into any shape. In the normal state of a rubber sample the average length of the threads corresponds to their most probable degree of coiling. On stretching the threads uncoil into longer and less probable shapes, only to return to their most probable condition when the stress is released.

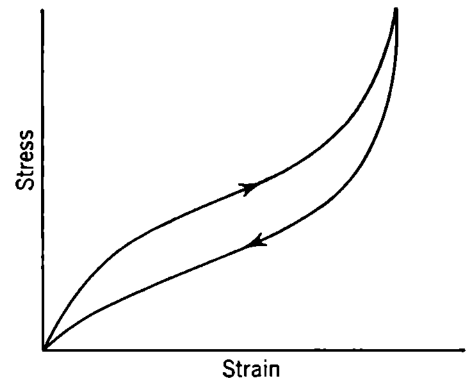


Fig. 4.9. Hysteresis.

★4.7. Stresses in Beams and Rods. When a beam bends under the application of a load, a complicated system of longitudinal stresses is set up. If the beam is considered as consisting of a large number of horizontal strips, some of these are lengthened while others are shortened. It is clear, therefore, that the bending of beams is controlled by Young's modulus. The detailed theory will not be given in this book, but a typical formula, useful in engineering, will be given without proof. If a beam of rectangular cross section, of length l, breadth b, and depth d, is subjected to a weight W at its middle while supported at its ends, the amount of sagging at the middle is

$$s = \frac{Wl^3}{4Ybd^3} \tag{4.5}$$

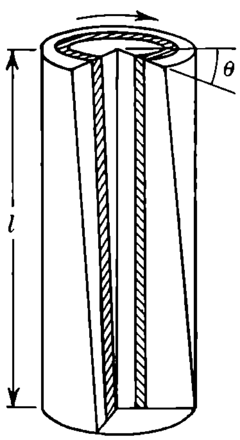


Fig. 4.10. Torsion.

This assumes the beam itself to be weightless. Under its own weight W it sags an amount

$$s = \frac{5Wl^3}{32Ybd^3} \tag{4.6}$$

When the ends of a rod or wire of circular cross section are subjected to torque, as is the case, for example, with the drive shaft of an engine, it suffers a deformation called a "twist" or a *torsion*. A strain of this type is completely reducible to a shear, as Fig. 4.10 will illustrate. The rod may be regarded as composed of cylindrical shells, one of which is shaded in the figure (where one quadrant is cut away to permit the strain to be exhibited). If each of these shells were flattened out, it would form a slab of rhombic shape, indicating a shear, and the angle of shear is the same for all slabs. The following formula may be derived for the torque L that, when applied to one end of the rod (the other being held firmly), will produce a shear θ:

$$L = \frac{\pi N r^4}{2l}\theta \tag{4.7}$$

Here r is the radius of the rod and l its length; the angle θ must, of course, be measured in radians.

To give an example, a 100-mil copper wire of length 2 ft has a torque of 1 lb in. applied at one end, the other end being clamped fast. According to Eq. (4.7),

$$1 \text{ lb in.} = 3.14 \times 6 \times 10^6 \text{ lb/in.}^2 \frac{(0.05 \text{ in.})^4}{2 \times 24 \text{ in.}} \theta$$

Notice that the units cancel on the two sides of the equation, and the shear or twist θ becomes 0.406 radian = 23.3°.

***4.8. Poisson's Ratio.** We return briefly to the longitudinal strain, or stretch, depicted in Fig. 4.5, which is controlled by Young's modulus. The longitudinal strain was seen to be $\Delta l/l$. There exists, however, in addition to the longitudinal strain a lateral strain responsible for the decrease in cross section of the rod or wire, and of magnitude $\Delta r/r$. The ratio of lateral to longitudinal strain, *viz.*,

$$p = \frac{\Delta r/r}{\Delta l/l} \tag{4.8}$$

is called *Poisson's ratio*, in honor of the physicist and mathematician S. D. Poisson (1781–1840), who introduced it. This is not an idle definition, for it happens that p is nearly independent of the magnitude of the strain and interestingly related to other elastic constants.

The ratio p is, in fact, related to Y and M. It may be shown that

$$p = \frac{3M - Y}{6M} \tag{4.9}$$

We are now in a position to answer the question as to whether the volume of a wire changes as a result of a stretch. Referring again to Fig. 4.5, the unstretched volume is $\pi r^2 l$; the stretched volume $\pi(r - \Delta r)^2(l + \Delta l)$. Calling the former V, the latter $V + \Delta V$, we find

$$\frac{V + \Delta V}{V} = \left[1 - 2\frac{\Delta r}{r} + \left(\frac{\Delta r}{r}\right)^2\right]\left(1 + \frac{\Delta l}{l}\right)$$

which may be approximated by $1 + (\Delta l/l) - 2\,\Delta r/r$ if $\Delta r/r$ and $\Delta l/l$ are both much smaller than 1 (cf. Sec. 1.7), which is assumed. Hence ΔV is positive when $\Delta l/l > 2\,\Delta r/r$, negative when $\Delta l/l < 2\,\Delta r/r$. In other words,

V increases if $p < \frac{1}{2}$
V decreases if $p > \frac{1}{2}$

But according to Eq. (4.9) p must always be smaller than $\frac{1}{2}$ since Y is a positive quantity. We may conclude, therefore, that stretch always increases the volume of an object.

PROBLEMS

1. An iron wire 6 ft long and 30 mils in diameter carried a load of 5 lb. A microscope with a scale was then focused upon the lower end of the wire, and its positions were read under additional loads, as follows:

Additional Load	*Position*
0	0
1 lb	6.0×10^{-3} in.
2 lb	12.1×10^{-3} in.
4 lb	23.9×10^{-3} in.
6 lb	36.1×10^{-3} in.
9 lb	75.8×10^{-3} in.
12 lb	137.8×10^{-3} in.
13 lb	172.7×10^{-3} in.

Plot a stress-strain diagram. Calculate Young's modulus.

2. A steel cable of diameter 1 in. supports a load of 10 tons. Calculate stress and strain in the cable. If its length is 100 ft, what is its elongation?

3. A 6-ft steel column of 5 in. diameter carries a weight of 20 tons. Compute stress, strain, and shortening.

4. The stress at the elastic limit of a 50-mil aluminum wire is 8 tons/in.² If its length is 10 ft, what maximum elongation can it stand without suffering a permanent set? What weight will produce this maximum elongation?

5. Calculate the elongations of the two wires in Fig. 4.11.

6. A copper wire and a steel wire, both having a cross section of 0.2 in.², and of lengths 2 ft and 3 ft, respectively, are fastened end to end, and then subjected to a tension of 200 lb. Calculate the elongation of each wire.

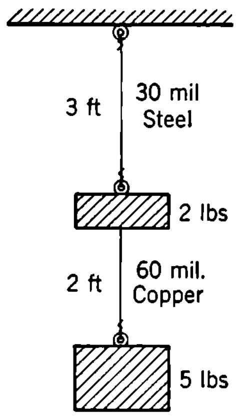


FIG. 4.11. Problem 5.

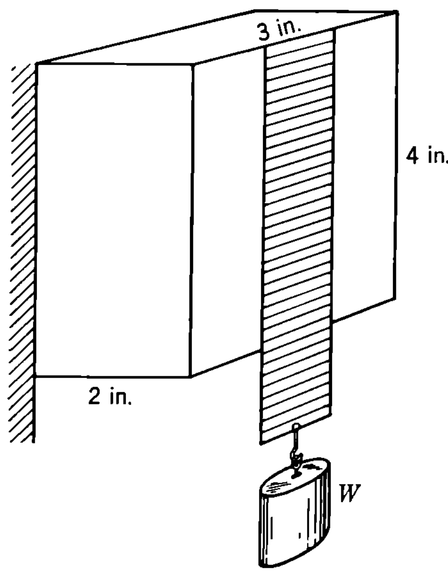


FIG. 4.12. Problem 9.

7. The ratio of the diameters of two wires of the same material is R. How much more will the smaller one be stretched under a given load?

8. Compute the stiffness of a spring that stretches 9 in. under a load of 1 lb. How much will it stretch under 50 gm?

9. A block of steel, rigidly fastened to a wall (see Fig. 4.12), has a strip brazed to its outer face, as shown. How much weight must be suspended from the strip to pull down the outer face of the block by 1 mm?

10. A force of 1 oz will displace the upper surface of a cube of jelly (side length 4 in.) by ½ in. Calculate the shear modulus for jelly.

★11. Shear moduli are usually determined by an arrangement like that shown in Fig. 4.13, W and θ being measured, and N then computed from Eq. (4.7). The rod has a twistable length l and a radius r. In an experiment the following values were found: $l = 3$ in., $r = \frac{1}{2}$ in., $R = 10$ in., $W = 10{,}300$ lb, $\theta = 10°$. Compute N.

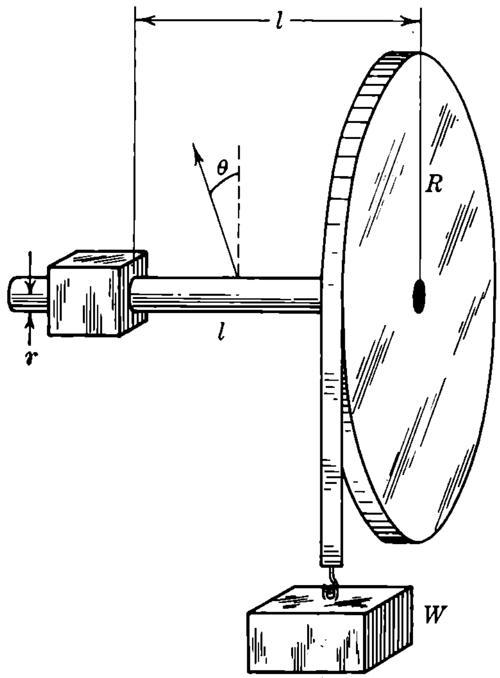


FIG. 4.13. Problem 11.

★12. How much weight is needed to twist the rod in Fig. 4.13 through 30°, if its $N = 6 \times 10^6$ lb/in.²?

★13. The steel drive shaft of an automobile is 5 ft long and has a diameter of 2 in. It is subject to a torque of 250 lb in. What is the angular displacement of one end with respect to the other?

14. The pressure (F per square inch) at the bottom of a lake is 150 lb/in.² Calculate the change in volume of a lead object normally occupying 1 gal when taken to the bottom of the lake. (One gallon = 231 in.³)

★15. Calculate the depression at the center of a steel beam of cross section 2 by 5 in., 10 ft long, and supported at its ends, when carrying a load of 1,000 lb at its center. (Neglect the weight of the beam.) (*a*) Take the depth of the beam to be 2 in. (*b*) Take the depth of the beam to be 5 in.

★16. Calculate the sag of the beam in Prob. 15 under its own weight, taking its density to be 0.28 lb/in.³

★17. Two beams of the same cross section, one having $b = 2$ in., $d = 4$ in. and the other with $b = d$, are loaded at the center. Compute the ratio of two loads that will produce the same deflection.

18. How may the values of the moduli listed in Table 4.1 be converted to kg/cm²?

19. Calculate the compressibility of aluminum, steel, water, and air in "relative volume change per atmosphere."

★20. Using Eq. (4.9), calculate Poisson's ratio for brass, gold, steel, and water. Show that the result for water means no volume change.

CHAPTER 5

DESCRIPTION OF MOTIONS; KINEMATICS

5.1. Introduction. To describe the motion of a body is to specify its successive positions in time. For an extended body this is not an easy thing to do, since attention must be paid to its orientation with respect to other objects as well as to momentary position of, let us say, its center of gravity. The matter becomes simpler when we consider small bodies, so-called "mass points," for which rotatory motion can be left out of account. In the present chapter consideration will therefore be limited to mass points.

The study of motions without reference to the forces that are responsible for these motions is called *kinematics* (Greek *kinein*, to move, as in "cinema"). Its purpose may be illustrated simply by means of an example. A stone is observed to fall freely. We know all about its motion if we can say where it is at any instant. This information, obtained by observation, may be collected in the form of Table 5.1, in which the first line lists successive instants after the motion has begun, the second the corresponding distances of the stone above ground.

Table 5.1

t	0	1 sec	2 sec	3 sec	4 sec	5 sec
y	400 ft	384 ft	336 ft	256 ft	144 ft	0 ft

Table 5.1 may be said to represent the kinematics of a falling stone.

But it is inconvenient to make a table of this sort for each kind of motion. The physicist therefore avails himself of more elegant methods provided by the mathematician. Two such methods can be used. The first is to make a *graph*, on which the distance from the ground y is plotted as ordinate against the time t as abscissa. The result is shown in Fig. 5.1. The other method is to find a mathematical function of t that will represent the distances in Table 5.1. How this can be done will be explained; for the present we shall merely state that the formula

$$y = 400 - 16t^2 \tag{5.1}$$

reproduces all the numbers in Table 5.1, as the student will verify.

Table 5.1, Fig. 5.1, and Eq. (5.1) are three equivalent ways of expressing the kinematical behavior of the stone. Of these, Eq. (5.1) is by far the most convenient. Usually, therefore, our desire will be to find the displacement as a mathematical function of t for any given kind of motion.

But the situation is not always so simple as that here discussed. If, instead of falling freely, the stone were thrown and hence traveled in a curved path, the entries in Table 5.1 would not suffice; for it then becomes necessary to specify the horizontal position (x coordinate) and the vertical position (y coordinate) for every instant t. The table then contains *three* rows, one for t, one for x, and one for y. The graphical representation will contain *two* curves, one in which x is plotted against t and one in which y is plotted against t. Finally we shall need two formulas like Eq. (5.1), one for x and one for y. For the present, however, we return to the simple case of rectilinear motion.

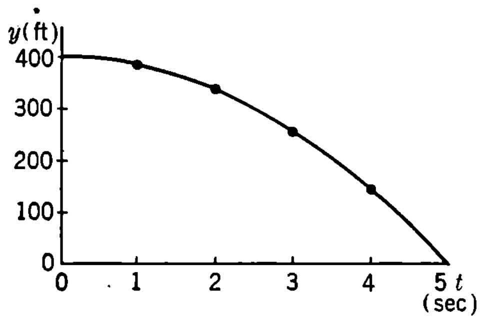


FIG. 5.1. Plot of the positions of a falling stone.

5.2. Velocity and Acceleration: Differentiation. Ordinarily, velocity is distance traveled per unit of time. It is measured in ft/sec, or miles/hr, or cm/sec, or km/min. This rather unsophisticated definition is quite satisfactory as long as the velocity does not change in time, *i.e.*, if the motion is *uniform*. For the motion described in Table 5.1 and for most other kinds of motion, this is not the case. We may then speak of the *average* velocity of the stone and compute it as the total distance traveled divided by the total time of travel. Using a bar over the symbol v to indicate average velocity, we thus obtain

$$\bar{v} = 400 \text{ ft}/5 \text{ sec} = 80 \text{ ft/sec}$$

But now it becomes clear that the average velocity, when computed for the whole 5 sec, is not the same as that for any of the partial intervals, for it is easily seen that the average velocity during the first second is 16 ft/sec, for the second 48 ft/sec, and so forth. In general, if t_1 and t_2 are any two instants, so that $t_2 - t_1$ is the time interval between them, and if y_1 and y_2 are the positions occupied by the stone at t_1 and t_2, respectively, the average speed during this interval is

$$\bar{v}_{21} = (y_2 - y_1)/(t_2 - t_1) \tag{5.2}$$

This formula can also be written the other way around:

$$y_2 - y_1 = \bar{v}_{21}(t_2 - t_1),$$

or, if we write s for the distance traveled, t for the time interval, and omit the subscripts on v,

$$s = \bar{v}t \tag{5.3}$$

To talk about average velocities is clearly not good enough in most problems of motion. It is necessary to make matters more precise by introducing the concept of *instantaneous* velocity. This is the *average velocity for a very small interval of time.* Let us see what happens to Eq. (5.2) when the difference $t_2 - t_1$ is made very small. It might at first seem as though $\bar{v}_{21}$ would then become very large; but this is erroneous, for $y_2 - y_1$ will also become very small, and the quotient remains, in fact, finite. Using the mathematical notation of a *limit,* we define the instantaneous speed v by the formula

$$v = \lim_{t_2 \to t_1} \frac{y_2 - y_1}{t_2 - t_1} \tag{5.4}$$

to be read: v is the limit, as t_2 approaches t_1, of the quotient

$$(y_2 - y_1)/(t_2 - t_1).$$

Another way of stating this definition is to write Δy for $y_2 - y_1$ and Δt for $t_2 - t_1$, so that

$$v = \lim_{\Delta t \to 0} \frac{\Delta y}{\Delta t}$$

The student will remember from his calculus course that the right-hand side of this equation is the *derivative* of y with respect to t,

$$\boxed{v = dy/dt}$$

This result shows the peculiar advantage gained by representing y as a function of t, as in Eq. (5.1). To obtain the instantaneous velocity, one simply differentiates y with respect to t. Hence, from Eq. (5.1),

$$v = -16 \times 2t = -32t \tag{5.5}$$

It will also be recalled that a derivative is a *rate of change;* hence velocity can be defined equally well as the *time rate of change of distance,* or the time rate of displacement. The negative signs in the last equation mean that y *diminishes* as time goes on.

The derivative of a function is known to be the *slope* of its graph. Hence the instantaneous velocity is the slope of the curve in Fig. 5.1. Uniform motion is represented by a function of constant slope: $y = \text{const} \times t$, and the constant, being the derivative, is the velocity.

The *unit* of instantaneous velocity is, of course, no different from that of average velocity. To say that the speed of a bullet is 50 miles/min implies that, *if* the bullet were to travel for 1 min at this velocity, it *would* go 50 miles.

The result of Eq. (5.5) shows that the velocity, too, is a function of the time. Whenever this is true the motion is said to be *accelerated* or non-uniform. The term "acceleration" will here be used to denote increase as well as decrease of velocity, although the latter is often called "deceleration." Instantaneous acceleration is the *time rate of change of velocity*,

$$a = \frac{dv}{dt} = \frac{d^2y}{dt^2} \tag{5.6}$$

Hence it is the *second* derivative of the position (or displacement) with respect to time. In the case of Eq. (5.5), $a = -32$ units; the negative sign indicates that the speed *increases* in the negative direction of y.

Since acceleration is change in velocity per unit of time, its unit must be a velocity unit divided by a time unit, such as the (foot per second) per second (abbreviated ft/sec^2) or the (centimeter per second) per second (cm/sec^2). Physically, we think of motion having an acceleration of a ft/sec^2 as motion in which the velocity increases by a ft/sec during each second.

In the motion of our example the acceleration turned out to be constant in time. This will not be the case in general. When it is true, as in free fall, the motion is said to be uniformly accelerated, or simply to have constant acceleration.

5.3. Integration. In the preceding section we have shown how velocity and acceleration may be found when the displacement is given as a function of the time, either in the form of a graph or in the form of a function. More important is the problem of finding velocity and displacement when the acceleration is given. This is because we often know the force acting on a particle, and by Newton's second law, discussed in Chap. 2, the force determines the acceleration. Let us consider, therefore, how we can pass from a knowledge of a to that of v and y.

The simplest case is that in which $a = 0$. For then $v = \text{const}$ and the displacement is vt. If v is not constant, another procedure must be used. Assume that an automobile starts from rest. An observer records speedometer readings at equal intervals—let us say every 10 sec—and it is desired to know the distance traveled after a certain period of time. First a table listing the displacements x for the various instants t is prepared. From the table a graph, like Fig. 5.2, is made. On this graph the ordinates drawn at 20 sec and at 30 sec, for example, are the velocities at these instants. The area of the shaded rectangle, being the product of the mean ordinate between these instants and the time interval, represents approximately the distance traversed during the 10 sec under consideration. If similar rectangles are drawn for all intervals, the sum of their areas is approximately the total distance covered. An uncertainty arises only because the mean ordinate is not well defined.

This uncertainty is eliminated if we subdivide the abscissas more and more finely, using a very large number of intervals and hence very narrow rectangles. But in that case the sum of the area of all the rectangles becomes identical with the area under the curve. Hence we conclude: *The area under the velocity curve between the instants* t_1 and t_2 *is the distance traveled between* t_1 *and* t_2. The student will recall from the calculus that this area is the integral

$$\int_1^{t_2} v\,dt$$

This, then, is the answer to our problem: To obtain the distance when v is given, we integrate with respect to the time. Thus

$$x = \int v\,dt \tag{5.7}$$

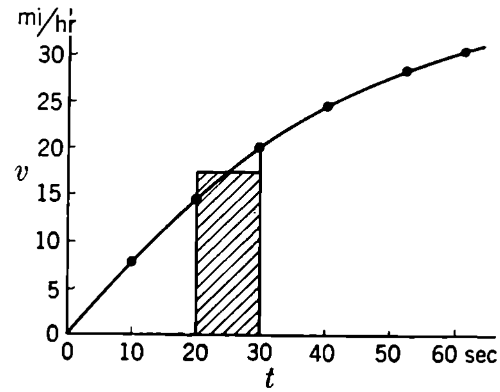


FIG. 5.2. Speed curve.

To make this integral definite we must insert the limits t_1 and t_2, and x will be the distance traveled between t_1 and t_2. More properly it is the difference between the displacement x_2 at the time t_2 and x_1 at t_1.

As x is related to v, so is v related to a. For a was defined as dv/dt; hence by inversion

$$v = \int a\,dt \tag{5.8}$$

This result, however, is not quite so simple as it looks, and care must be taken in its interpretation. If the limits of the integral are again t_1 and t_2, v is the increase in velocity between these instants, just as x in Eq. (5.7) was the increase in displacement that took place between t_1 and t_2. Hence it is better to replace Eqs. (5.7) and (5.8) by the following more explicit formulas:

$$\left.\begin{aligned} x_2 - x_1 &= \int_{t_1}^{t_2} v\,dt \\ v_2 - v_1 &= \int_{t_1}^{t_2} a\,dt \end{aligned}\right\} \tag{5.9}$$

Returning briefly to Fig. 5.2, we note that the slope of the curve at every instant is the acceleration. If a is plotted against t, the area under the curve will be the velocity.

★5.4. Worked Examples. *a.* The velocity of a train at various times is given in the following table:

Time	2:00	2:01	2:02	2:03	2:04	2:05	2:06	2:07	2:08
Velocity, ft/sec	0	8	16	24	30	34	37	38	38

It is desired to find the acceleration of the train and the distance it has traveled at any time.

On looking at the entries, several facts are at once apparent. During the first 3 min the train gains equal increments of velocity and therefore has constant acceleration, its value being 8 ft $\sec^{-1}$ $\min^{-1}$, which equals 480 ft/min^2 or $\frac{8}{60}$ ft/sec^2. After 3 min the acceleration becomes smaller, and after 7 minutes it has become zero. During the eighth minute the motion is uniform.

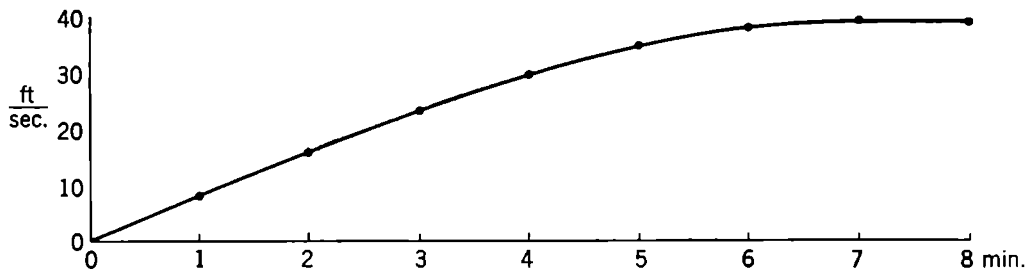


FIG. 5.3. Speed curve.

To get the distance covered we could proceed as follows: The *mean velocity* during the first minute is 4 ft/sec; hence the distance traveled in that minute is 4 ft/sec × 1 min = 240 ft. The mean velocity in the second minute is 12 ft/sec, and the distance covered is 720 ft. By adding up all these partial distances we obtain the total at any instant.

It is advantageous, however, to do this by means of a graph. The data of the table are plotted in Fig. 5.3. Up to the third minute the plot is straight, indicating that the slope, which is the derivative and hence the acceleration, is constant. The last part of the curve is straight and horizontal and therefore characteristic of uniform motion.

The distance is the area under the curve. Up to the end of the third minute we may compute it by using the formula for the area of a triangle; it is

$$\tfrac{1}{2} \times 24 \text{ ft/sec} \times 3 \text{ min} = 36 \text{ ft min/sec} = 36 \times 60 \text{ ft.}$$

To find the distance at a later time—where the curve is no longer straight—one would have to measure the area with a planimeter.

b. The work is simpler when the velocity is given as a mathematical function of t. Suppose that $v = c_0 + c_1t + c_2t^2$ where all the c's are known constants. We then have at once $a = dv/dt = c_1 + 2c_2t$. On the other hand the distance covered between the instants t_1 and t_2, which we may call $x_2 - x_1$, is seen to be

$$\begin{aligned} x_2 - x_1 &= \int_{t_1}^{t_2} v\,dt = \int_{t_1}^{t_2} (c_0 + c_1 t + c_2 t^2)dt \\ &= \left[c_0 t + \tfrac{1}{2} c_1 t^2 + \tfrac{1}{3} c_2 t^3 \right]_{t_1}^{t_2} \\ &= c_0 t_2 + \tfrac{1}{2} c_1 t_2^2 + \tfrac{1}{3} c_2 t_2^3 - c_0 t_1 - \tfrac{1}{2} c_1 t_1^2 - \tfrac{1}{3} c_2 t_1^3 \end{aligned}$$

If we are interested in the distance covered from the beginning, *i.e.*, since $t = 0$, we must put $t_1 = 0$ in the last result.

5.5. Integration (Continued). Students who are not conversant with definite integrals but who know that integration is the inverse of differentiation may obtain the same results in the following way: If $v = dx/dt$, then

$$x = \int v\,dt \tag{5.7}$$

Similarly, since $a = dv/dt$,

$$v = \int a\,dt \tag{5.8}$$

In order to evaluate these integrals the integrands [v in Eq. (5.7), a in Eq. (5.8)] must be given as functions of t. As an example, suppose $a = ct^n$, where n is some integer and c a given constant. Equation (5.7) then gives

$$v = \frac{ct^{n+1}}{n+1} + C$$

C being the constant of integration. If in this equation we put $t = 0$, it reads $v_{t=0} = C$ and this allows us to identify C with v_0, the speed at $t = 0$. Hence

$$v = c\,\frac{t^{n+1}}{n+1} + v_0 \tag{5.7'}$$

On substituting this into (5.7) we find

$$x = \int \left(c\,\frac{t^{n+1}}{n+1} + v_0 \right) dt = \frac{ct^{n+2}}{(n+1)(n+2)} + v_0 t + C$$

and by putting $t = 0$ we see that the present $C = x_0$. Hence we find

$$x = \frac{ct^{n+2}}{(n+1)(n+2)} + v_0 t + x_0 \tag{5.8'}$$

5.6. Free Fall. Prior to the time of Galileo Galilei it was believed that heavy bodies fall faster than light bodies. Galileo demonstrated—according to some accounts by dropping objects from the leaning tower of Pisa—that this belief is not substantiated by the facts. It is true, of course, and agrees with common observation that a feather falls more slowly than a steel ball, but this is not the case in a vacuum. Air resistance causes the difference: in a vacuum, all bodies fall "equally fast."

This does not mean, however, that bodies fall with constant velocity. Galileo's great discovery showed that they fall with *constant acceleration*. By painstaking observations made on inclines, which, as we shall see in Sec. 6.6, have the effect of reducing the acceleration to measurable values yet leaving it constant, Galileo determined the acceleration in free fall to be approximately 32.2 ft/sec², or 980 cm/sec². This value is usually denoted by g. It varies slightly from one location to another for reasons to be explained later. For present purposes we shall use the approximate values above.

The *laws of free fall* are now easily developed. Starting with

$$a = dv/dt = g, \text{ a constant} \tag{5.10}$$

we find at once by integration

$$v_2 - v_1 = \int_{t_1}^{t_2} g\,dt = g(t_2 - t_1) \tag{5.11}$$

If the body starts from rest, $v_1 = 0$ at the time $t_1 = 0$, and we have

$$v_2 = gt_2$$

which can be written

$$v = gt \tag{5.12}$$

since v_2 is the velocity at t_2 and there is no longer any need for retaining the subscript.

We can also integrate the equation $dv/dt = g$ by writing the indefinite integral

$$v = gt + \text{const} \tag{5.13}$$

provided that we do not forget the constant. The physical meaning of the constant of integration can always be seen from the context in which it appears. In the present instance the constant equals v when $t = 0$. This gives it away: it must be the initial velocity, for which we may write v_1 as above or, if we please, v_0. Upon using the latter choice, Eq. (5.13) takes the form

$$\boxed{v = v_0 + gt} \tag{5.14}$$

The distance covered is found by one further integration as seen from Eq. (5.7). Hence

$$y = \int (v_0 + gt)dt = v_0t + \tfrac{1}{2}gt^2 + \text{const}$$

or

$$y = y_0 + v_0t + \tfrac{1}{2}gt^2 \tag{5.15}$$

Observe also that Eqs. (5.14) and (5.15) could have been obtained from Eqs. (5.7′) and (5.8′) provided that we put $x = y$, $c = g$, and $n = 1$.

Equations (5.14) and (5.15) permit the calculation of velocity and distance in free fall at any *instant*. They will not answer the question: What velocity will a body acquire while it falls through a given *distance?* A formula providing the answer is obtained by eliminating the variable t between Eqs. (5.14) and (5.15). From Eq. (5.14),

$$t = \frac{v - v_0}{g}$$

When this is put into Eq. (5.15), there results

$$y - y_0 = \frac{v_0(v - v_0)}{g} + \frac{1}{2}g\frac{(v - v_0)^2}{g^2} = \frac{1}{2}\frac{(v^2 - v_0^2)}{g}$$

A simple way of writing this result is to put $y - y_0$, the distance of fall, equal to s and to multiply through by $2g$. Then

$$v^2 - v_0^2 = 2gs \tag{5.16}$$

This formula holds for *any* motion with constant acceleration, not only for free fall, provided that we replace g by the proper value a.

Equations (5.14), (5.15), and (5.16) contain the entire theory of uniformly accelerated motion, of which free fall is a special case. The student would do well to commit them to memory. Equation (5.14) means that the velocity increases by constant amounts in equal intervals; Eq. (5.15) expresses the fact that the area within a triangle equals half the product of its base by its height (see Fig. 5.4); Eq. (5.16) has no such simple interpretation.

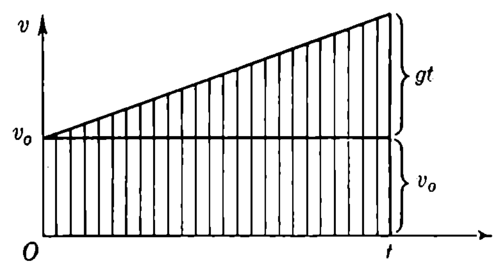


FIG. 5.4. Shaded area (distance y-y_0) = rectangle (v_0t) + triangle ($\tfrac{1}{2}t \cdot gt$).

***5.7. Worked Example.** A stone is thrown upward with an initial velocity of 64 ft/sec.

a. What is its velocity after 3 sec?

b. How high will it be after 3 sec?

c. How high will it rise?

d. How long will it stay in the air?

e. What velocity will it acquire after 10 ft of rise?

To answer any of these questions one must first specify what direction is to be taken as the positive one. Let us reckon y positive upward, as is customary. The acceleration g will then be *negative*, that is, $g = -32$ ft/sec². (This approximate value will here be used for convenience.)

Question a may be answered with the use of Eq. (5.14).

$$v = 64 \text{ ft/sec} - 32 \text{ ft/sec}^2 \times 3 \text{ sec} = -32 \text{ ft/sec}$$

Hence the velocity is downward and has a magnitude of 32 ft/sec.

The answer to b is based on Eq. (5.15). Here $y_0 = 0$, and

$$y = 64 \text{ ft/sec} \times 3 \text{ sec} - 16 \text{ ft/sec}^2 \times 9 \text{ sec}^2 = 48 \text{ ft}$$

When the stone is at its highest point, its velocity is zero. The corresponding time is observed from Eq. (5.14), which says

$$0 = 64 \text{ ft/sec} - 32 \text{ ft/sec}^2 \times t$$

and hence

$$t = 2 \text{ sec}$$

To find the position of the stone at this time we substitute 2 sec for t in Eq. (5.15), which then gives $y = 128 \text{ ft} - 64 \text{ ft} = 64 \text{ ft}$ as the answer to c.

Question d can be answered by noting that $y = 0$ when the stone returns to the ground. Thus, using Eq. (5.15), we have

$$0 = v_0t + \tfrac{1}{2}gt^2 = 0$$

This has two solutions, $t = 0$ and $t = -2v_0/g$. The first is trivial, for it gives the initial time at which the stone was projected. The second is the answer we want;

$$t = -\frac{128 \text{ ft/sec}}{-32 \text{ ft/sec}^2} = 4 \text{ sec}$$

We have already seen that the time of ascent was 2 sec and have therefore also verified the obvious fact that the time of descent equals the time of ascent.

The solution of question e requires the use of Eq. (5.16):

$$v^2 - (64 \text{ ft/sec})^2 = -64 \text{ ft/sec}^2 \times 10 \text{ ft}$$

Hence $v^2 = 3{,}456 \text{ ft}^2/\text{sec}^2$, and $v = 58.8$ ft/sec.

***5.8. Motion in a Plane. Velocity and Acceleration as Vectors.** Thus far only rectilinear motion has been considered. The most general kind of motion is not confined to a straight line; it takes place in three dimensions and requires a knowledge of x, y, and z at every instant. An example is the flight of a long-range projectile, which describes a curve lying predominantly in a vertical plane but departs slightly from this plane because of the earth's rotation. This case, however, is too difficult to be treated in this book. Since the departure from the original plane of flight is small and since most other motions of interest are confined to two dimensions, we shall restrict our attention to *motion in one plane*.

To this end it is necessary that we recognize at once the vector nature of velocity and acceleration. In Fig. 5.5 the curve represents the path of a moving particle. The points marked 1 and 2 are its positions at times t_1 and t_2. At every instant, the displacement of the particle from the origin is represented by a vector drawn from the origin, such as $\mathbf{r}_1$ or $\mathbf{r}_2$. The two rectangular components of the vector $\mathbf{r}$ are x and y, and these are functions of the time.

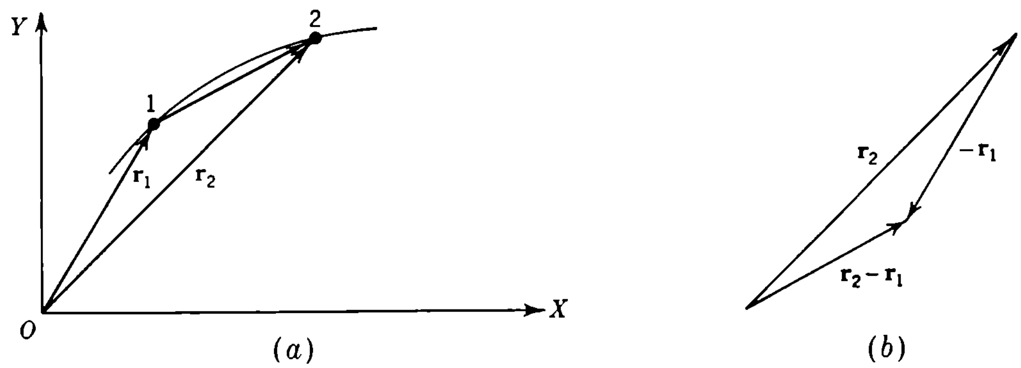


FIG. 5.5. Velocity as a vector.

Average velocity is defined exactly as in Eq. (5.2), except that the vector displacement $\mathbf{r}$ must be substituted for the one-dimensional distance y. Thus the average velocity of the moving particle between positions 1 and 2 is

$$\bar{\mathbf{v}}_{21} = \frac{\mathbf{r}_2 - \mathbf{r}_1}{t_2 - t_1} \tag{5.17}$$

The difference in the numerator is now a *vector difference;* hence $\bar{\mathbf{v}}_{21}$ is a vector. This vector difference, as shown in Fig. 5.5*b*, is the line drawn from the point 1 to the point 2. When its length is divided by $t_2 - t_1$, it represents the average velocity we are seeking. To find the average velocity on a curved path this graphical procedure must be used; there is in general no formula to take its place.

Next we define the *instantaneous* velocity, again in conformity with our previous development. Equation (5.4) is applicable, except for the substitution of $\mathbf{r}$ for y.

$$\mathbf{v} = \lim_{t_2 \to t_1} \frac{\mathbf{r}_2 - \mathbf{r}_1}{t_2 - t_1} \tag{5.18}$$

This tells us to form the vector difference $\mathbf{r}_2 - \mathbf{r}_1$ as the point 2 in Fig. 5.5*a* approaches more and more closely the point 1. But it is seen that, in the limit, the *direction* of this vector will be along the tangent to the curve at 1, and its length is infinitesimally small. If we divide its length by the other small quantity, $t_2 - t_1$, a finite quotient results whose magnitude equals the rate of progression of the particle, in ft/sec or cm/sec or a similar unit, along the curve. Hence the important theorem:

The instantaneous velocity of a particle moving on a curve is a vector, tangent to the curve at the point considered, and of magnitude equal to the rate of progression along the curve.

Rate of progression along the curve without regard to direction is called *speed;* it is the magnitude of the vector velocity.

Equation (5.18) may also be written in derivative form,

$$\mathbf{v} = d\mathbf{r}/dt \tag{5.18$'$}$$

but the limit process by which the *vector differential* $d\mathbf{r}$, explained in connection with Fig. 5.5, was derived must never be forgotten.

Acceleration, also, is a vector. Formally its definition is the same as before,

$$\mathbf{a} = d\mathbf{v}/dt \tag{5.19$'$}$$

but to see its meaning it must be written in more explicit form,

$$\mathbf{a} = \lim_{t_2 \to t_1} \frac{\mathbf{v}_2 - \mathbf{v}_1}{t_2 - t_1} \tag{5.19}$$

Figure 5.6*a* shows what is involved here. The velocities, constructed in accord with the theorem just proved, are drawn at the points 1 and 2. Equation (5.19)

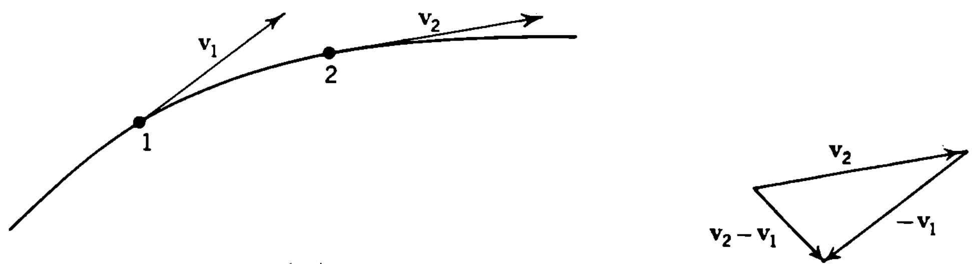


FIG. 5.6. Acceleration as a vector.

asks us to form their difference. This is done in Fig. 5.6*b*. The vector $\mathbf{v}_2 - \mathbf{v}_1$, when divided by $t_2 - t_1$, is the average acceleration during the interval from t_1 to t_2. The *instantaneous* acceleration at t_1 is obtained by letting the point 2 approach 1. While this is done, the difference $\mathbf{v}_2 - \mathbf{v}_1$ will change its direction (as it becomes smaller), and it is not possible to say in general what its limiting direction will be. Only for special types of motion, such as that to be discussed in the next section, can this result be specified.

If the motion is rectilinear, $\mathbf{v}_2$ and $\mathbf{v}_1$ are in the same straight line and the definition (5.19) reduces to the ordinary one, Eq. (5.6).

Finally we insert a brief comment about *relative* velocity. This is a term used to describe the motion of one object with respect to another. Assume that body A has, at a given instant, a velocity $\mathbf{v}_A$ and body B a velocity $\mathbf{v}_B$, both velocities being determined with respect to the same reference system, *e.g.*, the earth's surface. The relative velocity of B with respect to A is then $\mathbf{v}_B - \mathbf{v}_A$. For example, if an airplane has a velocity of 200 miles/hr north (relative to the ground) while a 50 mile/hr west wind is blowing, the velocity of the airplane relative to the air is seen to be $\sqrt{200^2 + 50^2}$ miles/hr in a direction making an angle $\theta = \tan^{-1} {}^{50}\!/_{200}$ with north.

Strictly speaking, *all velocities are relative.* The theory of relativity says, in fact, that it is impossible to find out whether or not a body is absolutely at rest.

5.9. Uniform Circular Motion; Acceleration. Perhaps the simplest kind of motion in a plane is the revolution of a particle at constant speed, like a stone in a slingshot or a dust particle on the rim of a flywheel. Its path is a circle, and it travels equal distances along the circle in equal times. This motion is called *uniform circular motion.*

Despite this misleading name, it is not *uniform* motion, for the particle —believe it or not!—has an acceleration. Its speed, to be sure, is constant; but since the velocity continually changes its direction, the *acceleration vector* is not zero. To see this, consider Fig. 5.7a.

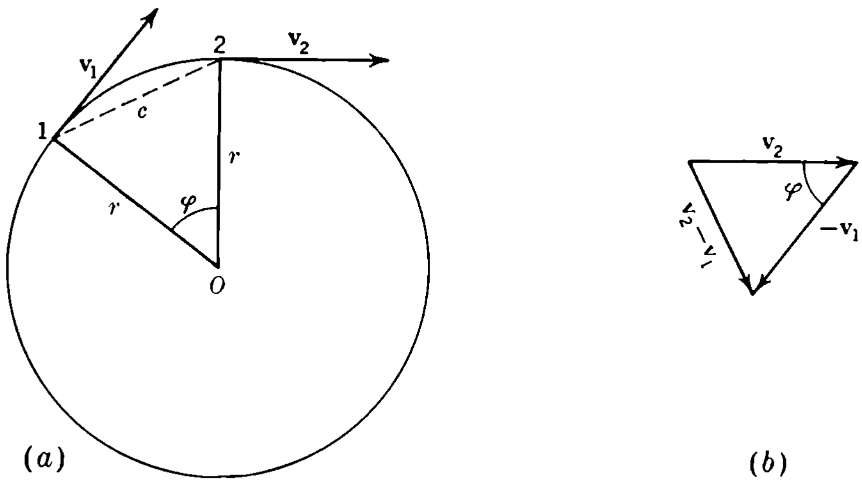


FIG. 5.7. Acceleration in uniform circular motion.

In position 1 the velocity of the particle is $\mathbf{v}_1$, tangential to the circle. At 2 it is $\mathbf{v}_2$, and the vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ have the same length, v. To find the acceleration [cf. Eq. (5.19)] we first form $\mathbf{v}_2 - \mathbf{v}_1$. This is done in Fig. 5.7b. Note that the angle φ in Fig. 5.7b is equal to φ in Fig. 5.7a, since the velocity vectors are perpendicular to the radii r. Considering similar triangles we can therefore say that

$$\text{Length of } \mathbf{v}_2 - \mathbf{v}_1/\text{length of } \mathbf{v}_1 = c/r$$

if c is the dotted chord in Fig. 5.7a. Now the length of $\mathbf{v}_1$ is v, and for the length of $\mathbf{v}_2 - \mathbf{v}_1$ we shall write Δv. Therefore

$$\Delta v = cv/r \tag{5.20}$$

In accordance with the definition of acceleration [see also Eq. (5.19)] the *magnitude* of $\mathbf{a}$ is

$$a = \lim_{t_2 \to t_1} \frac{\Delta v}{t_2 - t_1} \tag{5.21}$$

Let us see, then, what happens to Δv as $t_2 \to t_1$. Clearly the angle φ becomes smaller and smaller, and c in Fig. 5.7a, while becoming smaller, approaches coincidence with the arc s between the points 1 and 2. In other words, in the limit, $c \to s = v(t_2 - t_1)$. Hence, by Eq. (5.20).

$$\lim_{t_2 \to t_1} \Delta v = \frac{v^2(t_2 - t_1)}{r}$$

and in view of Eq. (5.21), which tells us that $a = \lim_{t_2 \to t_1} \frac{v^2(t_2 - t_1)}{r(t_2 - t_1)}$,

$$a = \frac{v^2}{r} \tag{5.22}$$

This gives the magnitude of **a**.

Its direction may be found as follows: When t_2 approaches t_1, the angle φ in Fig. 5.7b tends to zero. Thus the direction of $\mathbf{v}_2 - \mathbf{v}_1$, which is also the direction of **a**, tends toward perpendicularity to $\mathbf{v}_1$. Looking now at Fig. 5.7a we perceive that **a**, if it is perpendicular to $\mathbf{v}_1$, must point toward O. What has been proved is this:

A particle in uniform circular motion has an acceleration of magnitude v^2/r, which is always directed toward the center of the circle.

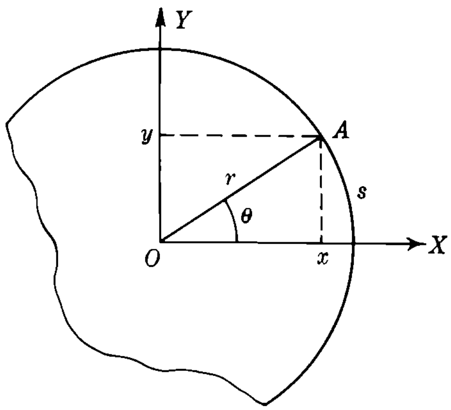


FIG. 5.8. x and y in uniform circular motion.

This acceleration, though of constant magnitude, changes its direction continually and is therefore not a constant vector. It is called *centripetal* acceleration (Latin *centrum*, center; *petere*, to seek).

To understand the physical meaning of this result one must note that the particle, if it had no acceleration, would move on a tangent in the direction of $\mathbf{v}_1$. To keep it on the periphery of the circle an acceleration toward the center must be imposed on it.

5.10. Uniform Circular Motion (Continued). Further insight into this important type of motion is gained if we consider several ways in which the position of the particle can be described. The simplest is to introduce an angle θ, as in Fig. 5.8, which measures the angular separation between the radius vector to the instantaneous position of the particle A and the horizontal or X axis. Assume θ to be expressed in radians. The rate of change of θ,

$$\omega = d\theta/dt, \tag{5.23}$$

is called the *angular velocity* of A. Though it may be shown to be a vector, its vector nature need not concern us here. If the motion is to be uniform circular, ω must be a constant.

The time interval during which A describes one revolution is called the *period;* it is

$$\boxed{P = 2\pi/\omega} \tag{5.24}$$

since a full revolution corresponds to an angle 2π. The reciprocal of the period is the number of revolutions, or cycles, per second; it is called the *frequency* of revolution. It is denoted by f.

$$\boxed{f = 1/P = \omega/2\pi} \tag{5.25}$$

In words, *the angular velocity is* 2π *times the frequency.*

Furthermore, if the radius of the circle is r and the arc from the X axis to A is s, we find from Fig. 5.8 that

$$\boxed{s = r\theta} \tag{5.26}$$

Since the speed of A along the circle is $ds/dt = v$, Eq. (5.23) when combined with Eq. (5.26) states that

$$\boxed{v = r\,d\theta/dt = r\omega} \tag{5.27}$$

Equations (5.23) to (5.27), although very simple and easily derived, are used so frequently in physics that the student might well ponder their meaning until thorough familiarity with them has been attained.

Instead of using the angle θ, one may also describe the motion by means of rectangular coordinates x and y, as drawn in Fig. 5.8. There we see that

$$\left.\begin{aligned} x &= r\cos\theta \\ \text{and} \qquad y &= r\sin\theta \end{aligned}\right\} \tag{5.28}$$

If we remember the calculus formulas

$$\frac{d(\cos\theta)}{d\theta} = -\sin\theta \qquad \frac{d(\sin\theta)}{d\theta} = \cos\theta$$

and also

$$\frac{d(\cos\theta)}{dt} = \frac{d(\cos\theta)}{d\theta}\frac{d\theta}{dt}$$

we find

$$\left.\begin{aligned} \frac{dx}{dt} &= -r\sin\theta\frac{d\theta}{dt} = -r\omega\sin\theta \\ \frac{dy}{dt} &= r\omega\cos\theta \end{aligned}\right\} \tag{5.29}$$

These are, respectively, the x and y components of the velocity **v**, namely, v_x and v_y. From these two equations we may derive

$$v = \sqrt{v_x^2 + v_y^2} = r\omega$$

which is the same as Eq. (5.27).

By differentiating Eqs. (5.29) once more we obtain the components of the acceleration,

$$\left.\begin{aligned} a_x &= \frac{d^2x}{dt^2} = -r\omega \cos\theta \frac{d\theta}{dt} = -r\omega^2 \cos\theta = -\frac{v^2}{r}\cos\theta \\ a_y &= \frac{d^2y}{dt^2} = -r\omega^2 \sin\theta = -\frac{v^2}{r}\sin\theta \end{aligned}\right\} \qquad (5.30)$$

Notice how precisely this agrees with the conclusions of the former section. Indeed,

$$a = \sqrt{a_x^2 + a_y^2} = \frac{v^2}{r}$$

In view of the relations (5.24), (5.25), and (5.27) the centripetal acceleration can also be expressed in the following equivalent forms:

$$a = v^2/r = \omega^2 r = 4\pi^2 f^2 r = 4\pi^2 r/P^2 = v\omega \qquad (5.31)$$

PROBLEMS

1. An end of a piston rod requires ½ sec to move from the center to the end of its path, a distance of 2 ft. What is its average velocity (*a*) during the first ½ sec; (*b*) during the first second; (*c*) during the first 1½ sec; (*d*) during the first 2 sec?

2. By observation a certain motion is found to be represented by the formula

$$x = 4\text{ ft} + 2\text{ ft/sec} \times t + 3\text{ ft/sec}^2 \times t^2 - 1\text{ ft/sec}^3 \times t^3$$

if t is expressed in seconds. Find velocity and acceleration at any instant, by differentiation. Make three graphs on squared paper, plotting x, v, and a against t from 0 to 4 sec.

3. By counting the number of squares under the a curve of Prob. 2, show that

$$\int_0^{4\text{ sec}} a\,dt = v - v_0$$

at 4 sec. Note that, when a goes negative, the area between the curve and the t axis must also be taken as negative.

4. By counting the number of squares under the v curve of Prob. 2, show that

$$\int_0^{4\text{ sec}} v\,dt = x - x_0$$

5. A body moves along the X axis with an acceleration,

$$a = +6\text{ cm/sec}^3 \times t$$

t being in seconds. At twelve (12:00) o'clock it is 50 cm to the right of the origin and has a velocity of 10 cm/sec directed to the *left*. Find velocity and position of the body at 12:00:02 and at 12:00:04.

6. Compute the average velocity of the body in Prob. 5 for the following three intervals: (*a*) from 12:00:00 to 12:00:02; (*b*) from 12:00:02 to 12:00:04; (*c*) from 12:00:00 to 12:00:04.

7. A certain motion is represented by $x = A \sin \omega t$, and $A = 10$ cm,

$$\omega = 2 \text{ radians/sec}$$

(*a*) Plot x against t. (*b*) Compute v, and plot v against t. (*c*) Compute a, and plot a against t.

8. Repeat the work in Prob. 7, but let $x = A \sin (\omega t - \alpha)$. Use the same values for A and ω, and take $\alpha = \pi/4$.

9. Make a graph of (*a*) a against t; (*b*) v against t; (*c*) x against t for uniformly accelerated motion.

10. A projectile is fired vertically upward with an initial velocity of 1,000 ft/sec. (*a*) Where will it be after 3 sec? (*b*) How long will it rise? (*c*) How high will it rise? (*d*) With what velocity will it strike the ground? (Neglect air resistance.)

11. A rock is thrown vertically downward from a cliff with an initial velocity of 100 ft/sec. When and with what velocity will it reach the ground 500 ft below? (Neglect air resistance.)

12. To find the depth of a well a stone is dropped into it. It is heard to strike the bottom 3 sec after release. (*a*) Compute the depth of the well, assuming sound to travel infinitely fast. (*b*) Compute the depth, assuming the velocity of sound to be 1,000 ft/sec. (This leads to a quadratic equation.)

13. A freight train, starting from rest, acquires a velocity of 10 miles/hr while moving 2 miles under constant acceleration. (*a*) Find its acceleration. (*b*) Express it in ft/sec^2.

14. What velocity does a body acquire when falling from rest through a distance of 100 m? (Neglect air resistance.)

15. By what amount will the velocity of a body, initially moving at a rate of 5 m/sec, increase while it is falling through 100 m?

16. The velocity of a body increases with distance according to the law $v = bx$, where b is a constant. Calculate the acceleration at any point x. (HINT: Note that $dv/dt = dv/dx \cdot dx/dt = v\, dv/dx$.)

17. Galileo's law of free fall may be written $v = \text{const} \times t$. In an earlier, erroneous formulation, Galileo proposed the law $v = \text{const} \times x$, x being the distance of fall. Show that this does not correspond to a constant acceleration; show that it leads to $a = Ae^{bt}$, where A and b are constants.

18. A particle describes uniform circular motion of period 2 sec on a vertical circle of radius 5 ft. At $t = 0$ it is at the highest point on the circle and moving to the right. What is its average velocity (*a*) during the first half second; (*b*) during the first second; (*c*) during one revolution?

19. The velocity of an airplane relative to the air is 250 miles/hr north. The wind has a velocity of 20 miles/hr from northwest. Find the velocity of the airplane relative to the ground.

★20. A boat is crossing a river of width 1 mile, heading upstream so that it moves toward a point exactly opposite its starting point on the other shore. The current has a velocity of 5 miles/hr, and the trip across takes 15 min. What direction does the keel of the boat take relative to the riverbank? What is the velocity of the boat relative to the water? What distance does it actually travel?

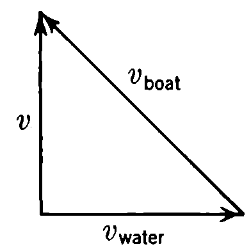


FIG. 5.9. Problem 20.

***21.** Suppose raindrops fall vertically to the ground? Why must a person running through the rain tilt his umbrella forward? (Determine the *relative* velocity of a raindrop with respect to the runner.)

22. Calculate the centripetal acceleration of the particle in Prob. 18.

23. A body describing uniform circular motion has a period of 5 sec and a speed of 10 ft/sec. Calculate its frequency, its angular velocity, and its centripetal acceleration.

24. Find the acceleration of a particle that moves on the periphery of a circle of 10 cm radius. At a given instant its speed is 25 cm/sec, and the speed is increasing at the rate of 20 cm/sec^2 along the path. (NOTE: The total acceleration equals the vector sum of centripetal acceleration and acceleration along the path.)

25. Christian Huygens (1629–1695) stated the following in the form of an anagram: If a body moves on the circumference of a circle with a speed equal to that which it would acquire by falling through half the radius of the circle, then its centripetal acceleration equals the acceleration in free fall. Prove this.

CHAPTER 6

FORCE AND ACCELERATION; DYNAMICS[1]

6.1. Introduction. Newton's First and Second Laws. Newton's first law states that, without application of an external force, a body moves with constant velocity. Velocity is here to be regarded as a vector; hence motion, in this simple case, takes place along a straight line. Perfectly force-free motion is rare indeed; the actual proper motion of the "fixed" stars, which are far away from other attracting matter, is perhaps the closest approximation to it. All ordinary motions are subject to some amount of frictional force and therefore illustrate Newton's first law only approximately.

According to Newton's second law the resultant force on a body (to be exact, on a particle) causes it to be accelerated, and the acceleration is proportional to the force. Both force and acceleration are vectors, and the word "proportional" here implies *equality of direction* and *proportionality of magnitudes;* for vectors are said to be proportional only if they have the same direction. By force is meant the sum, or resultant, of all forces acting *on* the body. The first law is seen to be a special case of the second, and we shall not consider it further.

In Chap. 3 on Statics a study was made of the conditions under which a body can be without acceleration. In Chap. 5 acceleration and velocity were analyzed from the point of view of kinematics, *i.e.*, without reference to forces. We have learned how velocities and positions of a particle can be derived when the acceleration is given. It is through this procedure that contact is made with Newton's second law. For the law tells us the acceleration when the forces are given, and on combining this knowledge with what was learned in the preceding chapter we are able to obtain all aspects of the motion from the given forces. The purpose of the present chapter is to study methods for doing this.

Newton's second law, which is perhaps the greatest generalization and the most useful theorem of all natural science, may be stated in the form

$$\mathbf{F} \propto m\mathbf{a}$$

where $\mathbf{F}$ is the resultant force, m the mass of the particle, and $\mathbf{a}$ its acceleration. But to say that one quantity is proportional to another is to say that it is equal to the other times a constant. Hence

$$\mathbf{F} = km\mathbf{a} \qquad (6.1)$$

[1] The student should review Sec. 2.2.

k being a constant, at present undetermined. Its value will depend upon the choice of units of **F** and m. The meaning of Eq. (6.1) will be apparent from the following experiment (Fig. 6.1):

A spring S rests on a smooth horizontal table. Its left end is fixed rigidly to an immovable wall, while its right end is attached to a mass m, able to slide without friction on the table.

We first consider what happens when *different forces* are allowed to act upon the same mass m. Let the mass be drawn to the right a distance b beyond x_0, which is the point at which the spring exerts no force. In accordance with Hooke's law the spring then pulls with a force proportional to b, say 2 lb. When the mass is now released, it will move back toward x_0 with an initial acceleration, which turns out to be, say, 5 ft/sec^2. Next the mass is drawn out a distance $2b$ beyond x_0. When it is released from this position, where the force is twice as great, the initial acceleration is found to be 10 ft/sec^2, in accordance with Eq. (6.1). Thus in general an initial displacement $n \times b$, which corresponds to a force of $n \times 2$ lb, produces an initial acceleration of $n \times 5$ ft/sec^2. Note that both **F** and **a** are to the left.

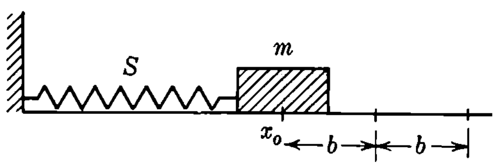


FIG. 6.1. Illustration of Newton's second law.

Second, we perform a set of experiments in which the force is held constant but is allowed to act on *different masses*. This may be done by drawing the spring out to the same distance b and letting the force of 2 lb, which it then exerts, successively accelerate different masses. When m is doubled, **a** will be 2.5 ft/sec^2; when m is halved, **a** will be 10 ft/sec^2, and so forth. This, too, is implied by Eq. (6.1).

In these experiments none of the masses moved with constant acceleration. For as the force diminished with diminishing extension of the spring, the acceleration became smaller. In particular, at the instant in which the masses passed through the point x_0, there was no force. Hence we conclude that at this instant the acceleration of all the masses was momentarily zero, their velocity was uniform.

6.2. The Meaning of Mass. The time has come when we must state more clearly what is meant by *mass*. Although both mass and weight are commonly measured in pounds, the two are not the same. The mass of an object is the same everywhere in the universe; its weight is different, for example, on the moon and on the earth.

Mass is identical with inertia; but this statement hardly defines it, since the term "inertia," although it has the right qualitative implications, is equally vague. Mass has to do with how difficult it is to *accelerate* a body. In the following we give what is called an *operational* definition of mass, a definition not merely by words but by reference to experimental operations, a definition that has the further advantage of permitting a quantitative determination of mass. It will be based on Newton's second law.

Suppose we are given three bodies, numbered 1, 2, and 3, and we are to find their masses. Attach each of them in turn to the spring of Fig. 6.1, draw out the spring to the same distance b beyond its neutral point, and measure the initial acceleration. This, while difficult practically, can nevertheless be carried out in principle. Call the observed accelerations of the three bodies a_1, a_2, a_3 (only magnitudes need here be considered). By Eq. (6.1), $F = km_1a_1 = km_2a_2 = km_3a_3$. Hence

$$\frac{m_2}{m_1} = \frac{a_1}{a_2}; \qquad \frac{m_3}{m_1} = \frac{a_1}{a_3} \tag{6.2}$$

The masses are inversely as the accelerations. All mass ratios may thus be determined. If we now choose m_1 to be a unit of mass, the masses themselves are defined by relations (6.2). By means of experiments of this type all masses can be determined through comparison with a unit mass or standard mass.

Standard masses are actually used in physics, although the comparison is usually made by more practical procedures. The standard *kilogram* is a platinum object kept at Sèvres in France; the standard pound is in Westminster, England. They are carefully protected from all corrosive influences.

6.3. Units of force. We have already become acquainted with one unit of force, the pound. It was seen to be the force with which gravity pulls upon 1 lb of mass. Hence 1 lb of force produces in 1 lb of mass an acceleration of $g = 32.2$ ft/sec².

Let us substitute this information in Eq. (6.1). It reads

$$1 \text{ lb-force} = k \times 1 \text{ lb-mass} \times 32.2 \text{ ft/sec}^2.$$

To balance the equation numerically we must give the constant k the numerical value $1/g = 1/32.2$. This is often inconvenient.

To avoid it we may introduce a new unit of force, so chosen that the constant k will have the value unity (and no physical dimension). This unit force must clearly be one that, when acting on a mass of 1 lb, will give it an acceleration of 1 ft/sec^2, as substitution in Eq. (6.1) immediately shows. It is called the *poundal*. Since 1 lb-force causes an acceleration of 32.2 ft/sec^2 in a mass of 1 lb, 1 lb-force must be equal to 32.2 poundals. The poundal is said to be the *absolute* unit of force, the pound the *gravitational* unit. When the force is expressed in absolute units, Newton's second law reads

$$\mathbf{F} = m\mathbf{a} \qquad \text{(absolute form)} \qquad (6.3)$$

When **F** is expressed in gravitational units, the factor $1/g$ must be inserted; that is to say, m must be replaced by W/g, where W is the weight in pound-force, numerically equal to the mass in pounds. Hence the formula reads

$$\mathbf{F} = \frac{W}{g}\mathbf{a} \qquad \text{(gravitational form)} \qquad (6.4)$$

In this equation, both F and W are measured in pound-force, while both g and a have the units of acceleration. This form of Newton's second law is commonly employed in engineering practice.

The same distinction between absolute and gravitational units occurs when masses and accelerations are expressed in the cgs system, where the unit of mass is the gram and the unit of length the centimeter. Here the absolute unit of force is the *dyne* (Greek, *dynamos*, force), defined as the force which, when acting on a mass of 1 gm, imparts to it an acceleration of 1 cm/sec². The gram-force, on the other hand, which is the gravitational unit of force in the cgs system, is 980 dynes (g = 980 cm/sec² in this system), because it produces an acceleration of 980 cm/sec² in a mass of 1 gm.

Again, in the mks system, a similar distinction is made. Here the absolute unit of force is the *newton*, defined as the force that, when acting on a mass of 1 kg, imparts to it an acceleration of 1 m/sec². The kilogram-force, which is the gravitational unit of force in the mks system, is 9.8 newtons, since g = 9.8 m/sec².

In all three systems Eq. (6.3) is valid when forces are expressed in absolute units, Eq. (6.4) when they are expressed in gravitational units.

We have thus arrived at six different units of force; they are given in Table 6.1. They are called absolute fps unit (poundal), absolute cgs unit (dyne), absolute mks unit (newton), and so forth. All three absolute units have the following advantages:

Table 6.1. *Units of Force*

	fps	cgs	mks
Absolute	poundal	dyne	newton
Gravitational	lb	gm	kilogram

1. They permit the use of Newton's second law in the simple form (6.3).

2. Their names, being different from the names of the corresponding mass units, avoid confusion between force and mass, which are fundamentally distinct.

Henceforth the student should learn to distinguish very clearly between a pound of force and a pound of mass, as well as between a gram of force and a gram of mass. Though the units are homonyms, their meanings are no more identical than the North Pole and the telephone pole. Whether the word pound, or gram, or ton (= 2,000 lb) refers to a mass or to a force will usually be clear from the context; when confusion is likely to arise, these words will be qualified by the addition of "force" or "mass."

Also it should be understood that the distinction between gravitational and absolute applies only to forces, not to masses. In this book we shall employ no gravitational unit of mass. For problem work involving the use of Newton's second law we recommend for the present the following procedure: First express all forces in absolute units, remembering that 1 gm-force equals 980 dynes and 1 lb-force is 32.2 poundals. Then use Newton's second law in the form (6.3). All answers will then be in absolute units. If the answers are wanted in gravitational units, we divide the results for forces by g.

★6.4. Worked Examples. *a.* A locomotive is capable of exerting a maximum pull of 2 tons. What velocity can it give to a freight train weighing 300 tons in a distance of 1 mile, if it starts from rest on a level track and the train is subject to a frictional force of 10 lb per ton of weight?

The force on the train is a forward pull of 4,000 lb less the backward pull due to friction of 300 × 10 lb = 3,000 lb; the resultant is 1,000 lb, or about 32,000 poundals. The mass of the train is 300 × 2,000 lb = 600,000 lb. Applying Newton's law in the form (6.3) we find

$$32{,}000 \text{ poundals} = 600{,}000 \text{ lb} \times a$$
$$a = \tfrac{4}{75} \text{ poundals/lb} = \tfrac{4}{75} \text{ ft/sec}^2$$

for a poundal, according to its definition, is 1 lb-mass × 1 ft/sec². To find the velocity acquired by the train over 1 mile of distance, we use Eq. (5.16), familiar from Sec. (5.5),

$$v^2 - v_0^2 = 2as$$

In this instance $v_0 = 0$. Hence

$$v^2 = \tfrac{8}{75} \text{ ft/sec}^2 \times 5{,}280 \text{ ft} = 563 \text{ (ft/sec)}^2$$
$$v = 23.7 \text{ ft/sec}$$

b. An airplane of total weight 1 ton is subject to the following forces: the thrust of its propeller, $\mathbf{F}_1 = 2{,}000$ lb (cf. Fig. 6.2); its weight $\mathbf{W}$; the action of the air upon it, $\mathbf{F}_2 = 1{,}000$ lb; find its momentary acceleration.

The first step is to calculate the resultant force on the airplane. This can be done either graphically or by components. W is, of course, 2,000 lb. Using the method of *components* we obtain for the horizontal component of the resultant (in the direction of the propeller force)

$$R_x = F_1 - F_2 \cos 30° = 1{,}134 \text{ lb}$$

and for the downward component

$$R_y = W - F_2 \sin 30° = 1{,}500 \text{ lb}$$

The resultant is a force of $\sqrt{(1{,}134)^2 + (1{,}500)^2}$ lb = 1,880 lb in a direction making an angle θ with F_1, and $\tan \theta = 1{,}500 \text{ lb}/1{,}134 \text{ lb} = 1.32$. Thus from trigonometry tables θ is approximately 53°. The acceleration is in this direction. Its magnitude is computed from Eq. (6.3) after conversion to absolute units. Since 1,880 lb = 60,540 poundals, we have

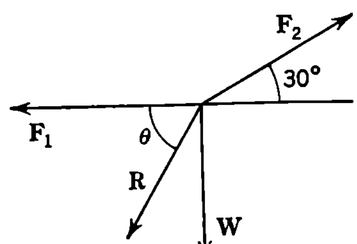


FIG. 6.2.

$$60{,}540 \text{ lb} \times \text{ft/sec}^2 = 2{,}000 \text{ lb} \times a$$
$$a = 30.3 \text{ ft/sec}^2$$

Notice that the mass of the airplane, not explicitly given, is numerically equal to its weight.

c. The moon revolves about the earth at a distance of 3.84×10^5 km and with a period of 28 days. Its mass is 7.36×10^{22} kg. Find the force with which the earth attracts it.

In accordance with Sec. 5.10 the moon has a centripetal acceleration $a = 4\pi^2 r/P^2 = 0.260 \text{ cm/sec}^2$. To provide this the earth must attract it with a force given by

$$ma = 1.91 \times 10^{25} \text{ dynes}$$

6.5. The Force of Gravity. Gravity, which means the earth's attraction, gives all falling bodies the same acceleration. This is a fact of observation. Does it mean that the force of gravity is the same for all bodies? Newton's law says no; for if $F = ma$, the force of gravity must be proportional to the mass of the falling object. This simple conclusion has given rise to much philosophic speculation and to considerable confusion. One wonders indeed how the force of gravity manages to adjust itself so precisely to the mass of an object, since most forces are quite independent of the masses on which they act. The elastic force of the spring considered in Sec. 6.1 depends wholly on the condition of the spring, not on the mass to which it is attached. As already stated, the force of gravity acting on an object is its weight. What has been said is that weight is proportional to mass.

Physicists have at times doubted that this proportionality can be exact. But in 1830 the famous mathematician *Friedrich Bessel* proved that it is at least as exact as measurements could determine. There was no satisfactory theoretical explanation of this remarkable coincidence until the advent of the general theory of relativity during the present century.

Table 6.2. *Values of g at Various Points on the Earth's Surface*

Location	Latitude	Elevation, m	g, cm/sec²
North Pole	90°00′	0	983.1
Karajak Glacier, Greenland	70°27′	20	982.5
Chicago, Ill	41°47′	182	980.3
New York, N.Y	40°49′	38	980.3
Mt. Hamilton, Calif	37°20′	1,282	979.7
Panama Canal Zone	8°55′	6	978.2
Equator	0°00′	0	978.1

The weight of a body of mass m is clearly

$$\mathbf{W} = m\mathbf{g} \tag{6.5}$$

when expressed in absolute units. If m is in grams and g is taken to be 980 cm/sec², the numerical result for W will be in dynes. If m is in pounds and g = 32.2 ft/sec², W will be in poundals. We have already used this fact in Example b of the preceding section. The direction of **W** is that of **g**, that is, downward.

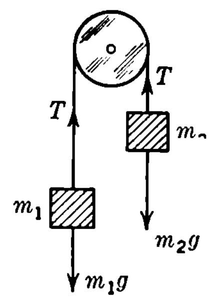


FIG. 6.3. Atwood machine.

A simple device known as the Atwood machine (George Atwood, 1746–1807) represents an application of these principles. It consists of a pulley, here regarded as frictionless and massless, with a flexible cord running over it (see Fig. 6.3). Attached to the ends of the cord are the masses m_1 and m_2. The larger one of these will be accelerated downward, the other upward. By determining the accelerations and the two masses it is possible to find g with considerable accuracy. The theory of the machine is as follows:

Newton's second law is applied to each of the two masses separately. The first is subject to two forces, $m_1\mathbf{g}$ and **T, T** being the tension in the string supporting m_1. Their downward resultant is $m_1\mathbf{g} - \mathbf{T}$. In a flexible string this tension is communicated, with its value unaltered, to the other side of the pulley, where it acts on m_2. Calling the downward accelerations of the two masses a_1 and a_2 and applying Eq. (6.3) to each mass we find

$$\left.\begin{aligned} m_1g - T &= m_1a_1 \\ m_2g - T &= m_2a_2 \end{aligned}\right\} \tag{6.6}$$

But if the spring is inextensible, $a_2 = -a_1$. On making this substitution and then subtracting the second equation from the first we find

$$(m_1 - m_2)g = (m_1 + m_2)a_1$$

and hence

$$a_1 = \frac{m_1 - m_2}{m_1 + m_2} g$$

Notice that, if the masses are nearly equal, a_1 is quite small and therefore easily measurable.

6.6. Friction Again. The qualitative aspects of the force of friction were discussed in Sec. 3.12 (review this). We are now ready to consider how this force can be measured. Certain regularities, sometimes called *laws of friction*, are matters of common observation and will now be discussed. The term "laws" should be employed with caution, for the facts to be stated are true only within limits.

The first is illustrated by a brick that is being pushed over a rough surface. It is found that the force of friction is the same whether the base on which it rides is the narrow or the wide face of the brick. Thus, in general, *the force of friction is independent of the area of contact.*

Second, if a board, with its flat surface horizontal, is loaded by weights and then drawn over a rough horizontal surface, it is found that the retarding force of friction is proportional to the total weight. Hence *the force of friction is proportional to the total force normal (perpendicular) to the plane.* Now the force of friction is always parallel to the plane and will be symbolized by $F_{\|}$; the normal force, which may or may not be the weight of the object to be moved—there might be an additional push against the plane—will be called $F_{\perp}$. The relation stated may then be expressed

$$F_{\|} = \mu F_{\perp} \tag{6.7}$$

The constant μ, which depends on the nature of the *two* substances in contact, is called the "coefficient of friction." Its value has been determined for a great variety of pairs of substances, some of which are given in Table 6.3. In Sec. 3.12 attention was called to the circumstance that kinetic friction is a little smaller than static friction. To be accurate, one must therefore distinguish between μ_{kin} and μ_{stat}; the latter is slightly

Table 6.3. *Coefficients of Friction*

Metal on dry oak wood	0.50
Metal on wet oak wood	0.25
Metal on metal, without lubricant	0.18
Metal on metal, lubricated	0.03
Steel on agate	0.20
Leather on wood	0.35
Leather on metal	0.56

larger than the former, but for the purposes of this book the difference will be disregarded.

We now apply these considerations to a problem of frequent occurrence, the motion of a body sliding down an incline (cf. Fig. 6.4). If its mass is m and the angle of the incline θ, the force normal to the plane is $mg \cos \theta$, while the component of the force along the plane is $mg \sin \theta$. The component $mg \cos \theta$ upon the body is neutralized by an equal and opposite force exerted by the plane (not drawn in the figure). The frictional force, counteracting the force $mg \sin \theta$ that propels the body down the plane, is $\mu mg \cos \theta$ in accordance with Eq. (6.7). Newton's second law, when applied in the direction of motion, is then

$$mg \sin \theta - \mu mg \cos \theta = ma$$

Solving for a we find

$$a = g(\sin \theta - \mu \cos \theta) \quad (6.8)$$

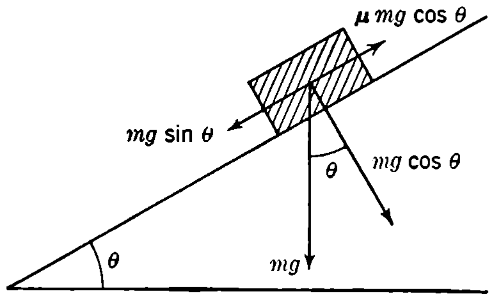


FIG. 6.4. Motion on an incline.

If $\theta = 90°$, $a = g$, which is what we should expect. But if $\theta = 0$, Eq. (6.8) would seem to predict an acceleration up the plane. This of course is nonsense; it must be remembered that friction always *opposes* motion and cannot produce motion of itself. Equation (6.8) is therefore correct only as long as it predicts motion in the direction of the accelerating force.

Under what condition will a be zero? According to Eq. (6.8) this happens when $\sin \theta - \mu \cos \theta = 0$. Solving this we find that

$$\tan \theta = \mu \quad (6.9)$$

The angle θ defined by this condition is called the *limiting angle of repose.* This last equation provides a useful method for measuring μ: one alters the angle of the incline from the value 0 to that at which sliding just begins. The tangent of this angle is the coefficient of (static) friction.

Finally Eq. (6.8) also tells us what happens in the absence of friction. When $\mu = 0$, $a = g \sin \theta$. This is the relation that enabled Galileo to determine g. The speed of falling bodies is so great that g is difficult to measure directly. But a can be made as small as is desired by making θ small and can be measured with ease. Motion on an incline is still uniformly accelerated, but the value of the acceleration is lessened.

6.7. Dimensional Analysis. It has frequently been emphasized that, in all the equations of physics, both sides must be measured in the same units. There is, however, an even more basic kind of homogeneity in these equations which arises from the fact that all quantities which are equated must have the same physical nature. The student who is quick in discerning this can often avoid mistakes. There is a systematic procedure for checking whether or not equations are correct from this point of view; it is called "dimensional analysis."

The three basic quantities in the science of mechanics are length, time, and mass. They will be designated by L, T, and M for the present purposes. Out of these three all other quantities can somehow be constructed. For example, velocity is a length divided by a time. It is said to have the *physical dimension* $[LT^{-1}]$. Thus the physical dimension of the quantity 30 miles/hr is $[LT^{-1}]$. The brackets are used to indicate that LT^{-1} represents *only* the physical dimension and that all numerics have been disregarded.

The dimension of an area is $[L^2]$, of an acceleration $[LT^{-2}]$, of a force in the absolute system $[MLT^{-2}]$, and so forth. When quantities are raised to some power, the physical dimensions are raised to the same power. The point of it all is that every equation used in physics must have the same dimensions on both sides. Let us test, for example, Eq. (5.16) for uniformly accelerated motion, $v^2 = 2as$. We have

$$[L^2T^{-2}] = [LT^{-2}][L]$$

and this is obviously correct.

It is easily seen that Newton's second law in its absolute form [Eq. (6.3)] is also correct, for in a sense it defines the physical dimensions of F. But in its gravitational form (6.4) it presents some puzzling features. Suppose we write

$$F = \frac{W}{g}a$$

and treat W as weight (a force!) and g as a number. We then have

$$[MLT^{-2}] = [MLT^{-2}][LT^{-2}]$$

which is incorrect. A correct dimensional equation results if the constant g is given its dimensions of acceleration. Henceforth, in all examples of dimensional analysis, Newton's law will be assumed to be written in its absolute form, so that ambiguity cannot arise.

Considerations of this sort become more useful as more and more new physical quantities, like energy, momentum, power, are introduced. Each of these has its distinctive dimensions.

6.8. Summary on Units of Force. In the present chapter we have introduced six units of force, and we have reviewed three units of mass.

1. In the cgs system the unit of mass is always the gram. The unit of force can be either the absolute one (dyne) or the gravitational one (gram-force). When the absolute unit is employed, Newton's second law takes the standard form

$$\mathbf{F} = m\mathbf{a}$$

A dyne is the force that will give a mass of 1 gm an acceleration of 1 cm/sec^2. A gram-force equals 980 dynes.

2. In the mks system the kilogram is the unit of mass. The unit of force can be either the absolute one (newton) or the gravitational one (kilogram-force). A newton is that force which will give a kilogram-mass an acceleration of 1 m/sec^2. A kilogram-force equals 9.8 newtons.

3. In the British system the unit of mass is always the pound (or one of its multiples or submultiples). The unit of force can be either the absolute one (poundal) or the gravitational one (pound-force). When the absolute unit is employed, Newton's second law takes the simple standard form above.

A poundal is the force that will give a mass of 1 lb an acceleration of 1 ft/sec^2. A pound-force equals 32.2 poundals.

Engineers often prefer to use gravitational units of force throughout their calculations. As we have seen, this necessitates the introduction of a factor $1/g$ in Newton's second law if the conventional units of mass are employed. One can, however, retain gravitational units of force *and* the simple form of the law provided that he introduces a new unit of *mass*. If 1 lb-force gives a mass of 1 lb an acceleration of 32.2 ft/sec^2, then 1 lb-force will give a mass of 32.2 lb an acceleration of 1 ft/sec^2. Therefore, when masses are expressed in multiples of 32.2 lb and F in pounds, the simple form [Eq. (6.3)] remains valid.

A mass of 32.2 lb is called 1 *slug*. A slug is that mass which acquires an acceleration of 1 ft/sec^2 under the application of 1 lb-force.

6.9. Limitations of Newton's Second Law. The preceding discussions have always taken for granted that there can be no doubt as to the meaning of a, the acceleration of a body. A little reflection shows, however, that a is always measured *relative to something else*. When a body falls, it has an acceleration g with respect to the earth. Since the earth rotates about its axis, a point manifestly at rest on its surface has itself an acceleration with respect to a fixed star and therefore the acceleration of a falling body with respect to that fixed star would be different from g.

To take another example, consider that an airplane is falling toward the earth. The passengers in the cabin have an acceleration zero with respect to the airplane but of g relative to the earth's surface.

Universal experience with moving bodies decides the following: *The acceleration that enters Newton's second law must be relative to some body* *that does not have an acceleration itself.* To be sure, this raises a host of other questions; for how can we be sure that a body which we innocently regard as having no acceleration may not be accelerated relative to objects so far away as to be invisible? Questions of this kind are dealt with in the theory of relativity. In this book we assume that we can always tell whether a body is accelerated or not.

Clearly, then, the earth is not a proper "system of reference" for measuring a, since every point on its surface except the poles has a centripetal acceleration $\omega^2 R$. With

$$\omega = 2\pi \text{ per day} = (2\pi/86{,}400)\ \text{sec}^{-1} \qquad \text{and} \qquad R = 6.4 \times 10^8\ \text{cm}$$

we obtain about 3.5 cm/sec² at the equator, where the acceleration is greatest. For many purposes this may be neglected; for precise work,

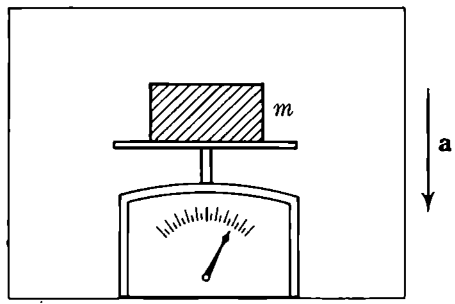


FIG. 6.5. Descending elevator.

however, it must be taken into consideration. More will be said about this in the next chapter. Here we assume *as an approximation* that the acceleration a appearing in Newton's law is referred to the earth's surface.

What is the apparent weight of an object placed on an elevator, when the elevator descends with an acceleration a? To answer this question, consider Fig. 6.5. Let the object of mass m be placed on the pan of a spring scale. The forces upon it are mg downward and the support of the balance pan upward. Call the latter F. The correct acceleration to be used in this example is that relative to the earth, a, not relative to the elevator (which is zero!). Hence we have

$$mg - F = ma$$

The apparent weight of the body is F, the force of the balance. This is $m(g - a)$. If $a = g$, the body has no apparent weight.

PROBLEMS

1. At the Lick Observatory, Mt. Hamilton, Calif., the acceleration of gravity is found to be 979.7 cm/sec². What is the value of g in ft/sec² at this location?

2. Two masses m_1 and m_2, connected by a flexible cord, are placed on a pair of smooth inclined planes as shown in Fig. 6.6. There is no friction at the point A. Find the acceleration of the system and the tension in the cord. If the system is to be in equilibrium, what must be the relation between the masses and the angles?

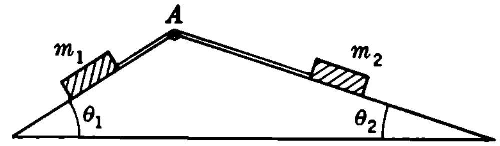


FIG. 6.6. Problem 2.

3. A 1,200-lb elevator is pulled upward by a cable with an acceleration of 5 ft/sec². Find the tension in the cable.

4. A block slides down an inclined plane with constant velocity when the slope of the plane is 10°. If the slope is increased to 25°, what will be the acceleration of the same block down the plane?

5. A mass m slides down a smooth inclined plane of slope θ and length d. Find its velocity at the bottom of the plane. Show that this is equal to the velocity it would have acquired in falling a distance equal to the height of the plane.

6. An Atwood machine (see Sec. 6.5) consists of two masses suspended by a flexible cord over a massless pulley. If one of the masses, weighing 11 lb, has an upward acceleration of 12 ft/sec², find the other mass.

7. An automobile weighing 2 tons is traveling 60 miles/hr. The brakes are applied, and it comes to rest with uniform deceleration after going 154 ft. Find the force exerted to stop the automobile.

8. A spring with stiffness $k = 500$ dynes/cm is used in an experiment like that described in Sec. 6.1. If, when the spring is extended 4 cm and released, it gives an instantaneous acceleration of 25 cm/sec² to the mass m attached to it, what is the magnitude of m?

9. A spring and four masses m_1, m_2, m_3, and m_4 are used in an experiment similar to that described in Sec. 6.1. The spring, when extended 5 cm, imparts to the masses instantaneous accelerations of 20 cm/sec², 5 cm/sec², 35 cm/sec², and 40 cm/sec², respectively. If m_1 is chosen as the standard mass, compute the masses of m_2, m_3, and m_4 in terms of it.

10. Find in terms of g, m_1, and m_2, an expression for the tension T in the cord of the Atwood machine described in Sec. 6.5.

11. Show that v^2/r has the dimensions of acceleration.

12. A body of mass 75 gm rests on a horizontal frictionless table. What horizontal force is required to give it an acceleration of 3 cm/sec²? If the force is applied downward at an angle of 60° with respect to the horizontal, what magnitude is necessary?

13. A body weighing 73 gm starts from rest at the top of an inclined plane, 2 m long, making an angle of 30° with the horizontal. If the body takes 8 sec to reach the bottom, what is the coefficient of sliding friction between it and the plane?

14. A wooden block is placed on an adjustable inclined plane, and the angle of inclination θ is increased gradually until the block just begins to slip. If $\theta = 21°$ at this point, what is the coefficient of friction between the block and the plane?

15. In an experiment like that which Galileo conducted (see Sec. 6.6) a 30-lb mass is observed to slide down a nearly frictionless inclined plane ($\theta = 20°$) with an acceleration of 11.00 ft/sec². Calculate the acceleration of gravity g from these data.

16. A block of mass m is projected up a smooth inclined plane of slope θ with an initial velocity v_0. How far up the plane will it travel? How much time will elapse before it is again at the foot of the plane?

17. A block of mass 3 lb slides down an inclined plane ($\theta = 25°$). If the coefficient of friction is 0.4, find the acceleration of the block in ft/sec^2; in cm/sec^2.

18. At a given instant a 2-ton airplane has an acceleration directed horizontally forward of 5 ft/sec^2. If the thrust of its propeller is 1,000 lb, find the magnitude and direction of the force that must be exerted by the air on the airplane to keep it in level flight.

19. A man twirls a mass of 0.5 lb in a circle on the end of a 3-ft string. If the mass makes one complete revolution in 1.5 sec, what is the tension in the string?

20. A rectangular block of mass 2 lb is accelerated along a surface by a force of 80 poundals applied at an angle of 60° with the surface. If the coefficient of friction is 0.5, how long must the force act to give the block a velocity of 20 ft/sec?

21. A 30-gm mass moving initially to the right is acted on by a constant force of 600 dynes to the left. It returns to its initial position in 5 sec after the force starts to act. Find its initial velocity. Where and when does it reverse its direction of motion?

22. Calculate the acceleration of the system shown in Fig. 6 7. The coefficient of sliding friction between the 4-lb mass and the inclined plane is 0.2. (Assume the pulleys to be frictionless.)

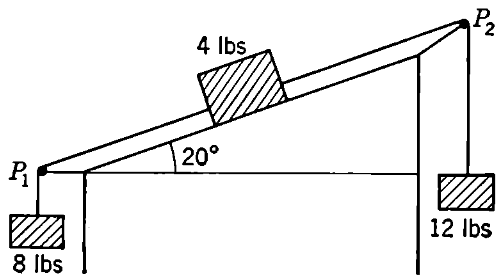


FIG. 6.7. Problem 22.

23. A 50-gm block is initially at rest on a table. If a horizontal force of 20,000 dynes acts on it for 2 sec, how far from its initial position does it finally come to rest? Take the coefficient of sliding friction to be 0.3.

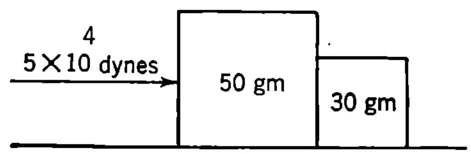


FIG. 6.8. Problem 24.

★**24.** A force of 5×10^4 dynes is exerted on a 50-gm block, which in turn pushes a 30-gm block (see Fig. 6.8). If the blocks move on a frictionless surface, what force does one exert upon the other?

If the coefficient of friction between the 50-gm block and the surface is 0.3 and that between the 30-gm block and the surface is 0.4, find the acceleration of the system and the force that one block exerts on the other.

★**25.** A 13-lb block is pulled along a smooth table by a string which runs over a small frictionless pulley and on which hangs a 3-lb block. If the system starts from rest and the 3-lb block is initially 6 ft from the floor, find the time that elapses before the 3-lb block hits the floor. What is the velocity of the system when this occurs?

26. A 5-lb mass is attached to a spring scale, which is suspended from the roof of an elevator. What does the scale read (*a*) if the elevator ascends with a constant velocity of 20 ft/sec; (*b*) if the elevator descends with an acceleration of 4 ft/sec^2; (*c*) if the elevator ascends with an acceleration of 2 ft/sec^2; (*d*) if the cable breaks and the elevator falls freely?

★**27.** A plumb bob hangs from the roof of a railroad car. What angle θ will it make with the vertical when the train has an acceleration of 8 ft/sec^2? Find the general expression giving θ as a function of a. (This device constitutes a simple form of accelerometer.)

★**28.** Using length, time, and force as basic dimensions, find the dimensions of mass.

CHAPTER 7

SPECIAL MOTIONS IN A PLANE

7.1. The Motion of a Projectile in a Vacuum. A body, treated here as a particle, is projected at an angle θ with the horizontal, its initial velocity being V (cf. Fig. 7.1). The origin of a rectangular system of coordinates is taken at the point of projection. At any instant during its flight (for instance, at the point P) the projectile is subject to only one force, *viz.*, its weight mg, and this acts vertically downward. If therefore we resolve its acceleration at P into one component $a_y = d^2y/dt^2$

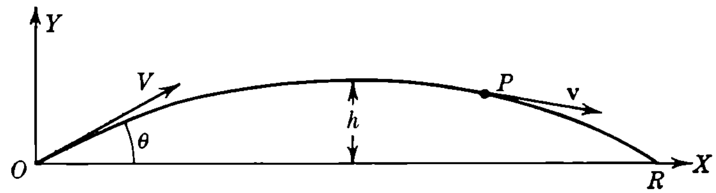


FIG. 7.1. Flight of a projectile.

along Y and another $a_x = d^2x/dt^2$ along X, we find from Newton's second law

$$m\frac{d^2y}{dt^2} = -mg \qquad m\frac{d^2x}{dt^2} = 0 \tag{7.1}$$

From these, velocity and position at any instant are obtained by integration. The vector **v**, which is tangent to the curve of flight called the *trajectory*, has components $v_y = dy/dt$ and $v_x = dx/dt$, and the displacement has components y and x. Integrating once we find

$$\frac{dy}{dt} = -gt + c_1 \qquad \frac{dx}{dt} = c_2 \tag{7.2}$$

The constants c_1 and c_2, which are not determined by the mathematical equations, must be chosen to fit the initial condition of our problem. We have assumed that, at $t = 0$, v_x was $V \cos \theta$ and v_y was $V \sin \theta$. Putting $t = 0$ in Eqs. (7.2) we see that

$$\left.\frac{dy}{dt}\right|_{t=0} = c_1 \qquad \left.\frac{dx}{dt}\right|_{t=0} = c_2$$

Hence $c_1 = V \sin \theta$ and $c_2 = V \cos \theta$, and Eqs. (7.2) take the definite form

$$\frac{dy}{dt} = v_y = -gt + V \sin \theta \qquad \frac{dx}{dt} = v_x = V \cos \theta \tag{7.3}$$

From these results it is apparent that the vertical velocity is the same as that in free fall, with the initial velocity $V \sin \theta$ superposed; the horizontal velocity remains constant and equal to its initial value. The absence of "interference" between the two motions is sometimes referred to as the "independence of vertical and horizontal motions." It is but a consequence of the vector nature of accelerations.

To obtain the instantaneous position of the projectile Eqs. (7.3) are integrated once more. Thus

$$y = -\tfrac{1}{2}gt^2 + V \sin \theta \, t + c_3$$
$$x = V \cos \theta \, t + c_4$$

The new constants c_3 and c_4 are again determined in conformity with the initial conditions. Since both y and x are zero at $t = 0$, the constants must vanish. This leaves us with

$$y = -\tfrac{1}{2}gt^2 + V \sin \theta \cdot t \qquad x = V \cos \theta \cdot t \tag{7.4}$$

Equations (7.4) are called the equations of motion of the projectile; they tell us its position at every t.

It is not easy to see from them, however, what the path or trajectory is. To obtain this information one must eliminate t between Eqs. 7.4. According to the second, $t = x/V \cos \theta$; when this is introduced in the first there results

$$y = -\frac{1}{2} g \frac{x^2}{V^2 \cos^2 \theta} + x \tan \theta \tag{7.5}$$

which is the equation of a parabola with its axis parallel to the Y axis.

7.2. Time of Flight, Range of Projectiles. The time of flight is easily obtained from Eqs. (7.4). The y coordinate is zero at two times—at the initial instant and when the projectile reaches the ground again. Correspondingly the equation

$$0 = -\tfrac{1}{2}gt^2 + V \sin \theta \cdot t$$

has two solutions, $t = 0$ and $t = 2V \sin \theta/g$. The latter must be the time of flight.

The time at which the missile reaches its highest point is that for which $v_y = 0$. From Eqs. (7.3) this is seen to be $t = V \sin \theta/g$, which is one-half the time of flight. Hence one may conclude that the time of ascent is equal to the time of descent for our frictionless projectile. If air resistance were taken into account, this would not be quite true.

The *range* (R in Fig. 7.1) is the distance x reached when the body returns to ground, that is, x at $t = 2V \sin \theta/g$. From Eqs. (7.4), this is

$$R = \frac{2V^2 \sin \theta \cos \theta}{g} = \frac{V^2}{g} \sin 2\theta \tag{7.6}$$

because $\sin 2\theta = 2 \sin \theta \cos \theta$. The range depends on the initial velocity V and on θ, the angle of elevation. If, for a given V, we wish to obtain the *maximum* range, θ must be taken to be 45°, because sin 90° is the maximum value of the sine function. One may also obtain R by putting $y = 0$ in Eq. (7.5) and then solving it for x.

The maximum elevation of the projectile, h, is the distance y at $t = V \sin \theta/g$; it may be found by putting this value of t into Eqs. (7.4).

$$h = -\frac{1}{2} g \frac{V^2 \sin^2 \theta}{g^2} + V \sin \theta \frac{V \sin \theta}{g} = \frac{1}{2} \frac{V^2 \sin^2 \theta}{g} \tag{7.7}$$

It can also be obtained from the equation for the path; if the maximum of Eq. (7.5) is determined by differentiation, the same result is found.

A bombsight is an optical device utilizing the basic principles of the preceding section, though with very great elaborations. To understand its function we consider an airplane in horizontal flight (Fig. 7.2). Assume its height above ground to be h and its velocity V. The target T is sighted forward and at an angle ϕ with the direction of flight. The bombardier must know at what value of ϕ he has to release the bomb. Neglecting air resistance the bomb describes the second half of a parabolic trajectory. If it takes t sec to fall the distance h, then

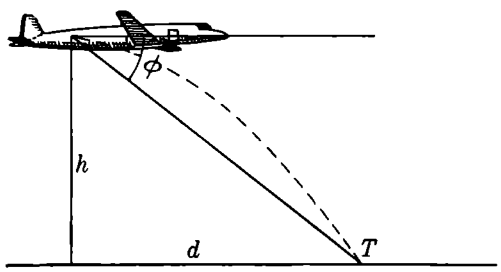


FIG. 7.2. Plane dropping bomb.

$$h = \frac{1}{2} gt^2 \qquad \text{or} \qquad t = \sqrt{\frac{2h}{g}}$$

During this time t it travels a distance $d = Vt$ along the horizontal, and

$$d = V \sqrt{\frac{2h}{g}}$$

But $\tan \phi = h/d$; hence the correct angle of release is

$$\phi = \tan^{-1} \frac{h}{d} = \tan^{-1} \sqrt{\frac{hg}{2V^2}} \tag{7.8}$$

An automatic bombsight not only sets a telescope for the correct angle ϕ (h and V are supplied to it directly through connections with altimeter and tachometer) but also makes adjustments for wind velocity, departure from horizontal flight, air resistance, and other factors.

The true trajectory of a body in air is qualitatively given by the solid curve in Fig. 7.3. For high velocities of projection the asymmetry is very considerable, the range is shorter than that given by Eq. (7.6), and the height attained smaller than predicted by Eq. (7.7). Even before Galileo's time interest in gunnery was great, and considerable information about actual paths of flight had been collected. But the analysis was based on the following interesting though erroneous assumption: The path, conceived to be the dotted trajectory in Fig. 7.3, consists of a straightline part, a, in which the motion is "violent," or nonnatural, and a curved downward part, b, representing natural motion in Aristotle's sense.

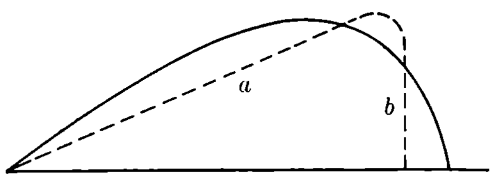


FIG. 7.3. Actual trajectory.

***7.3. Exterior Ballistics.** The subject that deals with the flight of shells and bullets in all its aspects is called *exterior ballistics* in contradistinction to interior ballistics, the study of the behavior of projectiles inside guns. The general problem of the motion of projectiles is too difficult and complicated for treatment in this book. However, the physical principles and the manner of attack are very interesting and can be briefly described. One main concern is, of course, to take account of the action of the resisting medium, air, upon the projectile.

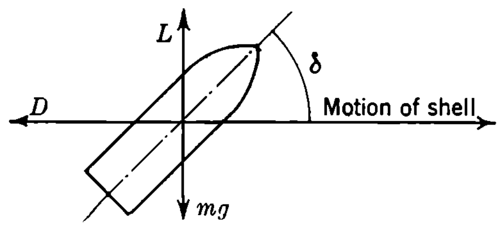


FIG. 7.4. Forces on a shell.

Air resistance depends upon the velocity and the shape and size of the moving object. In Fig. 7.4 we have drawn a shell in flight to the right. The axis of the shell makes angle δ, called the "angle of yaw," with the direction of motion. The three forces acting on it are its weight mg, the "drag" D due to air resistance, and the "cross-wind lift" L, which is mainly due to a difference in air pressure below and above the shell. With this information it is easy to set up the equations from which the motion must be found (Newton's second law),

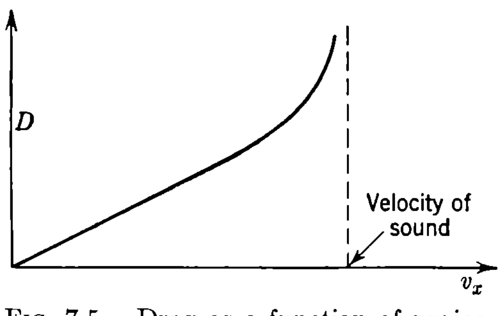


FIG. 7.5. Drag as a function of projectile velocity.

$$m\frac{dv_x}{dt} = -D$$

$$m\frac{dv_y}{dt} = L - mg$$

In these two harmless-looking equations, D and L are functions of v_x and also of δ. A qualitative plot of D for a given angle δ is shown in Fig. 7.5. It is interesting to note that D becomes very large for velocities approaching that of sound. The equations above cannot be solved in general by known procedures. So-called "numerical methods," which involve substitution of D and L from graphs and tables and integration by a process of summing, must be used in their solution. Modern electronic calculating machines have been, and will increasingly be, of very great help in the theory of ballistics.

***7.4. Vertical Motion of a Raindrop.** Actual bodies, when falling in air, are subject not only to their weight but also to a retarding force due to air resistance. This is greater the greater the velocity of the falling objects and for small speeds is proportional to v. For definiteness we consider a falling raindrop of mass m. Reckoning downward displacements as positive, Newton's second law states

$$m\frac{dv}{dt} = mg - bv \tag{7.9}$$

when b is written for the force of friction per unit of velocity. So long as $mg > bv$, dv/dt will be positive and v will increase. Finally, however, dv/dt will be zero, and the drop will fall with constant velocity. This so-called "terminal velocity" is attained when $bv = mg$; hence

$$v_t = \frac{mg}{b} \tag{7.10}$$

A similar result holds for all falling objects. A person falling from a very great height attains a terminal velocity of about 150 miles/hr. A parachute is a device that makes the constant b very large.

7.5. Centripetal and Centrifugal Forces. Many a fun-loving youth remembers vividly his experience on the horizontal wheel, now gradually disappearing from fairgrounds because it is thought to expose its occupants to undue risk. As the wheel spins faster and faster, a person sitting on it will be thrown off. And if you ask him, he will assert that he was pushed off by "centrifugal force." This, we shall find, is an error. But let us analyze the situation.

To make things as simple as possible, assume an object to be tied by a string to a vertical shaft S (Fig. 7.6) and to revolve about S at constant speed v in a horizontal plane. The length of the string is r, and the surface (perhaps a horizontal table top) on which the motion takes place is perfectly smooth. The mass has then no vertical acceleration; its weight is neutralized by the support of the surface.

We know from Sec. 5.9, however, that the mass has a *centripetal* acceleration of magnitude v^2/r. It is therefore not in equilibrium but requires a force that will supply this acceleration. By Newton's law this force must have the magnitude mv^2/r, in absolute units; and it must be in the direction of the acceleration, *i.e.*, *centripetal* (Latin *petere*, to seek). It is indeed provided by the string, which, being under tension, *pulls* the mass toward O. The tension thus equals mv^2/r, as a spring balance inserted between O and m would show.

It is true, of course, that the mass while rotating has a centrifugal tendency; for it flies off along a tangent when the string breaks. But this is precisely because, when the string ceases to supply the centripetal force, the mass must execute force-free motion along a straight line in accordance with Newton's first law. Hence there is certainly no centrifugal *force* acting on the mass m.

Similarly the person was not pushed off the wheel by any centrifugal force; he began to slide off because friction, which previously supplied the *centripetal* force needed to hold him in a condition of centripetal acceleration, was no longer sufficient to do so. In fact he began to slide off when mv^2/r (m is his mass) became greater than the force of friction. But the latter, as we know from Sec. 6.6, is given by μmg if μ is the coefficient of friction. Thus sliding occurs the instant when

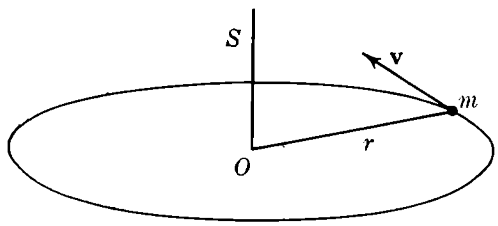


FIG. 7.6. Revolving mass.

$$\frac{mv^2}{r} = \mu mg \qquad (7.11)$$

or, since $v = \omega r$, when $\omega = \sqrt{\mu g/r}$. If the angular velocity and the distance from the center, r, were determined, the wheel could be used for measuring μ. From this result it is also apparent that a person near the center is harder to push off (requires a greater ω) than one near the edge of the wheel.

We have seen that the mass m in Fig. 7.6 is subject to a centripetal force only. Let us now consider the shaft S about which m is rotating. The string pulls on it; S is indeed subject to a *centrifugal* force at O. This centrifugal force is the reaction, in the sense of Newton's third law (which will be carefully considered in the next chapter), to the centripetal force on m; it does not act on m, nor does the centripetal force act on S. To summarize our conclusion: The revolving object is subject to a centri*petal* force, while the center to which the revolving object is tied experiences a centri*fugal* force. Usually, as in the present example, the centrifugal force on S is neutralized by another force due to the support holding S in place, so that the resultant force on S is zero. Hence S has no acceleration.

7.6. Motion on Curves. The same principles apply to the motion of automobiles and trains on curves. In Fig. 7.7*a* is drawn an assembly consisting of an axle, two wheels, and a load rounding a curve whose center of curvature is a distance R to the right and in the plane of the diagram. Whether it is moving toward or away from the reader is immaterial. The center of gravity C must be subject to a resultant force mv^2/R to the right if the vehicle is to stay in its course. But physically there are two forces acting on C: the weight mg, which is directed vertically downward, and the reaction of the road, F, whose direction and magnitude adjust themselves so as to make the motion possible. In

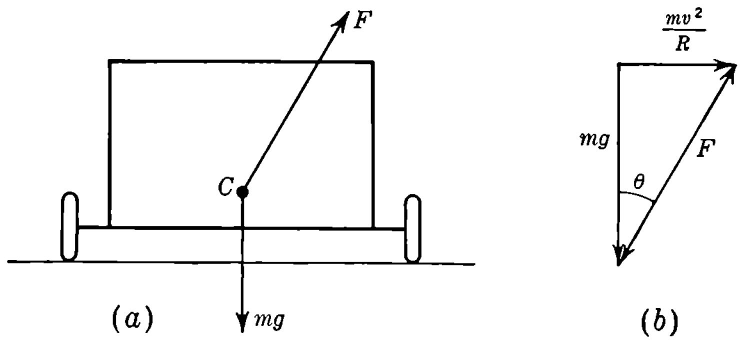


FIG. 7.7. Automobile rounding a curve.

other words, F will be such that, when it is added vectorially to mg, there results a force mv^2/R to the right, as shown in Fig. 7.7*b*.

Can the roadbed supply a force F? Since F has a component *parallel* to the road surface, friction must be relied upon to provide that part of F. In fact the parallel component of F is seen (from Fig. 7.7*b*) to be mv^2/R itself. On the other hand the maximum value of friction is μmg. Hence the car stays on the road if

$$\frac{v^2}{R} < \mu g \tag{7.12}$$

If the road is slippery, a large v and a small R are dangerous.

The banking of roads makes unnecessary the uncertain action of friction in keeping the car in its course. Clearly, if in Fig. 7.7 we make the road surface perpendicular to F, as in Fig. 7.8, friction is not called into play at all; the car exerts a normal thrust upon the road that is somewhat larger than its weight. Note that in Fig. 7.8 the vectors **F** and **mg** are exactly the same as in Fig. 7.7; only the orientation of the car has changed. The angle of banking, θ, is the same as that indicated in Fig. 7.7*b*.

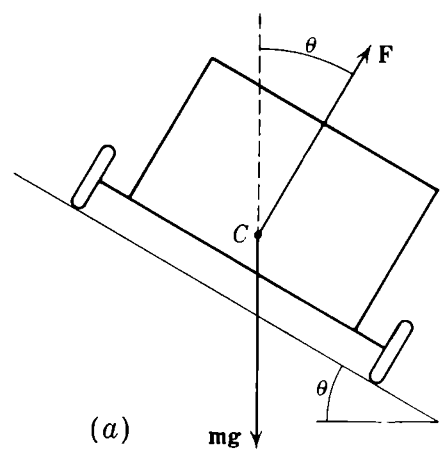


FIG. 7.8. Banked road.

$$\tan\theta = \frac{v^2}{Rg} \tag{7.13}$$

The angle θ, required for proper banking, varies with the speed. In practice a safe mean is chosen.

A more careful analysis of this problem involves consideration of torques as well as forces. Actually the force of the road is not applied at the center of mass as indicated in Fig. 7.7; it is applied at the places of contact between the tires and the road as indicated in Fig. 7.9. At each point of contact two force components are drawn. In order to obtain the result mv^2/R toward the right we must have

$$Y_1 + Y_2 - mg = 0$$
$$X_1 + X_2 = \frac{mv^2}{R}$$

Even though the car has a centripetal acceleration, we must still require that the sum of all torques about C be zero, for the car does not rotate in the plane of the diagram. This means

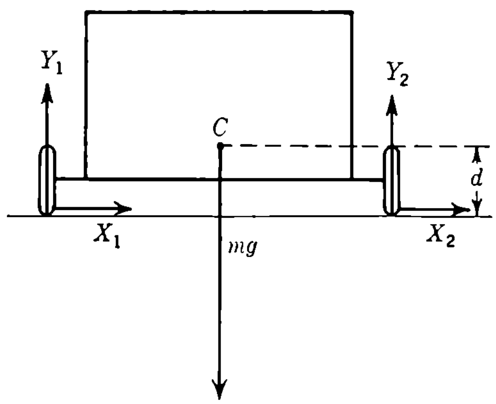


FIG. 7.9. Car on unbanked road.

$$X_1d + X_2d + Y_2\frac{l}{2} - Y_1\frac{l}{2} = 0$$

where l is the length of the axle and d the height of C above ground. On solving the last three equations we find easily

$$Y_1 = \frac{mg}{2} + \frac{mv^2}{R}\frac{d}{l}$$
$$Y_2 = \frac{mg}{2} - \frac{mv^2}{R}\frac{d}{l}$$

The reactions at the two wheels are not equal. When Y_2 becomes negative, the car will overturn. This happens when

$$\frac{v^2}{R} \geq \frac{gl}{2d} \tag{7.14}$$

Here we see the advantage of small d and large l.

***7.7. Effect of the Earth's Rotation upon g.** We have said that the weight of an object, mg, is the force of the earth's attraction upon it and is directed vertically downward. This is not exactly true, as the following careful treatment shows. In Fig. 7.10, P is a point on the earth's surface at latitude L. At P a body of mass m rests on the platform of a balance. *If the earth did not rotate,* the body would be in equilibrium under the action of two forces:

1. The attraction of the earth, mg_0, toward its center O.
2. The push of the platform, W_0, which is equal and opposite to mg_0.

Since the earth rotates, m has an acceleration toward Q; hence a resultant force mv^2/r along the dotted arrow must act upon it. Now the earth's attraction is unchanged by the rotation, hence force (1) is still mg_0 and still directed toward O. But W_0 is changed to W, the force that, when added (vectorially) to mg_0, will yield mv^2/r pointing toward Q. The *negative of* W is said to be the weight of the mass m and is called mg. Its direction is not toward O but toward M (the distance between O and M is vastly exaggerated in the figure; it is about 7 miles for $L = 45°$).

The vector relation between g and g_0 is indicated in Fig. 7.11. Using the cosine theorem we find from the figure

$$g^2 = g_0^2 + \frac{v^4}{r^2} - 2g_0\frac{v^2}{r}\cos L$$

or, in terms of the angular velocity ω of the earth and R, its radius,

$$g^2 = g_0^2 + R^2\omega^4\cos^2 L - 2g_0R\omega^2\cos^2 L$$

because $r = R\cos L$.

The second term on the right is much smaller than the others and can be neglected. Hence

$$g^2 \cong g_0^2\left(1 - 2\frac{R}{g_0}\omega^2\cos^2 L\right)$$

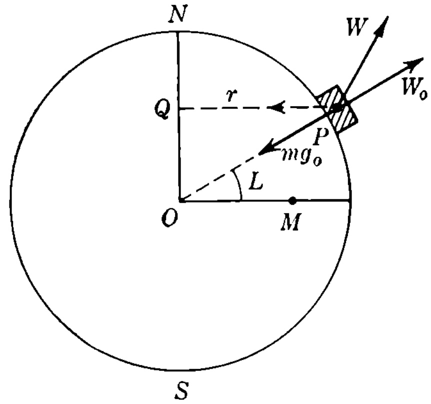


FIG. 7.10. Effect of earth's rotation upon "gravity."

approximately, or, in view of the binomial theorem,

$$g \cong g_0 - R\omega^2\cos^2 L \tag{7.15}$$

The reduction in the value of g_0 is greatest at the equator ($L = 0$), zero at the poles. The student should inspect Table 6.2 to see whether or not this effect is observable. The quantity g_0 is called the "true value of the gravitational acceleration"; it varies with altitude and to some extent with the local distribution of rocks in the earth's surface. For further information on the matter consult Chap. 11.

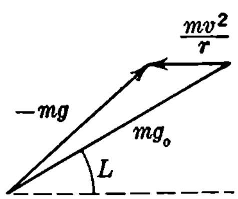


FIG. 7.11. Relation between g and g_0.

7.8. Further Examples.

If an airplane is to make a horizontal turn without slipping, it must assume a banking angle which is such that the reaction of the air upon the airplane, F, and its weight produce a resultant mv^2/r, as shown in Fig. 7.12.

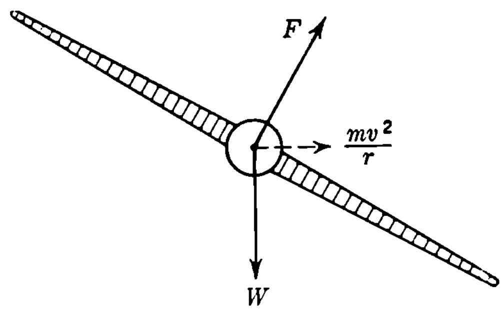


FIG. 7.12. Plane in curved flight.

In the cream separator a mixture of colloidal particles is whirled very rapidly. The denser particles, requiring a greater force mv^2/r than the others, exert a greater *centrifugal* force upon the medium tending to hold them in place. The medium, being fluid, cannot supply the reaction to this force, so that the particles migrate outward, the denser ones faster than the others. Since milk is denser than cream, there will be a greater concentration of milk near the outer wall of the centrifuge, the cream remaining near the center.

In modern organic chemistry and biology, high-speed rotors, called "ultracentrifuges," are used for similar purposes. Some of them rotate while suspended in air and attain angular velocities as high as 1 million rpm.

A simple conical pendulum is a mass, supported by a string, which is allowed to swing in a horizontal circle (cf. Fig. 7.13). The tension T and the weight mg have a resultant (dotted arrow) equal to $m\omega^2 r$. Resolving the forces we find

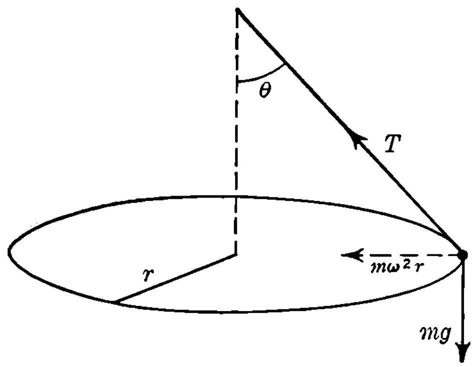


FIG. 7.13. Simple conical pendulum.

$$\left.\begin{aligned} mg &= T\cos\theta \\ m\omega^2 r &= T\sin\theta \end{aligned}\right\} \qquad (7.16)$$

But if l is the length of the string, $r = l \sin\theta$ and we obtain on eliminating T from Eqs. (7.16)

$$\cos\theta = \frac{g}{\omega^2 l} \qquad (7.17)$$

The greater ω, the greater will be the angle θ.

The trick of putting a substance (*e.g.*, water) in a pail and swinging it in a vertical circle without having the object drop out is familiar and easily explained. In the instantaneous position at the top of a swing the substance is subject to two downward forces, its weight and the push P of the bottom of the pail against it. These two, when added, must produce mv^2/r. Thus

$$mg + P = \frac{mv^2}{r}$$

When $v^2/r = g$, P is zero and the pail exerts no force on the object. For smaller velocities it will fall out. For unless the object is glued to the pail, P cannot be negative. But according to our equation P is negative whenever

$$mg > \frac{mv^2}{r}$$

Therefore the object will fall out if $v^2 < rg$.

PROBLEMS

1. A rifle with a muzzle velocity of 1,500 ft/sec is fired horizontally 1 ft above level ground. At the same instant another bullet is allowed to fall freely from a height

of 1 ft. Which bullet strikes the ground first? How far does the first bullet travel before striking the ground? (Neglect air resistance.)

2. A piece of ice slides down a roof 20 ft long and inclined 30° with the horizontal. The edge of the roof is 32.2 ft above the ground and 1 ft from the side of the building. If the ice starts from rest at the top of the roof and the coefficient of friction is 0.1, how far from the edge of the building does the ice hit the ground?

★3. A bombing airplane dives toward the ground at an angle of 60° with a speed of 420 miles/hr. The airplane releases its bomb at an altitude of 1,000 ft. (*a*) Find the time that has elapsed when the bomb hits the ground. (*b*) Find the distance between the point directly below the airplane when it releases the bomb and the point where the bomb strikes the ground.

★4. Using Eq. (7.5) determine the general expressions for the range and greatest height of the projectile as functions of θ. (HINT: When y is maximum, $dy/dx = 0$.)

5. By setting $dR/d\theta$ equal to zero in Eq. (7.6) show that maximum range is obtained when $\theta = 45°$.

6. A mass m_1, projected from the origin at time $t = 0$, is initially aimed at a mass m_2 at the point (x',y'). If m_2 is allowed to fall freely at $t = 0$, show that the masses will collide in mid-air! Find the expression for the time at which they collide.

7. A projectile is shot off with an initial velocity V making an angle θ with the level ground. The projectile reaches a maximum height h and returns to ground T sec after it is fired. Neglecting air resistance sketch roughly, with time t as the abscissa, the following graphs: (*a*) the height y vs. t; (*b*) the vertical component of velocity v_y vs. t; (*c*) the horizontal component of velicity v_x vs. t; (*d*) the acceleration a of the projectile vs. t.

★8. For a raindrop of diameter 2 mm, the value of b in Eq. (7.10) is approximately 0.01 gm/sec. Find its terminal velocity.

9. A circular curve of radius 300 ft is to be built in a highway. It is to be designed to accommodate automobiles traveling at 30 miles/hr. Determine the proper angle of banking.

10. A boy and his bicycle weigh 150 lb and travel with a speed of 10 ft/sec. If the coefficient of friction between the tires and the ground is 0.5, determine the radius of the smallest circle that the boy can execute on horizontal ground without slipping. What angle must he and the bicycle make with the vertical in executing this circle?

★11. A railroad train travels along a track coincident with the equator in a direction opposite to that of the rotation of the earth. When the train is at rest, a spring balance, suspended from the ceiling of one of the cars and supporting a certain mass, reads W' lb. If the train is traveling with speed v, show that the balance reads very nearly

$$W = W'\left(1 - \frac{2v\omega}{g}\right)$$

where ω is the angular velocity of the earth and g the acceleration of gravity.

12. A man swings an object of mass M in a vertical circle with angular velocity ω, the length of his arm being l. Find the force exerted by his arm (*a*) at the top of the circle; (*b*) at the bottom of the circle; (*c*) when the arm is horizontal.

13. A 2-ton airplane does a loop-the-loop with a speed of 200 miles/hr. What is the radius of the largest loop that the airplane can make without tendency to fall at the top?

★14. A man weighing 150 lb leaps from an airplane and floats to the ground supported by a 50-lb parachute. The retarding force of the air on the parachute is proportional to the velocity. The man reaches the ground with a constant velocity equal to that which he would have acquired in falling without a parachute from an 8-ft wall. What is the value of b for the man and parachute?

15. A lightproof soundproof railway car travels with a constant speed v around a circular track of radius R that is banked at an angle $\theta = \tan^{-1} v^2/gR$. (*a*) In what direction will a plumb line suspended from the ceiling of the car point? (*b*) What experiment can a man in the car perform to show that he is accelerated with respect to the earth? (Assume he does *not* know the value of g.)

★16. The balls of a steam-engine governor (see Fig. 7.14) weigh 10 lb apiece. Neglecting the weight of the arms find the angular velocity of the governor. What is the tension in the supporting arms?

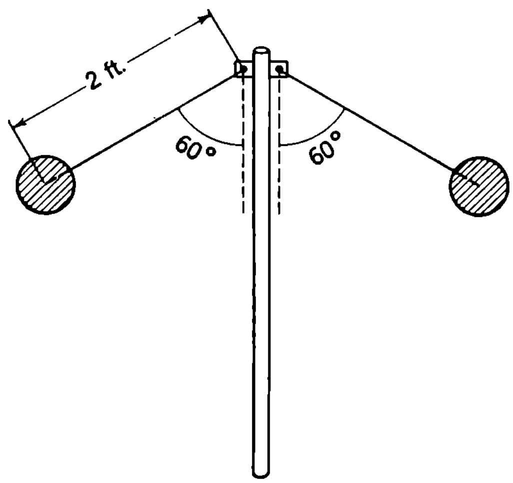


FIG. 7.14. Problem 16.

★17. A hemispherical bowl of radius a is rotated about its vertical axis with angular velocity ω (see Fig. 7.15). If a marble of mass m is placed in the rotating bowl, it ultimately settles at a point a distance d from the axis. Find d as a function of ω.

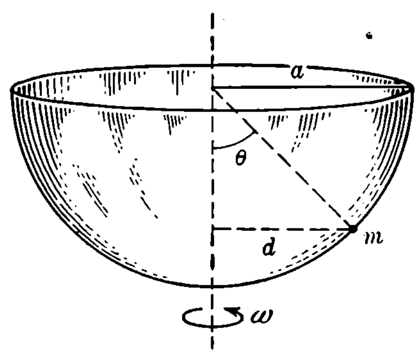


FIG. 7.15. Problem 17.

★18. Integrate Eq. (7.9) to obtain

$$v = \frac{mg}{b}\,(1 - e^{-bt/m})$$

CHAPTER 8

WORK, MOMENTUM, CONSERVATION PRINCIPLES

8.1. Newton's Third Law of Motion[1]. The usefulness of Newton's second law must now be apparent. It was the key to the solution of all the problems discussed in the preceding chapter and will continue to be of importance in the subsequent portions of this book. Before going on, however, attention must be given to the third law, which, when coupled with the second, provides the entire blueprint of the subject of particle mechanics. Thus far the third law has been mentioned in only one instance: it served to make clear the distinction between centripetal and centrifugal force.

"For every action there exists an equal and opposite reaction" is a simple statement of the third law. Action and reaction are literal translations from the Latin; each is meant to represent a *force*. A more modern version of the third law would be: *For every force there exists an equal counterforce.*

Those untrained in physics sometimes see a contradiction between the second and third laws of motion. If acceleration, they say, is proportional to the resultant force and the third law requires that every force have a counterforce, this pair adding up to zero, how can a body ever be accelerated?

The answer is simple. Force and counterforce *never* act on the same body. The resultant force producing acceleration is the sum of all forces *acting on* the body under consideration; this sum does not include the reactions. An example will make this clear.

A book rests on a table. Two forces act on the book, its weight mg and the upward push of the table top. It is in equilibrium because these two add up to zero. The reaction to mg is an equal and opposite force by which the book attracts the earth; as stated, it is a force on the earth. The reaction to the table's upward push on the book is a downward push by the book on the table.

The table is in equilibrium under the action of three forces: the downward push of the book, of magnitude mg; its own weight Mg; and the upward force exerted by the earth. These three add up to zero. Their reactions are: upward force by the table on the book; pull by table on earth, of magnitude Mg; push by table on earth of magnitude $(m + M)g$. This kind of analysis can be continued almost indefinitely; the student

[1]Review Sec. 2.2.

might be interested in seeing whether the earth is in equilibrium under the action of the forces thus far enumerated.

Not only does the third law hold for bodies in equilibrium; it is equally true when there are accelerations. When a diver gracefully approaches the water surface, the earth is actually coming up to meet him, though it moves only through an infinitesimal distance. For if the diver, let us say of mass 75 kg, has an acceleration g and therefore a force of $75{,}000g$ dynes acting on his body, an equal and opposite counterforce must be accelerating the earth upward. Now the mass of the earth is approximately 6×10^{24} kg; its mass times its acceleration must also equal $75{,}000g$ dynes. This makes its acceleration equal to $75{,}000g/(6 \times 10^{27})$, or about 1.27×10^{-23} cm/sec^2. The reader will doubtless be interested in computing the distance through which the earth moves in welcoming the diver, assuming a certain height of dive.

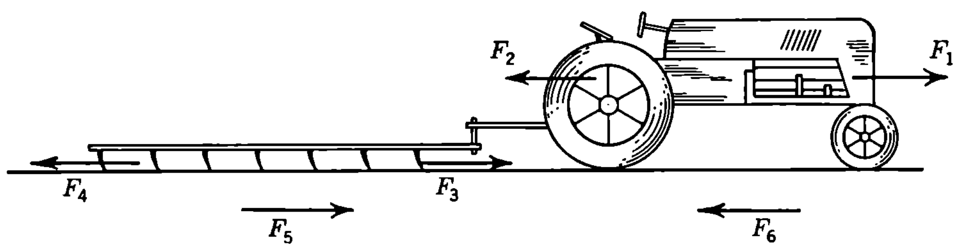


FIG. 8.1. Tractor pulling a harrow.

8.2. Example on Laws of Motion. Figure 8.1 represents a tractor of mass m_1 pulling a harrow of mass m_2 over the earth of mass M. The moving system is subject to a variety of forces, all of which can be resolved into vertical and horizontal components. Since there is no vertical acceleration, the sum of the vertical components of force must be zero and we may leave them out of our consideration here.

Through the agency of friction and the tractor motor the earth exerts a propelling force F_1 upon the tractor; the harrow, on the other hand, pulls on it with a force F_2. F_1 and F_2 are *not* action and reaction and need not be equal in magnitude.

The harrow is subject to a forward force F_3, exerted by the tractor, and a backward drag, due to the earth, F_4. These two need not be equal. But F_3 is the reaction to F_2; hence (the F's here refer only to the magnitudes of the forces)

$$F_3 = F_2 \tag{8.1}$$

Finally the earth is subject to a forward drag F_5 and a backward push F_6. These are the reactions to F_4 and F_1, respectively, and therefore

$$F_5 = F_4 \qquad F_6 = F_1 \tag{8.2}$$

Assume now that m_1 and m_2 move forward with a common acceleration a. Applying Newton's second law to the tractor we find

$$F_1 - F_2 = m_1 a \tag{8.3}$$

Applying it to the harrow,

$$F_3 - F_4 = m_2 a \tag{8.4}$$

Only when $a = 0$, that is, when the system moves with constant speed, will $F_1 = F_2$ and $F_3 = F_4$. In that case *all* the forces drawn in Fig. 8.1 are equal. But let us assume again that $a \neq 0$. The earth, too, must then have an acceleration A, which is given by

$$F_5 - F_6 = MA \tag{8.5}$$

If we add Eqs. (8.3) and (8.4) while using Eq. (8.1), we find

$$F_1 - F_4 = (m_1 + m_2)a \tag{8.6}$$

This result says, interestingly enough, that we should have been justified in treating tractor and harrow as *one system* provided that we considered only the *external* forces F_1 and F_4 as acting upon it; F_2 and F_3 are internal in the sense that they represent forces of one part of the system on another.

On applying Eqs. (8.2) to Eq. (8.5) we obtain

$$F_4 - F_1 = MA$$

and comparison of this result with Eq. (8.6) shows that

$$-MA = (m_1 + m_2)a$$

This is nothing but what is also found by applying Newton's third law to the two interacting systems, earth and tractor-harrow. It shows that, while m_1 and m_2 are accelerated forward, M is accelerated backward; furthermore

$$A = -\frac{m_1 + m_2}{M}\, a$$

From this one may show that, whatever motion is performed by the tractor-harrow combination in one direction, that same motion is executed by the earth in the reverse direction, but on a scale which is diminished in the ratio of $m_1 + m_2$ to M. In the same manner the motion of every object on the earth's surface produces its miniature counterpart in the earth itself.

8.3. Work and Energy. In the present section we shall introduce several new physical quantities. Their definitions will seem strange and unrelated to what has preceded. But let us learn them first in spite of their apparent strangeness. In the next section we shall show how beautifully they coordinate themselves with the laws of motion.

The first quantity is physical *work*. In everyday language, work is loosely defined as any kind of effort or exertion maintained for some time. In physics, *work is force times distance.* This statement is not very precise, for it does not say *what* force and *what* distance are to be multiplied. But it is good enough for cases in which a *constant* force moves an object in the direction in which the force is applied. When force F pushes a mass m over a horizontal plane through a distance d (Fig. 8.2), it does an amount of work equal to Fd. If F is measured in dynes and d in centimeters, the work is given in *ergs*. One erg (Greek *ergon*, work) is the work done when a force of 1 dyne acts through 1 cm. It is the absolute unit of work in the cgs system.

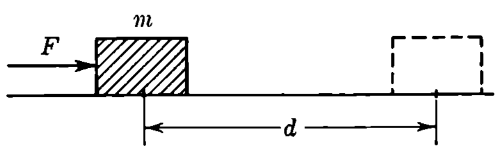


FIG. 8.2. Work = Fd.

Notice that the work done by F depends only on F and d. Neither the mass moved nor the time required for the motion matters. The work is the same whether the mass moved is a pound or a ton; whether the motion takes a second or a year. Furthermore it is quite immaterial what speed the mass acquires while the force acts on it.

Again, when a body falls through a height h, the work done by gravity upon it is mgh. This quantity may be expressed in terms of the velocity change that the body has undergone while it fell through h. If its initial velocity was v_0 and its final velocity v, we know from Sec. (5.5) that

$$v^2 - v_0^2 = 2gh$$

and on multiplying by $m/2$ we see that

$$mgh = \tfrac{1}{2}mv^2 - \tfrac{1}{2}mv_0^2 \tag{8.7}$$

The physical dimensions of work are, in accordance with our definition, those of force times distance,[1]

$$[W] = [MLT^{-2}][L] = [ML^2T^{-2}]$$

Equation (8.7), if correct, implies that $\tfrac{1}{2}mv^2$ also has the dimension of $[W]$; this is easily verified.

$$[\tfrac{1}{2}mv^2] = [M][LT^{-1}]^2 = [ML^2T^{-2}]$$

[1] Henceforth we shall use the symbol W for work, although it was previously often used for weight.

One may interpret Eq. (8.7) by saying: The work done by gravity equals the increase in the quantity $\frac{1}{2}mv^2$. As we shall see later this quantity, $\frac{1}{2}mv^2$, has a significance far more general than the simple example here chosen would indicate. It is called *kinetic energy* and is also measured in ergs, as shown.

8.4. General Definition of Work. The definition of work given in the preceding section is imperfect in two respects.

1. The force may change during the displacement d. We have not specified what value of F must then be taken in forming the product Fd.

2. The motion may occur in a direction different from that in which the force acts. In Fig. 8.2, for example, F might be applied at some angle with the horizontal. In that case, will the work done still be Fd? Or in the gravitational problem, if the body, instead of falling through a distance h, had moved horizontally through the same distance, should we still say that gravity had done an amount of work equal to mgh?

The difficulty referred to in item 1 can be removed only by dividing the total distance over which the motion occurs into segments so small that F may be regarded as constant over each of them. When the spring of Fig. 8.3 is drawn out from x_1 to x_2 by the force of the hand, this force must be proportional at any point such as P to the distance x from the neutral point O, since it is the reaction to the force exerted by the spring on the hand, which in turn obeys Hooke's law (cf. Sec. 4.1). Over each small interval, F is sensibly constant.

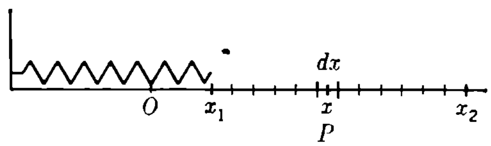


FIG. 8.3.

Mathematically what has been said amounts to this. Instead of defining work as force times distance we define the *element* of work done during an infinitesimal displacement dx as

$$dW = F\,dx \tag{8.8}$$

The work must then be found by integration, and this can be performed when F is known as a function of x. In the present instance, $F = kx$, where k is the stiffness of the spring.

Therefore

$$dW = kx\,dx$$
$$W = \int_{x_1}^{x_2} kx\,dx = \tfrac{1}{2}k(x_2^2 - x_1^2) \tag{8.9}$$

This represents the work done by the hand in pulling the spring from x_1 to x_2.

What should be done about item (2) is not quite so obvious. Clearly the work must be different in cases where the same force produces the displacement, where it acts at right angles and where it opposes the displacement. Theory becomes simplest if we accept the following definition: Let ds be an elementary displacement, F the force doing the work, and θ the angle between F and ds. Then the element of work is (cf. Fig. 8.4)

$$dW = F\,ds\cos\theta \tag{8.10}$$

and this agrees with Eq. (8.8) when $\theta = 0°$.

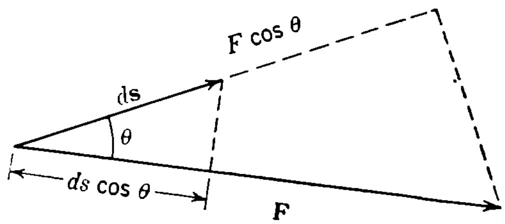


Fig. 8.4. $dW = F ds \cos\theta$.

Equation (8.10) permits other interpretations. We see from Fig. 8.4 that $ds \cos\theta$ is the component of the displacement vector in the direction of the force. Hence dW is the force times the displacement in the direction of the force. But we may say with equal justice that dW is the displacement ds times the force in the direction of the displacement, since $F \cos\theta$ is the component of force along $d\mathbf{s}$.

As a result of Eq. (8.10) *no* work is done when force and displacement are at right angles. Carrying a load horizontally involves no work against gravity in our technical sense of this term, regardless of how tiresome this process may be. Legal interpretations of what constitutes work are quite different from this. Holding a brick with outstretched arm for a time causes great fatigue but is not work. Only when this somewhat arbitrary definition is adopted can work be said to be equivalent to *energy*, as we shall soon show.

When F and ds are in opposite directions, $\theta = \pi$ and $\cos\theta = -1$. The work dW is then negative. The force then has *work done against it;* it is said to be doing *negative* work itself. Thus when I raise a mass m a distance h, I am doing positive work of amount mgh against gravity; *gravity* is doing $-mgh$ units of work. When a body falls a distance h, gravity does $+mgh$ units of work, this time against no other retarding force. Whenever work is done against no retarding force, kinetic energy is produced.

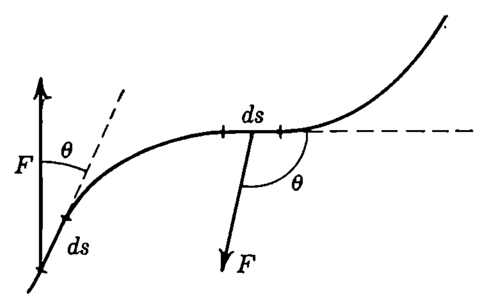


Fig. 8.5. In the calculation of W, θ is often a variable.

Suppose a force of varying (or unvarying) direction acts over a curved path, as in Fig. 8.5, where two sections ds are drawn. The definition (8.10) still holds, but in integrating, dW, the angle θ is a variable, and the calculation of W may be difficult. In such cases a modified form of Eq. (8.10), in which ds and $\cos\theta$ are expressed in Cartesian coordinates, is very useful.

From Fig. 8.6, $\theta = \theta_1 - \theta_2$; hence $\cos\theta = \cos\theta_1 \cos\theta_2 + \sin\theta_1 \sin\theta_2$. Therefore

$$dW = F\,ds(\cos\theta_1 \cos\theta_2 + \sin\theta_1 \sin\theta_2)$$

But $$F\cos\theta_1 = F_x \qquad F\sin\theta_1 = F_y$$

and $$ds\cos\theta_2 = dx \qquad ds\sin\theta_2 = dy$$

because $d\mathbf{s}$ is the vector whose rectangular components are dx and dy. Equation (8.10) may thus also be written in the symmetrical and simple form

$$dW = F_x\,dx + F_y\,dy \tag{8.11}$$

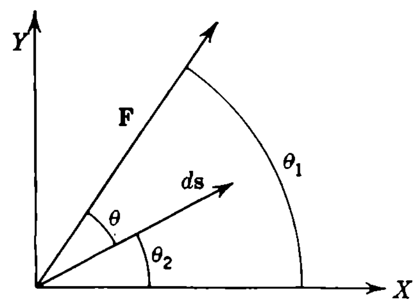


Fig. 8.6. $dW = F_x\,dx + F_y\,dy$.

8.5. Units of Work. The *erg* has already been defined as the work done when a force of 1 dyne acts through 1 cm in its own direction. The corresponding unit in the fps system is the *foot poundal:* it is the work done by one poundal acting through one foot. In the mks system the *newton meter*, or *joule* (after James Prescott Joule, 1818–1889), is defined as the work done by one newton acting through one meter. One joule clearly equals 10^7 ergs. These three are absolute units of work because they are based on the absolute units of force. To find the relation between them, we convert as follows:

$$\begin{aligned} 1 \text{ ft poundal} &= 1 \times (30.5 \text{ cm})(13{,}850 \text{ gm cm/sec}^2) \\ &= 422{,}400 \text{ gm cm}^2/\text{sec}^2 \\ &= 422{,}400 \text{ ergs} \end{aligned}$$

approximately, because 1 ft = 30.5 cm and

$$1 \text{ poundal} = 454 \times 30.5 \text{ dynes.}$$

There are three corresponding units of work that are derived from the gravitational units of force. The *gram centimeter* is the work done by one gram-force over one centimeter of distance. The *foot pound* is the work done by a pound-force over a distance of one foot. The *kilogram meter* is the work done by a kilogram-force over one meter. Clearly

$$\begin{aligned} 1 \text{ gm cm} &= 980 \text{ ergs} \\ 1 \text{ ft lb} &= 32.2 \text{ ft poundals} \end{aligned}$$

The multiplicity of force units, which was explained in Sec. 6.3, gives rise to similar and equally annoying multiplicities of units in all quantities that are derived from forces, as work in this instance. To avoid confusion the reader should review Sec. 6.3, and particularly Table 6.1. We now give a similar table for the units of work.

Table 8.1. *Units of Work*

	fps	cgs	mks
Absolute	ft poundal	erg	joule
Gravitational	ft lb	gm cm	kg m

In the calculation of work by Eq. (8.8) or (8.10) there is no ambiguity with respect to the unit in which W is expressed. Questions may arise in connection with an equation like Eq. (8.7), where the right-hand side, since it contains no force, is noncommittal as to gravitational or absolute units. *Here it must be remembered that the equation was derived with the use of the absolute form of Newton's second law; hence the work mgh on the left is in absolute units.* This will always be the case unless the opposite is expressly stated.

8.6. Worked Examples. *a.* The engine of a freight train exerts a constant force of 6 tons on the train. How much work does it do over a horizontal distance of 1 mile? The answer is clearly

$$W = 12{,}000 \text{ lb} \times 5{,}280 \text{ ft} = 6.34 \times 10^7 \text{ ft lb}.$$

b. A boy pulls a sled by a rope which is inclined at 30° with the horizontal and is under tension of 10 lb. How much work does he do in drawing the sled 100 ft? Answer:

$$W = 10 \text{ lb} \times 100 \text{ ft} \times \cos 30° = 866 \text{ ft lb}.$$

c. How much work is done when a mass m is moved a distance l along a smooth incline making an angle ϕ with the horizontal? The answer to this example is not unique. To move the mass up we require *at least* a force of magnitude $mg \sin \phi$ (component of the weight) along the plane. If the force is greater, the mass will move up with an acceleration, *i.e.*, it will acquire kinetic energy in addition to being raised. We assume here that the force has its minimum value, so that the body moves without acceleration or with an acceleration so small as to be negligible. The work is then $mg \sin \phi\, l$ since the force is constant. If m is expressed in grams, l in centimeters, and $g = 980$ cm/sec², the result is in *ergs*.

But we note that $l \sin \phi = h$, the height of the incline. What we have calculated is therefore the work mgh that would have been done in raising the mass vertically from A to B, Fig. 8.7.

One can also work this example with the use of Eq. (8.10). To be precise, we shall compute now the work done by gravity when the mass is allowed to slide slowly down the plane against a suitable retarding force. Then

$$W = mgl \cos \theta \tag{8.12}$$

where θ is the angle indicated in Fig. 8.7. But $\cos \theta = \sin \phi$; hence the same result is obtained. In this case, if there were no retarding force, the work done by gravity would obviously be the same, for the value of the force of gravity is mg regardless of whether or not there is an opposing force; but the velocity of the mass at the bottom would be different.

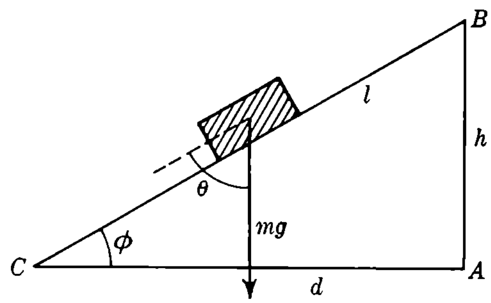


FIG. 8.7. Work done on an incline.

d. If the incline in the preceding example is rough, the coefficient of friction being μ, then the work done is (cf. Fig. 8.7)

$$W = (mg \sin \phi + \mu mg \cos \phi)l = mgh + \mu mgd \tag{8.13}$$

It is the same as though the object had first been moved from C to A over a surface of equal roughness and then lifted against gravity from A to B.

Suppose the train in Example *a* had moved a distance of 1 mile up a grade of steepness 10°. The answer computed previously was the work done against friction over one horizontal mile. The total work in this instance is the previous answer multiplied by cos 10° plus the weight of the train multiplied by 5,280 ft times sin 10°.

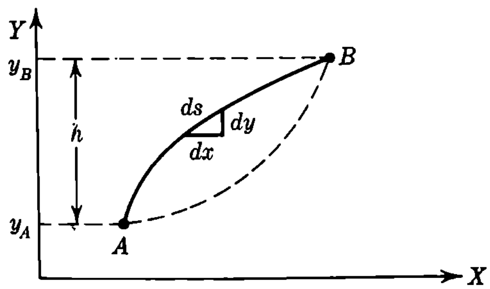


FIG. 8.8.

e. Calculate the work done in moving a mass m from A to B against gravity along the path drawn in Fig. 8.8 (solid curve). Consider the displacement along ds. The element of work, from Eq. (8.11), is

$$dW = F_x\,dx + F_y\,dy \tag{8.14}$$

But $F_x = 0$, and $F_y = mg$, a const. Therefore

$$dW = mg\,dy$$

and

$$W = mg \int_{y_A}^{y_B} dy = mg(y_B - y_A) = mgh \tag{8.15}$$

Curiously enough the result is the same as though the mass had been raised vertically through a distance h. The same result would have been obtained if the mass had been carried from A to B along the dotted path. *The work against gravity is independent of the path* and depends only on the difference in elevation of the initial and the final point.

Another interesting fact is this: In going from A to B along the solid curve I do mgh ergs of work against gravity; in returning along the same path I do $-mgh$ ergs of work; the total work I have done on return is zero. If, instead of returning along the solid curve, I choose to go back along the dotted line, the work done during return is also $-mgh$ ergs. Hence if I go from A to B along the solid and return from B to A along the dotted curve, the total work done on the round trip is zero. *The work done in carrying a mass around any closed circuit against gravity is zero.* Many forces, though not all, have this property of giving back in one part of the cycle the work done in the other. Friction is a notable exception.

8.7. Conservation of Mechanical Energy. Almost everyone has some acquaintance with the so-called "law," or "principle," of conservation of energy. "Energy cannot be created or destroyed." A body at a great height has energy of position; when it falls, this is converted into energy of motion. A coiled spring possesses potential energy, which transforms itself into kinetic energy when the spring is released. These things, and many other familiar instances, are true examples of the conservation of energy. But it is our purpose here to make the meaning of this principle, and the meaning of energy, more precise.

Let us start again with Newton's second law. First we assume the motion of the body of mass m to be along the X axis, and the force is taken to be a function of x, at present unspecified. Then

$$F = m\frac{dv}{dt}$$

Multiply both sides of this equation by dx, and note that

$$\frac{dv}{dt}\,dx = dv\,\frac{dx}{dt} = v\,dv$$

Therefore

$$F\,dx = mv\,dv$$

Suppose the displacement is from x_1 to x_2 and the velocity of the body of mass m at x_1 is v_1, the velocity at x_2 is v_2. We may then integrate on the left-hand side between the limits x_1 and x_2, on the right-hand side between the corresponding limits,

$$\int_{x_1}^{x_2} F\,dx = m\int_{v_1}^{v_2} v\,dv = \tfrac{1}{2}mv_2^2 - \tfrac{1}{2}mv_1^2 \qquad (8.16)$$

The first integral cannot be evaluated unless F is given as a function of x. Whatever this function may be, we can introduce the symbol U for the integral over $-F$, as follows:

$$U = -\int_0^x F\,dx \tag{8.17}$$

U will then also be a function of x; it is called the *potential energy* of the mass when it is in the position x.

Since
$$\int_{x_1}^{x_2} F\,dx = \int_0^{x_2} F\,dx - \int_0^{x_1} F\,dx$$

Eq. (8.16) reads

$$U|_{x_1} - U|_{x_2} = \tfrac{1}{2}mv^2|_{x_2} - \tfrac{1}{2}mv^2|_{x_1}$$

or, on transposing,

$$U|_{x_1} + \tfrac{1}{2}mv^2|_{x_1} = U|_{x_2} + \tfrac{1}{2}mv^2|_{x_2} \tag{8.18}$$

If, as before, we call the quantity $\frac{1}{2}mv^2$ the *kinetic energy*, Eq. (8.18) can be expressed verbally by saying:

The sum of potential and kinetic energies at x_1 is equal to the sum of potential and kinetic energies at x_2.

This analysis has been somewhat abstract. Let us see what it means. When a body is thrown upward, it passes a point a distance x_1 aboveground with velocity v_1 [we here call vertical distances x to conform to the notation of Eq. (8.18)] and a point a distance x_2 above ground with velocity v_2. In this case

$$U(x) = -\int_0^x (-mg)dx = mgx \tag{8.19}$$

(the minus sign before mg indicates that the force is downward, opposite to dx), and Eq. (8.18) reads

$$mgx_1 + \tfrac{1}{2}mv_1^2 = mgx_2 + \tfrac{1}{2}mv_2^2$$

The correctness of this equation should be verified by reference to Sec. 5.6.

As another example we consider a vibrating spring, for which

$$U = -\int_0^x (-kx)dx = \tfrac{1}{2}kx^2 \tag{8.20}$$

To understand this we may refer to Figs. 8.9a and b. In (a) the spring is in its neutral position, with the mass m at 0. In (b) the spring has been drawn out a distance x beyond 0. The force on the mass is to the left and of magnitude kx (cf. Hooke's law); if we reckon x positive to the right, the force on m is $-kx$.

Equation (8.18) reads in this case

$$\tfrac{1}{2}kx_1^2 + \tfrac{1}{2}mv_1^2 = \tfrac{1}{2}kx_2^2 + \tfrac{1}{2}mv_2^2 \tag{8.21}$$

Here x_1 and x_2 are any two values of x. The quantity

$$\boxed{E = \tfrac{1}{2}kx^2 + \tfrac{1}{2}mv^2} \tag{8.22}$$

is therefore constant throughout the vibratory motion of the mass. It is said to be *conserved.*

8.8. Meaning of Potential Energy; Motion in Two Dimensions. What is the physical significance of the quantity U, which mathematical considerations have led us to introduce? On recalling the definition of work [Sec. 8.4, Eq. (8.8)] it is seen from Eq. (8.17) that U is the *negative* of the work done *by* the force F in a displacement from 0 to x. Or, to put it the other way around, it is the work done *against* the force F in this displacement. This is the meaning of mgx in connection with gravitation and of $\tfrac{1}{2}kx^2$ for the elastic spring.

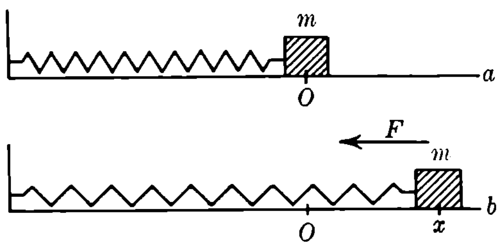


FIG. 8.9.

Equation (8.18) tells us that the sum of U and $\tfrac{1}{2}mv^2$ does not change, although U and $\tfrac{1}{2}mv^2$ change their values individually. We therefore say that the constant sum, each term of which has the physical meaning and the dimension of work, represents a store of work within the moving body, *i.e.*, *energy* (Greek *en*, in; *ergon*, work). The quantity U is potential energy; $\tfrac{1}{2}mv^2$ is kinetic energy. In the motions thus far considered we observe a continual interconversion of potential ($E_{pot} = U$) into kinetic energy (E_{kin}).

Consider an elastic ball bouncing up and down. At the instant when it strikes the floor its E_{pot} is zero, and its E_{kin} is a maximum. At the top of its bounce E_{pot} is a maximum, and E_{kin} is zero. The vibrating spring has maximum E_{pot} and zero E_{kin} at either end of its path, zero E_{pot} and maximum E_{kin} at the middle. In fact we can use Eq. (8.21) to find the velocity of the end point of the spring at any moment if we know its maximum extension x_{max}. For we note that

$$\tfrac{1}{2}kx^2 + \tfrac{1}{2}mv^2 = \tfrac{1}{2}kx_{max}^2 + 0$$

so that the velocity v at x is

$$v = \sqrt{\frac{k}{m}(x_{max}^2 - x^2)} \tag{8.23}$$

Energy is sometimes said to be work resident within a body; the energy that a body possesses equals the work it is capable of doing. From this point of view potential energy is latent work due to position or configuration, and kinetic energy is latent work due to motion. Indeed $\frac{1}{2}mv^2$ is the work a body does while its speed is reduced from the value v to zero. This is obvious from Eq. (8.16) if we put $v_1 = 0$.

The principle of conservation of mechanical energy, Eq. (8.18), which may also be put in the form

$$E_{pot} + E_{kin} = \text{const}$$

does not always hold. It fails, for example, when the spring is subject to frictional forces or when the bouncing ball is not completely elastic. In such cases heat is generated. But it will be shown later that heat, too, is a form of energy, so that, when a term corresponding to it is added onto E_{kin} and E_{pot}, the sum of the three is constant. $E_{kin} + E_{pot}$ is said to be the mechanical energy of a system. Thus, while the mechanical energy of a system is not always conserved, the total energy, which includes all forms of energy, is constant under all known conditions.

Conservation of mechanical energy holds not only for the case of one-dimensional motion discussed in the preceding section. If motion takes place in two dimensions, Newton's law states

$$F_x = m\frac{dv_x}{dt}$$
$$F_y = m\frac{dv_y}{dt}$$

Multiply the first of these by dx, the second by dy, and add.

$$\begin{aligned} F_x dx + F_y dy &= m\left(\frac{dv_x}{dt}dx + \frac{dv_y}{dt}dy\right) \\ &= m(v_x dv_x + v_y dv_y) \end{aligned} \tag{8.24}$$

If we now put

$$-\int_{x=0,\,y=0}^{x,y} (F_x dx + F_y dy) = U \tag{8.25}$$

so that U is a function of the two variables x, y, we can integrate Eq. (8.24), obtaining

$$U\Big|_{x_2y_2}^{x_1y_1} = \tfrac{1}{2}m(v_x^2 + v_y^2)\Big|_{x_1y_1}^{x_2y_2} = \tfrac{1}{2}mv^2\Big|_{x_1y_1}^{x_2y_2} = \tfrac{1}{2}m(v_2^2 - v_1^2)$$

and on transposing we have

$$U\Big|_{x_1y_1} + \tfrac{1}{2}mv_1^2 = U\Big|_{x_2y_2} + \tfrac{1}{2}mv_2^2$$

This is again the law of conservation of mechanical energy.

There is an important question, however, that we have not answered. The student who is thoroughly conversant with the calculus will doubt, perhaps, whether U as defined by Eq. (8.25) makes any sense. That equation represents what is called a *line integral* by mathematicians. To say that U is a function only of the variables x, y, which appear as its upper limit, is not generally correct. Only when it is correct may one introduce a potential energy and speak of the conservation of mechanical energy at all. It may be shown that it is proper for most forces encountered in mechanics.

If F_x or F_y depended on the velocity of motion as well as on the position of the moving particle, an integration of Eq. (8.24) would not lead to the conservation law. Forces depending on v, like friction and others that fail to yield a constant mechanical energy, are said to be nonconservative forces.

***8.9. Summary on Work and Energy.** The infinitesimal work done when a force acts through a distance ds can be expressed in two ways,

$$dW = F\,ds\cos\theta$$

and

$$dW = F_x\,dx + F_y\,dy$$

where θ is the angle between the positive direction of F and ds. When the force is constant and in the direction of the displacement, the first formula takes the simple form

$$W = Fd$$

Energy is a quantity having the same physical dimension as work, *viz.*,

$$[ML^2T^{-2}].$$

It is numerically equal to the work that a body in its given condition can do when its position and its speed are suitably changed.

There are many different kinds of energy such as mechanical energy, heat, sound, light, electric, and magnetic energy. *Mechanical* energy consists of two forms, kinetic and potential. The kinetic energy of a moving object is *always* given by the formula $\frac{1}{2}mv^2$. The potential energy, generally defined by

$$E_{\text{pot}} = U = -\int F\,dx$$

has different forms for objects under the action of different forces. For a mass subject to gravitation,

$$U = mgh \text{ (abs units)}$$

For a body acted on by an elastic spring of stiffness coefficient k,

$$U = \tfrac{1}{2}kx^2 \text{ (abs units)}$$

The principle of conservation of *mechanical* energy says that the total mechanical energy is constant.

$$E = E_{\text{pot}} + E_{\text{kin}} = \text{const}$$

It is not generally true, but only when forces like friction, which convert mechanical energy into other kinds (usually heat), are absent.

The principle of conservation of energy says that the total energy of an isolated physical system is constant. It involves consideration of all kinds of energy, not only mechanical, and is believed to be always true. A steam engine transforms heat into mechanical energy; the electrical generator transforms mechanical into electric energy; a vibrating string transforms mechanical energy into sound, and so forth.

This principle is one of the most useful generalizations of science and has led to numerous discoveries. There is no reason to suppose that its validity is confined to the nonliving world. All available evidence points to its operation in biological processes as well.

The potential energy under gravity, mgh, depends on the choice of level from which the height h is measured. A convenient and common choice is the ground at sea level. It should be emphasized, however, that only *difference* in potential energy and hence differences in height occur in actual applications of the energy principle. This is clear from Eq. (8.18), or the equation following Eq. (8.19), which may be written

$$\tfrac{1}{2}m(v_1^2 - v_2^2) = mg(x_2 - x_1)$$

Hence the choice of zero level for the potential energy has no effect on the physical situation.

★8.10. Efficiency of Machines.[1] The action of machines, analyzed in Sec. 3.13 from the point of view of forces, can also be understood by an application of the principle of the conservation of energy.

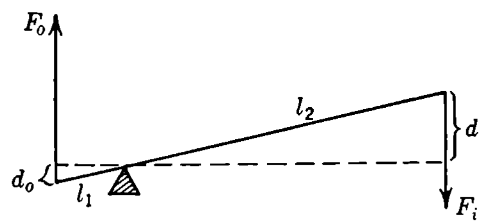


FIG. 8.10. The lever.

For example, if a lever (see Fig. 8.10) is in equilibrium and friction is absent, there can be no waste of energy when both arms are given a slight displacement. Hence, as F_o acts through a distance d_o and F_i through d_i, the output work must equal the input work.

$$F_o d_o = F_i d_i$$

But since $d_o/d_i = l_1/l_2$, this condition is

$$\frac{F_o}{F_i} = \frac{l_2}{l_1}$$

as was previously found [Eq. (3.13)]

For the wheel and axle (Fig. 3.21) the energy principle says

$$F_i r_2 \theta = F_o r_1 \theta$$

[1] (Review Sec. 3.13.)

θ being the angle through which the wheel is turned. This leads to a mechanical advantage $M = r_2/r_1$ as before.

For the differential pulley we have (Fig. 3.22)

$$F_o \frac{r_2 - r_1}{2} \theta = F_i r_2 \theta$$

Here $(r_2 - r_1)\theta$ is the length by which the two ropes holding the weight are shortened; one-half of it is the height by which the weight is raised. Again this formula leads to the result previously obtained.

$$M = \frac{2r_2}{r_2 - r_1}$$

The *efficiency* of a machine is the ratio of its *work output* to its *work input.*

$$\text{Eff} = \frac{W_o}{W_i}$$

In the preceding instances we have assumed the efficiency to be 1, or 100 per cent. Friction always reduces it to a lower value. If the theoretical as well as the actual mechanical advantage M_a of a machine are known, its efficiency can be computed.

Let F_o be the "theoretical" output force, which obeys the relation

$$F_o d_o = F_i d_i$$

The actual work input is $F_i d_i$, but the actual work output is $F_o' d_o$, where F_o' is the "actual" output force and $F_o' < F_o$. Hence

$$\frac{W_o}{W_i} = \frac{F_o' d_o}{F_i d_i} = \frac{F_o'/F_i}{d_i/d_o} = \frac{F_o'/F_i}{F_o/F_i}$$

by virtue of the preceding equation. Since F_o'/F_i is M_a, we have

$$\text{Eff} = \frac{M_a}{M}$$

as was stated in Sec. 3.13.

8.11. Power. A small motor can do a great deal of work if it is given a long time; a large one may be able to do the same amount in a small time. It is said to be more powerful. Obviously, then, power refers to the speed with which work can be done or with which energy can be delivered.

Power is the time rate of doing work. In symbols,

$$P = \frac{dW}{dt} \tag{8.26}$$

If we recall Eq. (8.10), we find

$$P = F\frac{ds}{dt}\cos\theta = Fv\cos\theta \tag{8.27}$$

where θ is the angle between F and v. In accordance with Eq. (8.26) one can define an *average* power,

$$\bar{P} = \frac{W}{t_2 - t_1}$$

provided W is the work done between the instants t_1 and t_2. This is of course the same as P when P is constant. In general P may change in time, so that it becomes necessary to distinguish between average and instantaneous power output. The obvious units of power are collected in Table 8.2, which is self-explanatory. In addition to these the student

Table 8.2. *Units of Power*

	fps	cgs	mks
Absolute	ft poundal/sec	erg/sec	joule/sec = watt
Gravitational	ft lb/sec	gm cm/sec	kg m/sec

must learn the following which are widely used:

1. The joule/sec, which equals 10^7 ergs/sec (cf. Sec. 8.5), is called the *watt* (after James Watt, 1736–1819, who developed the steam engine). A kilowatt is 1,000 watts. While this is most often employed in electrical practice, it is a perfectly good unit for all forms of power.

2. The horsepower (hp) introduced by James Watt (who experimented with horses!) equals 550 ft lb/sec, or 33,000 ft lb/min. To find the conversion factor between horsepower and watts we write

$$\begin{aligned} 1\text{ hp} = 550\text{ ft lb/sec} &= 550(30.5\text{ cm})(454 \times 980\text{ dynes})/\text{sec} \\ &= 7.46 \times 10^9\text{ ergs/sec} \\ &= 746\text{ watts} \end{aligned}$$

The kilowatt (kw) is so common a unit that it has seemed desirable to introduce an *energy* unit based on it. Since, for uniform rate of working, energy delivered is power times time, the energy unit is the *kilowatt hour* (kwhr). It is the work done when one kilowatt of power is delivered for one hour.

$$1\text{ kwhr} = 3.6 \times 10^6\text{ joules}$$

Energy, or work done, is a commodity; power, being a time rate, has no price. Thus one does not pay for electric power—one pays for *energy* at a fixed price per kilowatt hour.

As a simple application of Eq. (8.27) we calculate the power developed by a locomotive that exerts a pull of 30,000 lb and moves at 20 miles/hr. Here $\cos\theta = 1$. Since 20 miles/hr = 29.3 ft/sec, we find

$$P = 30{,}000 \text{ lb} \times 29.3 \text{ ft/sec} = 8.79 \times 10^5 \text{ ft lb/sec} = 1{,}598 \text{ hp}$$

If it takes 25 hp to drive an automobile at 20 miles/hr on a level road, the force of friction F can be computed as follows:

$$F \times 29.3 \text{ ft/sec} = 25 \times 550 \text{ ft lb/sec}$$
$$F = (25 \times 550)/29.3 \text{ lb} = 469 \text{ lb}$$

8.12. Momentum and Impulse. In Sec. 8.7 an equation expressing conservation of mechanical energy was obtained from Newton's second law of motion; the method was to multiply both sides of the law

$$F = m\frac{dv}{dt}$$

by dx and then to integrate. Another interesting result emerges when both sides are multiplied by dt and the integration is carried out. We then have

$$\mathbf{F}\,dt = m\,d\mathbf{v}$$

and on integration between t_1 and t_2

$$\int_{t_1}^{t_2} \mathbf{F}\,dt = m\mathbf{v}_2 - m\mathbf{v}_1 \tag{8.28}$$

In writing these equations, account has been taken of the *vector character* of the quantities **F** and **v**. This is necessary because a vector such as **F**, when multiplied by a scalar like dt, yields a product that is still a vector, having the direction of **F**. Similarly the vector $m\mathbf{v}$ has the direction of **v**. Work, which is the product $F\,dx$ or, in general, the quantity $F_x dx + F_y dy$, is *not* a vector. Why this is so is shown in courses on vector analysis and will not be proved here.

Equation (8.28) takes on significance when we examine the meaning of the integral

$$\boxed{\mathbf{I} = \int_0^t \mathbf{F}\,dt} \tag{8.29}$$

If **F** were constant, it would be the force multiplied by the time during which it acts. The quantity **I** can be the same for a small force of long duration as for a large force applied briefly. It is a measure of the temporal effectiveness of the force and is called the *impulse* of the force **F**. In order to calculate **I** in detail one must know the way in which **F** depends on t.

Using Eq. (8.29) we can write Eq. (8.28) in another way.

$$\mathbf{I}\big|_{t_2} - \mathbf{I}\big|_{t_1} = m\mathbf{v}_2 - m\mathbf{v}_1 \tag{8.30}$$

The quantity $m\mathbf{v}$ appearing on the right-hand side is called *momentum*. The result may therefore be stated verbally by saying: Change in impulse equals change in momentum.

Usually $t_1 = 0$; that is, the force starts acting on the mass m at $t = 0$. The equation then states that the momentum change produced by a force is equal to the impulse of the force. For example, when a force of 1 poundal acts for 2 sec on a body of 10 lb mass initially at rest, it produces a speed v, which is given by

$$Ft = mv$$

or

$$1 \text{ poundal } 2 \text{ sec} = 10 \text{ lb} \times v$$

$$v = \tfrac{1}{5} \text{ (poundal sec)/lb} = \tfrac{1}{5} \text{ ft/sec.}$$

Notice that Eq. (8.30), being derived from the absolute form of Newton's law, *is valid only in absolute units.*

The units of impulse and of momentum, two quantities that have the same physical dimension $[MLT^{-1}]$, are of course the same. They have no special name but might be the gm cm/sec, or the lb ft/sec, or the kg m/sec. They must be carefully distinguished from units of power that have the same appearance because of the ambiguity of the words gram and pound.

8.13. Conservation of Momentum. By far the most interesting application of the impulse-momentum equation (8.30) is the simple one in which the quantity on the left-hand side of it is zero. This can happen in several ways.

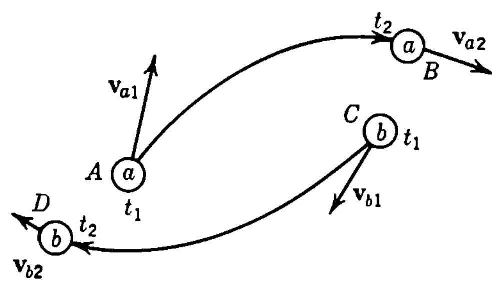


Fig. 8.11.

First let us suppose there is no force whatever. Then $\mathbf{I}$ is zero, and equality of $m\mathbf{v}_1$ and $m\mathbf{v}_2$ simply implies the validity of Newton's first law. Nothing new is to be learned from this instance; there is obviously conservation of momentum in the absence of force.

Much less trivial is the situation in which two bodies *interact*, *i.e.*, exert forces on each other. Call these bodies a and b, and consider their motion as it is indicated in Fig. 8.11. In its motion from A to B, particle a is at every instant subject to a force $\mathbf{F}_a$ due to body b, a force that may be elastic, electric, magnetic, or even due to bodily collision. Also, $\mathbf{F}_a$ need not be constant in time.

While body a moves from A at time t_1 to B at time t_2, body b moves from C to D. It is acted on by a force $\mathbf{F}_b$ due to body a, and, by Newton's third law,

$$\mathbf{F}_b = -\mathbf{F}_a \tag{8.31}$$

On applying Eq. (8.30) or, better, Eq. (8.28) to bodies a and b in turn, we find

$$\int_{t_1}^{t_2} \mathbf{F}_a dt = m_a \mathbf{v}_{a2} - m_a \mathbf{v}_{a1}$$
$$\int_{t_1}^{t_2} \mathbf{F}_b dt = m_b \mathbf{v}_{b2} - m_b \mathbf{v}_{b1}$$

If we add these two equations and use Eq. (8.31), there results

$$m_a \mathbf{v}_{a2} - m_a \mathbf{v}_{a1} + m_b \mathbf{v}_{b2} - m_b \mathbf{v}_{b1} = 0$$

or

$$m_a \mathbf{v}_{a2} + m_b \mathbf{v}_{b2} = m_a \mathbf{v}_{a1} + m_b \mathbf{v}_{b1} \tag{8.32}$$

The quantity on the left-hand side is the momentum of bodies a and b (vector sum) after, the quantity on the right-hand side the momentum before the interaction. When the two bodies are regarded as forming one physical system, the forces $\mathbf{F}_a$ and $\mathbf{F}_b$ are *internal* to this system in the sense that they are exerted by one part of the system upon another. One may thus state Eq. (8.32) in words by saying:

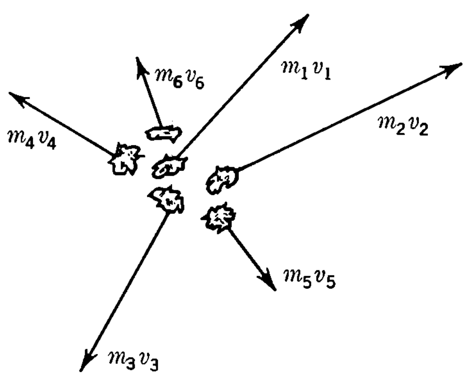


FIG. 8.12. Bursting bomb. $\Sigma m_i v_i = 0$.

The total (vector) momentum of a physical system that is subject only to internal forces remains constant.

If *external* forces, *i.e.*, forces due to bodies other than a and b, were also active, Eq. (8.32) would not be true, for the total force on a is not the negative of the total force on b. Actually, interacting bodies are always subject to gravity, which is an external force. But it often happens—for example, in collisions and explosions—that the forces of interaction are much greater than the force of gravity, so that application of the present result is warranted as an approximation. Later we shall show that external forces merely affect the motion of the center of mass (cf. Sec. 9.3) of the system and that the relative motion of a and b is given by Eq. (8.32) precisely.

***8.14. Explosions, Rockets.** The explosion of a stationary bomb is subject to the law of conservation of momentum. Since its momentum before bursting is zero, the vector momentum of all its fragments, when added, must still yield zero (see Fig. 8.12). Thus the $m\mathbf{v}$ vectors drawn in Fig. 8.12 must form a closed polygon. If the bomb broke into two fragments only, these would fly off in exactly opposite directions.

Suppose a shell, of mass M, traveling with velocity $\mathbf{V}$, bursts in mid air. The total momentum of its parts after the explosion still equals the vector $M\mathbf{V}$. If it breaks into two parts, one of mass $M/3$ and the other of mass $2M/3$, and if the smaller one flies off at an angle of 30° with the forward direction, the velocity of the second can be found by this rule. Let V = 100 ft/sec, and let the velocity of the smaller fragment be 300 ft/sec. We then see from Fig. 8.13 that, since

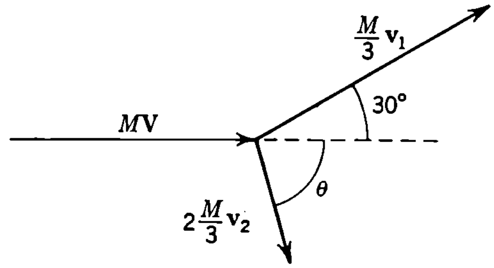


Fig. 8.13. Exploding shell.

$$[(M/3)\mathbf{v}_1] + [(2M/3)\mathbf{v}_2] = M\mathbf{V},$$

$$\frac{M}{3} v_1 \sin 30° = \frac{2}{3} M v_2 \sin \theta$$

$$\frac{M}{3} v_1 \cos 30° + \frac{2}{3} M v_2 \cos \theta = MV$$

The solution of these equations is

$$\tan \theta = \frac{v_1 \sin 30°}{3V - v_1 \cos 30°} = \frac{300 \text{ ft/sec} \times \frac{1}{2}}{300 \text{ ft/sec} - 300 \text{ ft/sec} \times 0.866}$$

$$\tan \theta = 3.73 \qquad \theta = 75°$$

$$v_2 = \frac{1}{2}\frac{v_1 \sin 30°}{\sin \theta} = \frac{1}{2}\frac{300 \text{ ft/sec} \times \frac{1}{2}}{0.966} = 77.7 \text{ ft/sec}$$

The principle of conservation of momentum has innumerable applications, the most interesting of which at present is probably to rocket propulsion. The

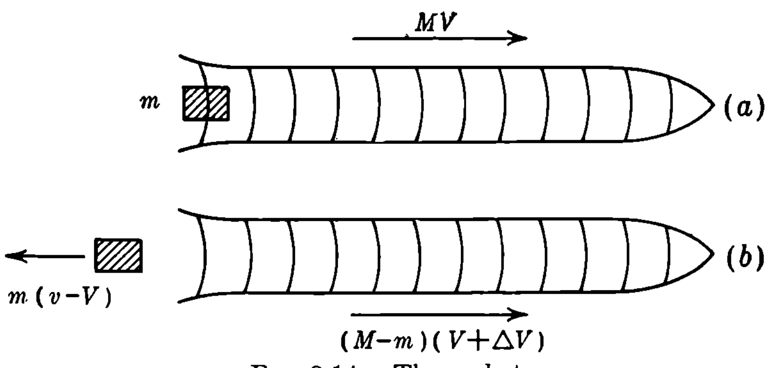


Fig. 8.14. The rocket.

rocket ejects from its tail a blast of gas and moves ahead, driven by the force of recoil from the blast. Contrary to widespread opinion, it needs no air to react against, for it kicks itself ahead from its own exhaust gases; the atmosphere is in fact an obstacle and reduces the efficiency of the rocket. As a rule, gases are expelled continuously to provide motive power; for simplicity let us first assume, however, that the expulsion is intermittent.

If the rocket of mass M, while moving to the right with velocity V, ejects from its tail a charge of mass m with velocity v relative to the rocket, the velocity of the rocket will increase by an amount ΔV. According to our principle (cf. Fig. 8.14),

$$(M - m)(V + \Delta V) - m(v - V) = MV$$

which reduces to

$$(M - m)\Delta V = mv \tag{8.33}$$

The increase in kinetic energy resulting from the explosion, ΔE, is

$$\tfrac{1}{2}m(v - V)^2 + \tfrac{1}{2}(M - m)(V + \Delta V)^2 - \tfrac{1}{2}MV^2 = \Delta E$$

But only part of this, *viz.*, the amount

$$\tfrac{1}{2}(M - m)(V + \Delta V)^2 - \tfrac{1}{2}MV^2 = \Delta E'$$

is the increase in kinetic energy of the rocket; the remainder $\tfrac{1}{2}m(v - V)^2$ is wasted as energy of the blast. The waste is zero when $V = v$, that is, when the exhaust mass is expelled with the velocity of the moving rocket. This corresponds to maximum efficiency and occurs when the exhaust gases are at rest relative to the earth.

In general the efficiency is given by $\Delta E'/\Delta E$. In calculating this ratio we simplify ΔE and $\Delta E'$ with the use of Eq. (8.33) and finally obtain

$$\frac{\Delta E'}{\Delta E} = 1 - \frac{(M - m)(v - V)^2}{Mv^2}$$

This shows that the efficiency of the rocket is m/M, and hence very small, when the rocket is stationary ($V = 0$). It is a maximum when $V = v$ as already noted. For higher speeds it decreases again, for then even the exhaust has a net velocity in the direction of the rocket's flight. This is an entirely possible condition, a rocket can go faster than its exhaust.

The same results hold for continuous propulsion. Jet propulsion differs from rocket action only inasmuch as air is taken in at the front of the missile and used in a combustion chamber to oxidize the fuel, which then produces the directed blast.

The most remarkable rocket produced thus far was the German V-2. It was propelled by the combustion products of alcohol. Since it passed through the stratosphere, it had to carry its own oxygen. As a propellant, it contained 1,200 gal of liquid oxygen and the same quantity of alcohol. Its explosive charge was sufficient to produce a crater 10 ft deep and 40 ft in diameter. Some data concerning the V-2 are given in Table 8.3.

In continuous propulsion, mass is sent forth at a uniform rate dm/dt and with constant velocity v. The thrust on the rocket is, by Newton's second law,

$$F = \frac{d}{dt}(mv) = \frac{dm}{dt}v$$

Table 8.3. *The V-2 Rocket*

Length	46 ft	Time of flight	6 min
Diam	5 ft 5 in.	Time of propulsion	70 sec
Total wt	27,000 lb	Max speed	1 mile/sec
Explosive	1,800 lb (TNT)	Terminal speed	½ mile/sec
Range	215 miles	Max power	500,000 hp

A given thrust can be obtained by varying dm/dt and v, but keeping their product constant. Now the *power* wasted on the exhaust products is

$$Fv = \frac{dm}{dt} v^2 = \frac{F^2}{dm/dt}$$

In order to keep this quantity small, dm/dt must be made large. This means that the rocket is wasteful either of power or of mass; it cannot be frugal with respect to both. In designing rockets a compromise must therefore be made. Indeed, it is usually desirable to save mass rather than energy.

***8.15. Ballistic Pendulum.** The speed of a rifle bullet is sometimes measured by a device called a "ballistic pendulum." It consists of a block of soft wood or lead (mass M) suspended by a light rod or by strings, as in Fig. 8.15. A rifle bullet of mass m and speed v is fired into the block, and the arc s over which the center of M swings out is measured.

When the bullet enters the block, the latter moves forth with a velocity V in accordance with the principle of conservation of momentum,

$$mv = (M + m)V \qquad (8.34)$$

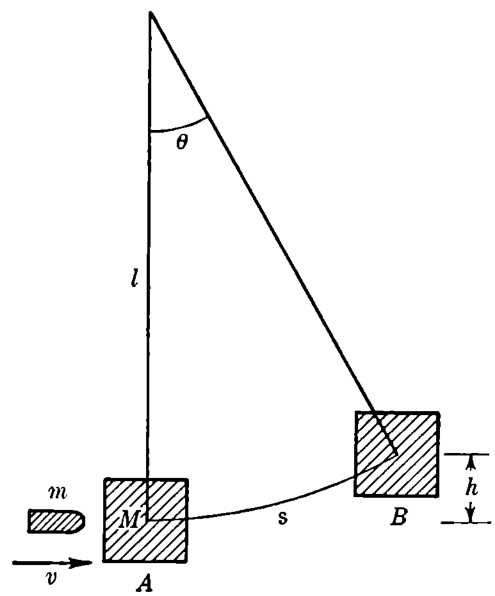


FIG. 8.15. Ballistic pendulum.

The principle of conservation of mechanical energy may *not* be applied to the *impact*, for the bullet will certainly generate heat and noise while lodging itself in the block. From Eq. (8.34) v could be determined if V were known. But V can be found by applying the principle of conservation of energy to the motion of $(M + m)$ from A, where the energy is all kinetic, to the end of its swing B, where it is all potential. Thus

$$\tfrac{1}{2}(M + m)V^2 = (M + m)gh$$

Now $h = l(1 - \cos \theta)$, so that

$$V = \sqrt{2gl(1 - \cos \theta)} = 2\sqrt{gl} \sin \frac{\theta}{2}$$

The angle $\theta = s/l$. Hence V is known, and v can be determined from Eq. (8.34) provided that m and M are given.

If we mistakenly assumed that kinetic energy is conserved before and after impact, we should have in addition to Eq. (8.34) the following:

$$\tfrac{1}{2}mv^2 = \tfrac{1}{2}(M + m)V^2$$

After dividing the left side by mv and the right-hand by its equal, $(M + m)V$, we have

$$v = V$$

When this is substituted back into Eq. (8.34), there results $m = M + m$, which is nonsense. If the bullet lodges itself in the block, mechanical energy *cannot* be conserved; most of the energy of the bullet is transformed into heat.

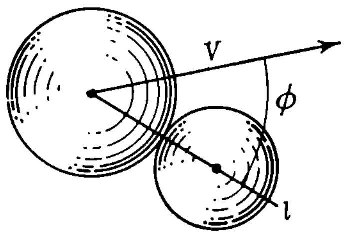


FIG. 8.16. Oblique impact.

8.16. Collisions. Conservation of momentum is the clue to all collision problems. These are usually difficult. To predict, for example, what happens in detail when two bricks collide in mid-air is a task that defies the best mathematician. Our attention will be confined to spheres; and even there we shall not treat the general case in which the line of centers, l, makes an arbitrary angle ϕ (cf. Fig. 8.16) with the velocity of the impinging sphere. Such an impact is called *oblique*. We shall here be interested in *head-on* collisions, for which $\phi = 0$. In that case there is no change in the direction of motion.

Figure 8.17 shows two balls of masses m_a and m_b with velocities $\mathbf{u}_a$ and $\mathbf{u}_b$ *before* impact. Since $\mathbf{u}_a > \mathbf{u}_b$, they will collide. After collision, their velocities are $\mathbf{v}_a$ and $\mathbf{v}_b$. We wish to find $\mathbf{v}_a$ and $\mathbf{v}_b$ when the velocities before impact and the masses are given.

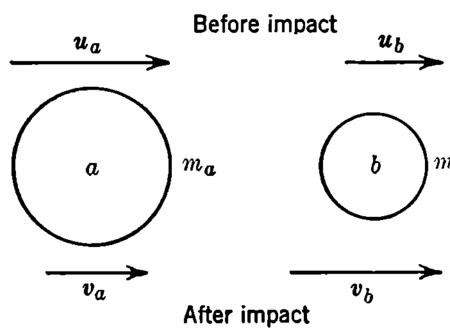


FIG. 8.17. Head-on collision of spheres.

Since momentum of both spheres before impact equals momentum after impact,

$$m_a u_a + m_b u_b = m_a v_a + m_b v_b \tag{8.35}$$

This is always true. *Kinetic energy* is conserved only if the impact is *perfectly elastic*, as is nearly true for billiard balls and for steel spheres. We assume this to be the case. Then

$$\tfrac{1}{2}m_a u_a^2 + \tfrac{1}{2}m_b u_b^2 = \tfrac{1}{2}m_a v_a^2 + \tfrac{1}{2}m_b v_b^2 \tag{8.36}$$

If we write Eqs. (8.35) and (8.36) in the form

$$m_a(u_a - v_a) = m_b(v_b - u_b)$$
$$m_a(u_a^2 - v_a^2) = m_b(v_b^2 - u_b^2)$$

and divide, we obtain

$$u_a + v_a = v_b + u_b$$

or

$$u_a - u_b = v_b - v_a \tag{8.37}$$

This means (cf. Fig. 8.17) that the *relative velocity of separation equals the relative velocity of approach.* It is now easy to see why kinetic energy cannot be conserved in the case of the ballistic pendulum, where the bullet remains lodged in the block and hence the speed of separation is zero.

When $m_a = m_b$, Eq. (8.35) reads

$$u_a + u_b = v_a + v_b$$

When this is solved simultaneously with Eq. (8.37), we find

$$v_b = u_a \qquad v_a = u_b$$

The masses have merely exchanged their velocities on impact. This result is familiar to the billiard player, who knows that, when a moving ball makes a head-on collision with a stationary one, the former stops.

***8.17. Worked Example.** A 5-gm rifle bullet is fired into the block of a ballistic pendulum, which weighs 10 kg. The block is suspended from cords 80 cm long and is observed to move along an arc of 5 cm.

a. Find the velocity of the bullet.

Using the notation of Sec. 8.15, we find

$$\theta = \frac{5}{80} = \frac{1}{16} \qquad \sin\frac{\theta}{2} = \frac{\theta}{2}\text{ (very nearly)} = \frac{1}{32}$$

Hence $$V = 2\sqrt{gl}\sin\frac{\theta}{2} = 2\sqrt{980 \times 80}\,\frac{1}{32}\text{ cm/sec} = 17.5\text{ cm/sec}$$

Now $$v = \frac{M + m}{m}V\text{ [cf. Eq. (8.34)]} = 2{,}001V = 3.5 \times 10^4\text{ cm/sec}$$

b. Find the work done by the bullet in penetrating the block.

The energy of the bullet was

$$\tfrac{1}{2} \times 5 \times (3.5 \times 10^4)^2\text{ ergs} = 3.08 \times 10^9\text{ ergs}$$

The energy of block and bullet after impact is

$$\tfrac{1}{2} \times 10{,}005 \times (17.5)^2\text{ ergs} = 1.53 \times 10^6\text{ ergs}$$

The difference is the work done in penetrating the block. It is nearly all of the initial energy.

PROBLEMS

1. A 500-lb horse draws a sleigh weighing 1,200 lb along a level, snow-covered road with an acceleration of 3 ft/sec². The coefficient of sliding friction is 0.1. Ignoring vertical force components, find the following: (*a*) the force exerted by the horse on the sleigh; (*b*) the force exerted by the sleigh on the horse; (*c*) the forces exerted by the earth on the sleigh and on the horse.

2. A mass, which may only move along the X axis, is subject to a force $F = -kx^3$ dynes. Calculate the work that must be done on the mass in order to move it from $x = -2$ cm to $x = +1$ cm without acceleration.

★3. A mass in the XY plane is subject to component forces $F_x = -4x$, $F_y = -7y$ dynes. Calculate the work required to move the mass (without acceleration) from the origin to the point (3,3) along the line $y = x$. Show that the same amount of work must be expended in moving the mass along the X axis to the point (3,0) and then parallel to the Y axis to the point (3,3). (Let 1 unit length = 1 cm.)

4. A 30-ton railway car traveling with a speed of 5 ft/sec hits a spring bumper at the end of the track. Calculate the kinetic energy of the car before collision and the distance that it compresses the spring. ($k = 4 \times 10^5$ poundals/ft.)

5. A pile driver raises a mass of iron (called the "monkey") a given height and then allows it to drop upon the pile, driving it into the ground. If the mass of the monkey is 1 ton and the height 20 ft, find the kinetic energy of the monkey at the instant it strikes the pile. Assuming the pile resists with a constant force of 8 tons, calculate the distance it is forced into the ground.

6. A 2-ton automobile requires 15 hp to maintain a speed of 30 miles/hr on a level road. What horsepower must it develop to travel up a 5° grade with the same speed?

7. A 150-lb man climbs a circular stairway to a height of 20 ft in 20 sec. How much energy does he expend? What horsepower does he develop?

8. An airplane engine developing 65 hp makes 2,000 rpm. What torque is it delivering?

9. An airplane in level flight at 240 miles/hr develops 1,500 hp. Determine the force that the air exerts in opposition to the forward motion of the airplane. (Assume that the propeller is 50 per cent efficient.)

10. When an elastic spherical ball strikes a horizontal surface at an angle with the vertical, the horizontal component of its momentum remains unchanged (provided that the surface is frictionless). Show that under this condition the angle of reflection is equal to the angle of incidence.

★11. The pulley system in Fig. 8.18 has an efficiency of 0.8. How much work must be put in to raise the mass $M = 20$ lb a height of 5 ft? What is the magnitude of the input force? What is the theoretical mechanical advantage; the actual mechanical advantage?

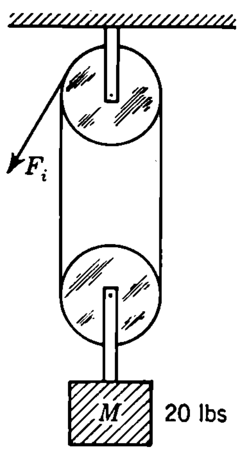


FIG. 8.18. Problem 11.

★12. An automobile jack, operated by a handle 2 ft long, utilizes a screw of pitch ½ in. When a 20-lb force is applied at the handle, the output force is 2 tons. Calculate (*a*) the theoretical mechanical advantage; (*b*) the actual mechanical advantage; (*c*) the efficiency of the machine; (*d*) the input force required to raise a 450-lb mass 3 in.

★13. A man operating a windlass raises a 30-lb mass from the bottom of a 50-ft well in 80 sec. If the windlass is 90 per cent efficient, how much energy does the man expend? What horsepower does he develop?

14. A force that is constant in direction but variable in time is given by the relation $F = 10^3 t^3$ dynes. If this force acts on a mass of 35 gm from $t_1 = 0.3$ sec to $t_2 = 0.7$ sec, calculate the change in momentum of the mass.

15. A pitched baseball weighing 0.5 lb is hit as it crosses the plate horizontally with a speed of 60 mph. It flies off horizontally at 90 mph. Calculate the change in momentum of the ball. Find the force exerted by the bat on the ball, assuming that it is constant and that the bat and ball are in contact for 0.1 sec.

16. A 100-ton cannon fires a 100-lb shell, imparting to it a velocity of 3,000 ft/sec. (*a*) Find the recoil velocity of cannon (assuming that it is completely free to recoil). (*b*) Find the work done on the shell by the expanding gases. (*c*) Find the work done on the cannon by the expanding gases. (*d*) Calculate the efficiency of the cannon, assuming that no energy is lost in heat.

17. Two cars of equal mass approach each other at right angles. The first is going at 40 mph, the second at 20 mph. They lock together during collision and travel off at a common speed. Find their velocity (magnitude and direction) after collision.

18. A 10-ton railway car traveling along a level track at 15 mph strikes a stationary 10-ton car. The two couple together and proceed up a 1° incline in the track. How far do they rise on the incline? (Neglect friction.)

19. A simple pendulum 10 ft long is displaced 30° from the vertical and then released. Determine the velocity of the bob at the instant it passes through the lowest point in its path. (Neglect any friction due to suspension or air resistance.)

***20.** A ballistic pendulum, 6 ft in length and using a 10-lb block, swings through an angle of 30° when a 0.05-lb bullet is fired into the block. Calculate the velocity of the bullet.

***21.** Suppose a pendulum of length l is suspended along a flat vertical surface (see Fig. 8.19). When it swings to the left, it utilizes its full length; but when it swings to the right, its effective length is $l/2$. If $\theta = 40°$ is the maximum angle of swing to the left, find ϕ, the maximum angle of swing to the right.

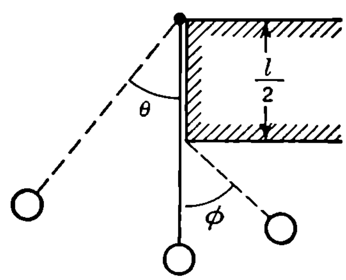


FIG. 8.19. Problem 21.

***22.** Show that only $m/(M + m)$ times the initial energy of the bullet is converted into kinetic energy of the block of a ballistic pendulum.

23. The spring in a BB gun has a stiffness $k = 2 \times 10^6$ dynes/cm. Calculate the work that must be done to compress the spring 20 cm from its neutral position. When the trigger is pulled, the spring is allowed to expand, propelling a BB weighing 0.25 gm. Find the kinetic energy and velocity of the BB as it leaves the gun.

CHAPTER 9

DYNAMICS OF A RIGID BODY

9.1. Motion of a System of Two Particles. Newton's laws as stated hold only for small particles. Nevertheless we have often applied them to extended objects, apparently without justification. In the present chapter we supply the reasons why such treatment was proper; we investigate the motion of objects called *rigid bodies.*

A stick when thrown in the air does not behave like a small stone, for it will in general execute an apparently irregular rotation while

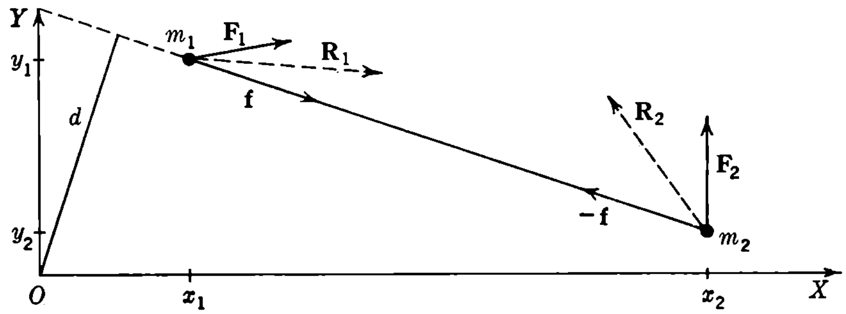


FIG. 9.1.

progressing. One may at first think that its motion is so complicated as to defy description. In fact, however, its features are astonishingly simple. There is one point in the stick—indeed in every rigid body—that moves precisely as does the stone, tracing out a smooth parabola while the whole body rotates about this key point. It is called the *center of mass.* To recognize its importance and to locate it is the first object of our quest.

A rigid body may be considered as a collection of particles, all held in fixed positions relative to one another. This view is in harmony with modern conceptions as to the structure of matter, according to which a solid consists of atoms or molecules arranged at definite places in a crystal "lattice." To simplify the analysis we consider first only two such particles, held together by forces like those in a rigid body, and we inquire how the pair moves when forces are applied to each of the partners.

In Fig. 9.1 are drawn two small masses m_1 and m_2, together with the origin O of a coordinate system. Each mass is assumed to be subject to forces. The forces are of two kinds: there are *internal* forces exerted by one particle on the other, and there are *external* forces due to outside agencies (gravitation, pushes, and pulls). In a rigid body the internal forces between molecules are of an electrical origin; they are *central* forces, *i.e.*, they act along the line joining the particles. Being subject to Newton's third law, they are denoted by $\mathbf{f}$ and $-\mathbf{f}$. In Fig. 9.1 the external force on m_1 is $\mathbf{F}_1$, that on m_2 is $\mathbf{F}_2$. The total force on m_1 is $\mathbf{R}_1$; that on m_2 is $\mathbf{R}_2$.

By Newton's second law,

$$m_1\mathbf{a}_1 = \mathbf{R}_1, \qquad m_2\mathbf{a}_2 = \mathbf{R}_2 \tag{9.1}$$

But $\qquad \mathbf{R}_1 = \mathbf{F}_1 + \mathbf{f}, \qquad$ and $\qquad \mathbf{R}_2 = \mathbf{F}_2 - \mathbf{f}$

and hence if we add the first two equations we find

$$m_1\mathbf{a}_1 + m_2\mathbf{a}_2 = \mathbf{F}_1 + \mathbf{F}_2 \tag{9.2}$$

This result is interesting, for the forces appearing here are external and measurable; the internal forces, which are unknown, fortunately have dropped out. The sum $\mathbf{F}_1 + \mathbf{F}_2$ is simply the total external force on the pair and will be called $\mathbf{F}$. Equation (9.2), being a vector equation, is equivalent to three scalar equations,

$$\begin{aligned} m_1 \frac{d^2x_1}{dt^2} + m_2 \frac{d^2x_2}{dt^2} &= F_x \\ m_1 \frac{d^2y_1}{dt^2} + m_2 \frac{d^2y_2}{dt^2} &= F_y \\ m_1 \frac{d^2z_1}{dt^2} + m_2 \frac{d^2z_2}{dt^2} &= F_z \end{aligned}$$

These can be simplified by introducing the quantities

$$\left.\begin{aligned} \bar{x} &= \frac{m_1x_1 + m_2x_2}{M} \\ \bar{y} &= \frac{m_1y_1 + m_2y_2}{M} \\ \bar{z} &= \frac{m_1z_1 + m_2z_2}{M} \\ M &= m_1 + m_2 \end{aligned}\right\} \tag{9.3}$$

for then they read

$$\left.\begin{aligned} M\frac{d^2\bar{x}}{dt^2} &= F_x \\ M\frac{d^2\bar{y}}{dt^2} &= F_y \\ M\frac{d^2\bar{z}}{dt^2} &= F_z \end{aligned}\right\} \tag{9.4}$$

Inspection of Eqs. (9.3) shows that $\bar{x}$ is the *mass-weighted mean* of x_1 and x_2, $\bar{y}$ is the mass-weighted mean of y_1 and y_2, and so forth. To understand this terminology we refer to a more familiar situation. Suppose we are given two boxes of nails and are to determine the *mean* length of the nails. For this purpose we might measure the length of one nail from *each* box (all nails in a box have the same length) and take the average: $l = (l_1 + l_2)/2$. But this is hardly the thing to do if the two boxes contain different numbers of nails. The proper procedure then is to *weight* the length l_1 by the number of nails in the first box and l_2 by the number of nails in the second box, thus obtaining

$$l = (n_1 l_1 + n_2 l_2)/(n_1 + n_2)$$

This is called the "weighted mean," the weights n_1 and n_2 being the numbers of nails of each kind present. In the definitions of $\bar{x}$, $\bar{y}$, and $\bar{z}$ the *masses* of the particles are used as weights. The three coordinates $\bar{x}$, $\bar{y}$, $\bar{z}$ define a point called the center of mass of m_1 and m_2. The vector displacement of this point from O will be called $\bar{\mathbf{r}}$; it has components $\bar{x}$, $\bar{y}$, $\bar{z}$. Clearly Eqs. (9.4) can then be combined into the one vector equation

$$M \frac{d^2\bar{\mathbf{r}}}{dt^2} = \mathbf{F} \tag{9.5}$$

This result will be interpreted in Sec. 9.3.

9.2. Motion of a System of Two Particles (Continued). Next we consider the torques produced by $\mathbf{R}_1$ and $\mathbf{R}_2$ about the origin O (cf Fig. 9.1). Again we resolve the forces into external ($\mathbf{F}_1$ and $\mathbf{F}_2$) and internal ($\mathbf{f}$ and $-\mathbf{f}$) components. And now we find that the torques due to $\mathbf{f}$ and $-\mathbf{f}$ cancel each other.

To see this we examine Fig. 9.1. The torque produced by $\mathbf{f}$ is $-fd$; that produced by $-\mathbf{f}$ is $+fd$. *The torques due to internal forces about any axis add up to zero.* This result is true because the internal forces act along the line joining the particles; *i.e.*, they are "central" forces.

Thus in computing the total torque on our pair we need not include the internal force at all, and we find, using Eq. (3.5),

$$L = x_1F_{1y} - y_1F_{1x} + x_2F_{2y} - y_2F_{2x}$$

Now put

$$x_1 = \bar{x} + \xi_1 \qquad x_2 = \bar{x} + \xi_2$$
$$y_1 = \bar{y} + \eta_1 \qquad y_2 = \bar{y} + \eta_2$$

so that ξ_1 and η_1 are the coordinates of m_1 *with respect to the center of mass* of the pair, and so forth. Then

$$\begin{aligned} L &= \bar{x}(F_{1y} + F_{2y}) - \bar{y}(F_{1x} + F_{2x}) + \xi_1 F_{1y} - \eta_1 F_{1x} + \xi_2 F_{2y} - \eta_2 F_{2x} \\ &= \bar{x}F_y - \bar{y}F_x + \xi_1 F_{1y} - \eta_1 F_{1x} + \xi_2 F_{2y} - \eta_2 F_{2x} \end{aligned} \tag{9.6}$$

since $\mathbf{F}_1 + \mathbf{F}_2 = \mathbf{F}$, the total external force.

We now hasten to put some physical meaning into these mathematical results.

9.3. Significance of the Center of Mass. Equation (9.5) looks very much like Newton's second law. It differs in three respects: (1) M is not the mass of a single particle but the mass of the pair. (2) The acceleration $d^2\bar{\mathbf{r}}/dt^2$ is not the acceleration of either mass but the acceleration of an abstract point, the center of mass. (3) The force $\mathbf{F}$ is not the force on any one particle but the *total external force* on the pair. Expressed in words, Eq. (9.5) says:

The center of mass C of a pair of particles moves under the action of external forces as if all the mass were concentrated at C and as if all forces were applied at C.

The center of mass of a pair of small spheres stuck on the ends of a light stick may be found very simply (Fig. 9.2). For convenience we can take the origin at the position of mass m_1 and the x coordinate along the stick. We then find

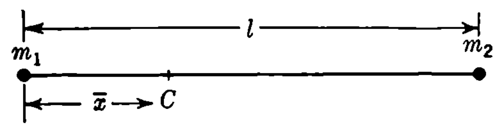


FIG. 9.2. Center of mass.

$$\bar{x} = \frac{m_1 \times 0 + m_2 l}{m_1 + m_2} = \frac{m_2}{m_1 + m_2} l$$

The student should show that the same point C is located even when a different origin is chosen.

We have proved that, if this stick were thrown, C would move on a parabolic path; this point would behave in all respects like a stone of mass $m_1 + m_2$. Whatever else may take place is contained in Eq. (9.6). It says that the torque on the pair consists of two parts:

(1) $$\bar{x}F_y - \bar{y}F_x$$

the torque about O that would result if the entire external force were applied at the center of mass, and

(2) $$(\xi_1 F_{1y} - \eta_1 F_{1x}) + (\xi_2 F_{2y} - \eta_2 F_{2x})$$

the torque of the actual forces $\mathbf{F}_1$ and $\mathbf{F}_2$ about the center of mass. Hence the center of mass moves as though it were subject to the whole external force, both with respect to translation and rotation; in addition to this the pair will rotate about C if the actual external forces produce a torque about C.

If the external forces are applied at the center of mass, there will be no rotation. For example, if the force of the hand is applied to the stick of Fig. 9.2 at C, it will not rotate while being thrown. When it has left the thrower's hand, the force of gravity will act upon it at the *center of gravity*. But this is coincident with the center of mass, as will be shown in the next section. Hence the stick, thus projected, will not rotate in flight.

All these results are perfectly general. They hold for a rigid body containing not two but a very large number of particles. But we shall not prove them again. We shall merely repeat that part of our analysis which led to Eqs. (9.4), assuming that the number of particles is great.

9.4. Center of Mass in General.[1] Of all the particles making up the rigid body that have masses m_1, m_2, m_3 . . . we select for consideration a typical one, say the ith (cf. Fig. 9.3). Its mass is m_i, and its momentary position is indicated by the displacement vector $\mathbf{r}_i$ drawn from the origin, with components x_i, y_i, and z_i.

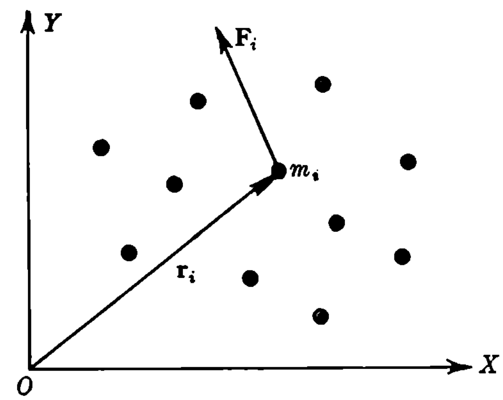


FIG. 9.3.

This particle is subjected to an *external* force $\mathbf{F}_i$ and also to a great number of internal forces, which may be represented by arrows (not drawn in Fig. 9.3) pointing along the lines from m_i to all the other particles. Let the *resultant* of all these internal forces acting on m_i be $\mathbf{f}_i$. The motion of this particular mass is then described by

$$\mathbf{F}_i + \mathbf{f}_i = m_i \frac{d^2\mathbf{r}_i}{dt^2}$$

A similar analysis can be made for every other particle, and a similar equation results. Let us add all these equations. We obtain

$$\sum_i \mathbf{F}_i + \sum_i \mathbf{f}_i = \sum_i m_i \frac{d^2\mathbf{r}_i}{dt^2}$$

But by Newton's third law

$$\sum_i \mathbf{f}_i = 0$$

since all internal forces occur as equal and opposite members of pairs.

[1] Review Secs. 2.10 and 3.7.

Furthermore, $\sum_i \mathbf{F}_i$ is the total external force on the rigid body, regardless of where it is applied; we shall call it $\mathbf{F}$. Hence

$$\mathbf{F} = \sum_i m_i \frac{d^2\mathbf{r}_i}{dt^2} \tag{9.7}$$

If we wish to write this equation in the same form as Eq. (9.5), we must take for the position vector of the center of mass the following:

$$\bar{\mathbf{r}} = \frac{1}{M} \sum_i m_i \mathbf{r}_i, \qquad \text{where } M = \sum_i m_i$$

or, in component form,

$$\left.\begin{aligned} \bar{x} &= \frac{1}{M} \sum_i m_i x_i \\ \bar{y} &= \frac{1}{M} \sum_i m_i y_i \\ \bar{z} &= \frac{1}{M} \sum_i m_i z_i \end{aligned}\right\} \tag{9.8}$$

Equations (9.3) represent a special case of this. The center of mass is still the *mass-weighted mean* of all the particle displacements, but in a more general sense. The law of motion of the rigid body becomes

$$\boxed{\mathbf{F} = M \frac{d^2\bar{\mathbf{r}}}{dt^2}} \tag{9.9}$$

The center of mass of the body moves as if all its mass were concentrated there and all forces applied there.

In Sec. 3.7 we introduced a point called *center of gravity* [Eq. (3.8)] by means of a set of equations analogous to Eqs. (9.8). They involved the *weights* of the particles instead of their *masses*. If $w_i = mg$, as we know to be the case near the earth's surface, the factor g cancels from Eq. (3.8) and the *center of gravity becomes identical with the center of mass*. However, one should recognize that this is not invariably true. For very large objects such as mountains and air masses, g is not the same at every particle, the cancellation of g does not take place, and the two "centers" are not coincident. Clearly the center of mass is the more fundamental concept. It is always a *fixed* point within a rigid body.

9.5. Center of Mass of a Continuous Mass Distribution; Examples. In practice it is impossible to carry out the summations over all particles that appear in Eqs. (9.8) for a body of ordinary size. Rigid bodies may, however, be thought of as having a *continuous* distribution of mass, and the summations then become integrations.

In Fig. 9.4 we have depicted a rod, along which the *mass per unit length* may change as we pass from one end to the other. Let the mass per unit length be λ gm/cm; in general λ will be a function of x. We wish to calculate

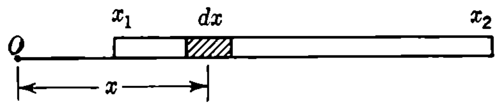


FIG. 9.4. Calculating the center of mass of a rod.

$$\bar{x} = \frac{1}{M} \sum_i m_i x_i$$

For this purpose we divide the rod into small segments, one of which, of length dx, is shaded in the figure. The mass of this segment, m_i, equals $\lambda\, dx$; its distance from O, formerly called x_i, is x. Thus $m_i x_i$ becomes $\lambda\, dx\, x$, and

$$\sum_i m_i x_i = \int_{x_1}^{x_2} \lambda x\, dx$$

In a similar way,

$$M = \int_{x_1}^{x_2} \lambda\, dx,$$

so that

$$\bar{x} = \int_{x_1}^{x_2} \lambda x\, dx \Big/ \int_{x_1}^{x_2} \lambda\, dx$$

To carry out these integrations one must know λ as a function of x.

a. The simplest case is that of a uniform rod of constant density, for which λ = const. Then

$$\bar{x} = \frac{\lambda[x^2/2]_{x_1}^{x_2}}{\lambda[x]_{x_1}^{x_2}} = \frac{1}{2}\frac{x_2^2 - x_1^2}{x_2 - x_1} = \frac{x_2 + x_1}{2}$$

The center of mass lies at the center of the rod.

b. Let the rod be composed of two different substances, so that $\lambda = \lambda_1$, a constant, between x_1 and x', $\lambda = \lambda_2$, another constant, between x' and x_2. Then

$$\bar{x} = \frac{\lambda_1 \int_{x_1}^{x'} x\, dx + \lambda_2 \int_{x'}^{x_2} x\, dx}{\lambda_1 \int_{x_1}^{x'} dx + \lambda_2 \int_{x'}^{x_2} dx}$$
$$= \frac{1}{2}\frac{\lambda_1(x'^2 - x_1^2) + \lambda_2(x_2^2 - x'^2)}{\lambda_1(x' - x_1) + \lambda_2(x_2 - x')}$$

If we notice that the masses of the two parts of the rod are $M_1 = \lambda_1(x' - x_1)$ and $M_2 = \lambda_2(x_2 - x')$, this result can be written

$$\frac{M_1[(x_1 + x')/2] + M_2[(x' + x_2)/2]}{M_1 + M_2}$$

This means that the center of mass of the composite object can be computed by putting all the mass of one part (M_1) at its center of mass $(x_1 + x')/2$, all the mass of the second part (M_2) at its center of mass $(x' + x_2)/2$, and then finding the center of mass of these two "particles."

c. To find $\bar{x}$ for a plane lamina in the shape of an isosceles triangle (cf. Fig. 9.5) we divide it into vertical strips of width dx. The length of each strip is $2x \tan \alpha$, and its mass

$$2x \tan \alpha \, dx \, \mu$$

if μ is the mass of the lamina per square centimeter. This will be assumed constant. For convenience we choose the origin at the apex.

Then
$$\bar{x} = \frac{2\int_0^h \mu \tan \alpha \, x^2 dx}{2\int_0^h \mu \tan \alpha \, x \, dx} = \frac{2}{3} h$$

By a similar method $\bar{y}$ could be computed. This, however, is unnecessary. For it is easily seen that every vertical strip has $\bar{y}$ at its center, and we conclude "from symmetry" that $\bar{y} = 0$.

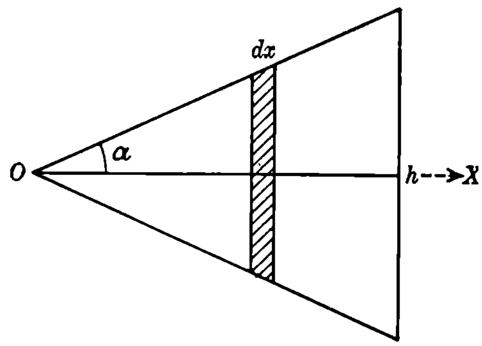


FIG. 9.5. Finding the center of mass of an isosceles triangle.

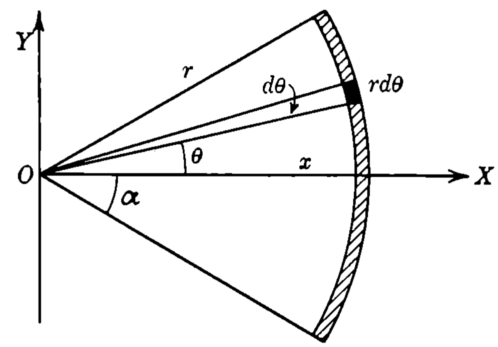


FIG. 9.6. Finding the center of mass of a circular arc of wire.

d. The center of mass of a circular arc (Fig. 9.6) is best calculated in polar coordinates. Here again, $\bar{y} = 0$;

$$\bar{x} = \frac{\int_{-\alpha}^{\alpha} r \cos \theta \, r \, d\theta \, \lambda}{\int_{-\alpha}^{\alpha} r \, d\theta \, \lambda} = r \frac{\sin \alpha}{\alpha}$$

As before, the mass per unit length of the arc, λ, cancels from the final result.

***9.6. Center of Mass of Three-dimensional Bodies.** To determine C for an irregular three-dimensional object by integration is often a hopelessly difficult problem. It would involve a threefold integration between complicated limits and is rarely carried out. The best procedure in such instances is to regard the body as decomposed into regular solids for which C is known and then to compound these parts in accordance with the rule that the composite center of mass results from treating the individual masses as particles, placed at their proper centers of mass.

For bodies of uniform composition the center of mass is identical with the geometric center, called the *centroid*.

To show how C can be determined for a regular body, we treat the case of a hemisphere of density (mass per unit volume) ρ gm/cm^3, reference being made to Fig. 9.7. If the radius of the hemisphere is a, the circular slab of thickness dx, shaded in the figure, has an area πy^2; and since

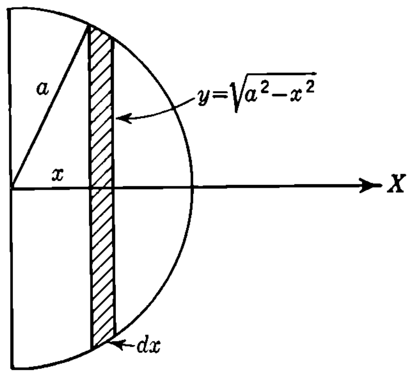


FIG. 9.7. Center of mass of a hemisphere.

$$x^2 + y^2 = a^2$$

this is $\pi(a^2 - x^2)$. Hence

$$\bar{x} = \frac{\int_0^a x\pi(a^2 - x^2)dx\,\rho}{M} = \frac{\pi\rho a^4}{4M}$$

But we know that M, the mass of the hemisphere, is $\frac{1}{2}\rho \times \frac{4}{3}\pi a^3$. The result is therefore

$$\bar{x} = \tfrac{3}{8}a$$

The other two coordinates of C, $\bar{y}$ and $\bar{z}$, are zero.

Finally it is well to establish the fact that C is always a *fixed point in the body*, that it does not change when a different origin of coordinates is chosen. Assume that $\bar{x}$ has been determined with respect to a given origin O. Now take an origin x_0 units of length to the left of O, on the negative X axis. We call the coordinates with respect to this new origin x'. Then

$$\bar{x}' = \frac{\sum_i m_i x_i'}{\sum_i m_i} = \frac{\sum_i m_i(x_i + x_0)}{\sum_i m_i} = \frac{\sum_i m_i x_i}{\sum_i m_i} + \frac{x_0 \sum_s m_i}{\sum_i m_i} = \bar{x} + x_0$$

The new center of mass has a coordinate x_0 units of length greater than the old. But since the origin has been shifted by an equal amount in the opposite direction, this specifies the same point in the body. The same holds for $\bar{y}$ and $\bar{z}$.

9.7. Rotation of a Rigid Body about a Fixed Axis. An account has been given of the general principles that govern the motion of bodies. To apply them in detail to actual problems, as, for example, the behavior of the earth under all forces exerted upon it by the sun, the moon, and the other planets, is an extremely difficult matter. One of the main difficulties arises from the fact that the earth does not spin about a fixed axis in space. In many interesting motions, however, this complication does not exist. The revolution of a flywheel, for example, occurs about an invariable axis. The rolling of a cylinder down a plane involves rotation about an axis, which, while moving in space, does not change its direction. These cases can be treated more simply and will now be considered.

The clue to their treatment is a law very similar to Newton's second law; one may in fact be derived from the other. In Fig. 9.8 we have drawn in outline a rigid body, which we assume free to rotate about a fixed axis through O. The axis extends at right angles to the plane of the paper, so that each particle of the body moves in a circle in a plane parallel to that of the figure. Unspecified forces attack all particles of the body, and we wish to find, ultimately, how the body moves under their action.

We select for consideration the particle whose mass is m_i. Let $\mathbf{F}_i$ be the force acting upon it. This may be resolved into two components, F_{i1} acting outward along r_i and F_{i2} acting at right angles to r_i. The first of these produces no torque about O; hence the torque about O to which m_i is subject will be

$$L_i = r_i F_{i2} \tag{9.10}$$

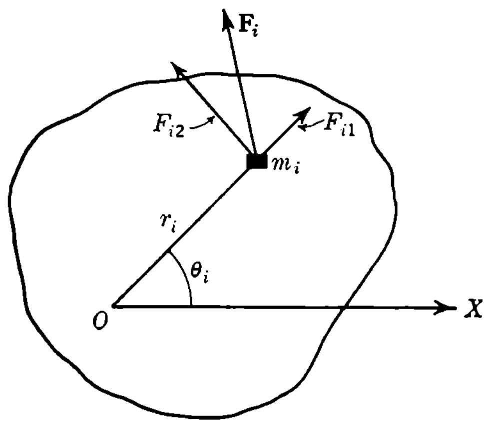


Fig. 9.8.

Now, by Newton's second law, $F_{i2} = m_i a_i$, if a_i is the component of the acceleration of the particle along the direction of F_{i2}. But the whole body has an angular acceleration, whose magnitude we will take to be α, about this axis; therefore

$$a_i = r_i \alpha$$

and

$$F_{i2} = m_i r_i \alpha \tag{9.11}$$

On putting this in Eq. (9.10) we have

$$L_i = m_i r_i^2 \alpha$$

To obtain the torque L acting on the whole body this expression must be summed over all particles, and this leads to the final result

$$L = \left(\sum_i m_i r_i^2\right)\alpha \tag{9.12}$$

an equation very similar to $F = ma$. Like the latter, Eq. (9.12) holds only when L is expressed in absolute units (centimeter dynes, or foot poundals, or meter newtons).

9.8. Moment of Inertia. That L is the rotational counterpart of F has already been noted. Also α, the angular acceleration, is the analogue of a. If m is the inertia of a body relative to motion under forces, Eq. (9.12) says that the quantity $\sum_i m_i r_i^2$ is the inertia of a body relative to rotation under torques. It is called the *moment of inertia* of the body and denoted by the symbol I. The fundamental law for rotation about a fixed axis therefore reads

$$\boxed{L = I\alpha} \tag{9.13}$$

It is important that we familiarize ourselves with the precise meaning of I. The quantity r_i appearing in the summation is the distance of the ith particle from the axis of rotation, $r_i^2 = x_i^2 + y_i^2$; it is not its distance from the origin, which would be $(x_i^2 + y_i^2 + z_i^2)^{1/2}$. Since the mass-weighted mean of the squares of all particle distances from the axis of rotation is $\sum_i m_i r_i^2 / \sum_i m_i$, the moment of inertia is this mean multiplied by the total mass of the body. Its numerical value for a given body depends on the axis chosen. For instance, I for a sphere is different with respect to an axis through its center from what it is with respect to an axis passing through an off-center point. The quantity $\sqrt{I/M}$, which has the physical dimension of a length, is called the *radius of gyration*. It represents the distance from the axis of rotation at which a single mass M would produce the same moment of inertia as the body.

The simplest application of Eq. (9.13) is to a single particle moving about an axis as a stone revolving about the finger at the end of a string. Its moment of inertia is simply mr^2, and Eq. (9.13) says

$$L = mr^2\alpha$$

But here $L = Fr$ if F is the force applied tangentially to the stone, and $\alpha = a/r$. Hence $Fr = mra$, which is nothing new.

A body that has all its mass disposed at equal distances from its axis of rotation—as, for example, a bicycle wheel (we neglect the mass of the spokes)—has an I that equals its total mass times the square of its radius, for

$$I = \sum_i m_i r_i^2 = r^2 \sum_i m_i = Mr^2$$

***9.9. Calculation of Moments of Inertia.** The moment of inertia of a continuous distribution of mass is clearly

$$I = \int r^2 dm$$

For some regular bodies this can be easily computed. The trick is to express m and r in terms of the same variable.

a. Uniform Rod (Fig. 9.9). If the mass per centimeter is λ, $dm = \lambda\, dx$, $r = x$, and

$$I = \lambda \int_0^l x^2 dx$$

provided that the axis is chosen through the end point. Hence $I = \lambda l^3/3$. Since the mass of the rod is $\lambda l = M$, this can be written

$$I = \tfrac{1}{3} M l^2 \tag{9.14}$$

b. Flywheel about Its Central Axis (Fig. 9.10). The cardinal rule for computing moments of inertia is to select an element dm, which has all its mass equidistant from the axis of rotation. In the present instance this is clearly a thin cylindrical shell of thickness dr. If the mass per unit volume is ρ, $dm = 2\pi r\, dr\, l\rho$, provided that the thickness of the flywheel is l. Thus

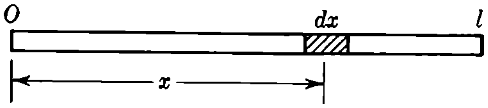


FIG. 9.9. Moment of inertia of a rod.

$$I = \int_0^R 2\pi l \rho r^3 dr = \frac{\pi}{2} l \rho R^4$$

and since $M = \pi R^2 l \rho$, this becomes

$$I = \tfrac{1}{2} M R^2 \tag{9.15}$$

Other examples may be treated by the same method. But there are sometimes short cuts based on mathematical tricks, one of which will here be illustrated.

c. Sphere about an Axis through Its Center. Clearly it makes no difference whether we compute I about the X axis, the Y axis, or the Z axis, so long as each passes through the center. The results must be equal, though the mathematical expressions look different.

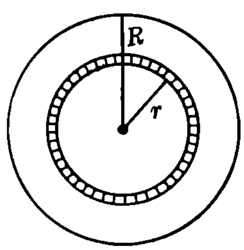


FIG. 9.10.

$$I_x = \sum_i m_i(y_i^2 + z_i^2)$$
$$I_y = \sum_i m_i(z_i^2 + x_i^2)$$
$$I_z = \sum_i m_i(x_i^2 + y_i^2)$$

if I_x is the moment computed about X, and so forth. If we add all these, we get three times the quantity we want; therefore

$$3I = 2 \sum_i m_i(x_i^2 + y_i^2 + z_i^2) = 2 \int_0^R (x^2 + y^2 + z^2) dm = 2 \int_0^R s^2 dm$$

Here s is the distance of a mass point from the center of the sphere, not from the axis of rotation. For a spherical shell of radius s and thickness ds,

$$dm = 4\pi s^2 ds \cdot \rho.$$

Hence

$$3I = 2\int_0^R s^2 \times 4\pi\rho s^2 ds = 2 \times 4\pi\rho \frac{R^5}{5}$$

and $$I = \tfrac{8}{15}\pi\rho R^5$$

Since $M = (\tfrac{4}{3})\pi R^3\rho$, this is equivalent to

$$I = \tfrac{2}{5}MR^2 \tag{9.16}$$

Table 9.1. *Moments of Inertia*

Rod..................	Transverse axis through center	$\tfrac{1}{12}Ml^2$
Circular disk..........	Perpendicular axis through center	$\tfrac{1}{2}MR^2$
Circular cylinder......	Transverse axis through center	$\tfrac{1}{4}M(R^2 + \tfrac{1}{3}l^2)$
Rectangular block.....	Axis through center, perpendicular to face with side lengths a and b	$\tfrac{1}{12}M(a^2 + b^2)$
Sphere, solid..........	Axis through center	$\tfrac{2}{5}MR^2$
Thin spherical shell....	Axis through center	$\tfrac{2}{3}MR^2$
Right cone, radius of base R.............	Axis of figure	$\tfrac{3}{10}MR^2$
Ellipsoid of axes $2a$, $2b$, $2c$.................	Axis $2a$	$\tfrac{1}{5}M(b^2 + c^2)$

9.10. Theorem of Parallel Axes. Calculations like those of the preceding section can be simplified by an interesting theorem. Often we know the moment of inertia of a body about an axis through the center of mass and wish to find it about some other axis. This can be done almost without calculation. In Fig. 9.11, which represents a section through a rigid body, O is the point at which the axis of rotation, which is perpendicular to the plane of the figure, pierces this plane. C is the point where an axis parallel to the one through O and passing through the center of mass intersects the figure. The distance between O and C is $\bar{r}$, and for convenience we choose the X axis so that it passes through both O and C. The moment of inertia of the body about O is

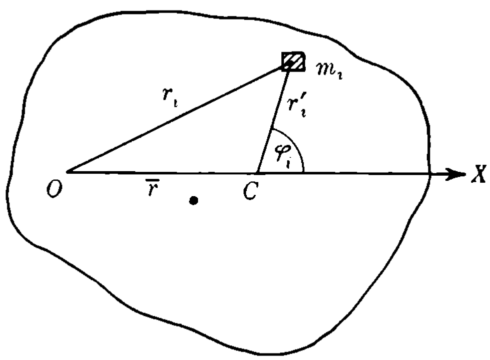


FIG. 9.11. Theorem of parallel axes.

$$I_o = \sum_i m_i r_i^2 \tag{9.17}$$

Now from Fig. 9.11 (using the cosine law)

$$r_i^2 = (r_i')^2 + \bar{r}^2 + 2\bar{r}r_i' \cos \varphi_i,$$

and $r_i' \cos \varphi_i = x_i'$, the x coordinate of m_i relative to the center of mass. Hence

$$I_o = \sum_i m_i r_i'^2 + \sum_i m_i \bar{r}^2 + 2\bar{r} \sum_i m_i x_i' \tag{9.18}$$

The first sum on the right is I_c, the moment of inertia about the center of mass; the second is M, the mass of the body, times $\bar{r}^2$, the square of the distance between the two parallel axes; and the last sum vanishes because $\sum_i m_i x_i'/M$ represents the distance of the x component of C from C [see Eqs. (9.8)], which is zero. Therefore,

$$I_o = I_c + M\bar{r}^2 \tag{9.19}$$

***9.11. Worked Examples.** *a.* A grindstone of radius 1 ft and mass 10 lb, initially at rest, is being turned by means of a crank having an 8-in. arm (cf. Fig. 9.12). What constant force applied perpendicular to the arm will give the stone an angular velocity of 2 rps in 30 sec? From Eq. (9.15),

$$I = \tfrac{1}{2} \times 100 \text{ lb} \times (1 \text{ ft})^2 = 50 \text{ lb ft}^2.$$

If the wheel has an angular speed of 2 rps, a point on the

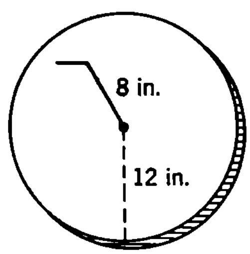


FIG. 9.12.

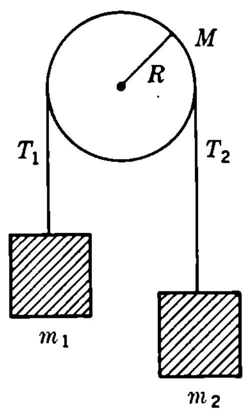


FIG. 9.13.

circumference must have a linear speed of $2 \times 2\pi r$ ft/sec, r being the radius of the wheel. If this speed is to be acquired under a constant acceleration in 30 sec, this acceleration is $2 \times 2\pi r/30$ ft/sec^2, and this corresponds to an angular acceleration

$$\alpha = 2 \times 2\pi/30 \text{ radians/sec}^2 = 0.419 \text{ radian/sec}^2.$$

Now we apply the equation

$$L = I\alpha = 50 \text{ lb ft}^2 \times 0.419 \text{ sec}^{-2} = 21.0 \text{ poundal ft}$$

But L equals the force times $\tfrac{2}{3}$ ft, so that

$$F = \frac{21.0 \text{ poundal ft}}{\tfrac{2}{3} \text{ ft}} = 31.5 \text{ poundals} = 0.978 \text{ lb}$$

b. A pulley of mass 200 gm carries two masses of 1 kg and 1.5 kg, respectively, as indicated in Fig. 9.13. The string is perfectly flexible and inextensible. We wish to find the acceleration of the masses.

Since the pulley has mass and is accelerated, the tensions on its two sides are *not* equal. Indeed the pulley behaves in accordance with the equation

$$L = (T_1 - T_2)R = I\alpha \tag{9.20}$$

Let us denote the downward acceleration of m_1 by a. The mass m_1 is then subject to the relation

$$m_1 a = m_1 g - T_1$$

while m_2 obeys

$$-m_2 a = m_2 g - T_2$$

Subtracting the last from the next to the last equation we find

$$(m_1 + m_2)a = (m_1 - m_2)g - (T_1 - T_2)$$

According to Eq. (9.20), $T_1 - T_2 = \frac{I\alpha}{R} = \frac{1}{2}\frac{MR^2}{R}\frac{a}{R} = \frac{1}{2}Ma$ because of Eq. (9.15). Therefore

$$(m_1 + m_2)a = (m_1 - m_2)g - \frac{1}{2}Ma \qquad a = \frac{m_1 - m_2}{m_1 + m_2 + \frac{1}{2}M}g$$

It is interesting to note that the *size* of the pulley does not enter the problem; R drops out of the final answer. Numerically, we have

$$a = \frac{(1{,}500 - 1{,}000)\text{gm}}{(1{,}500 + 1{,}000 + 100)\text{gm}} 980 \text{ cm/sec}^2 = 188 \text{ cm/sec}^2$$

The difference $T_1 - T_2$ is 100×188 dynes, or 19.2 gm. These results should be compared with those of Sec. 6.5.

c. Find the moment of inertia of a ring or hoop about an axis perpendicular to its plane and passing through its circumference.

As we have seen, I with respect to an axis through the center is MR^2. From the theorem of parallel axes [Eq. (9.19)], we must add to this the quantity $M\bar{r}^2$, that is, MR^2. Hence the answer is $I = 2MR^2$.

PROBLEMS

1. The mass per unit length of a thin rod of length l is proportional to the cube of the distance from one end. If the constant of proportionality is k, find the position of the center of mass of the rod.

★**2.** Find the position of the center of mass of a homogeneous right circular cone of altitude h and base radius a.

3. Applying the theorem of parallel axes, find (a) the moment of inertia of a sphere about a tangential axis and (b) the moment of inertia of a circular cylinder about a generatrix.

4. Find the center of mass of that portion of a thin, homogeneous elliptical plate ($x^2/a^2 + y^2/b^2 = 1$) that lies in the first quadrant.

5. The legs of the table shown in Fig. 9.14 are 2 by 2 by 30 in. and weigh 10 lb apiece. The top is a rectangular slab, 40 by 60 by 2 in., weighing 60 lb. Locate the center of mass of the table.

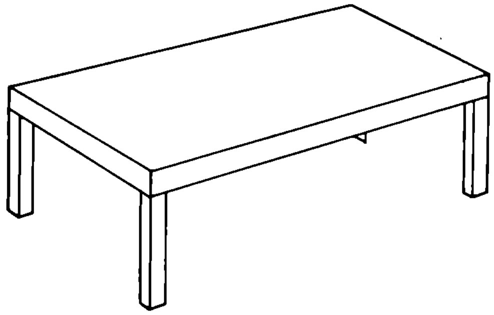


FIG. 9.14. Problem 5.

6. The pattern shown in Fig. 9.15 is cut from a thin homogeneous lamina. Find the position of its center of mass.

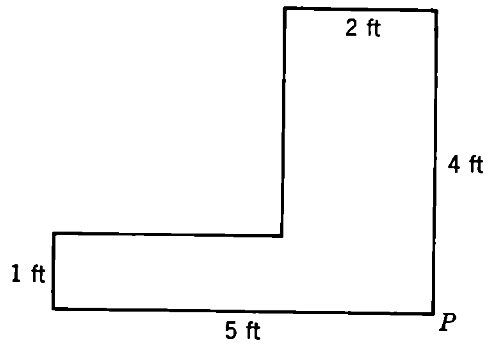


FIG. 9.15. Problem 6.

***7.** The moment of inertia of a circular lamina about a transverse axis through its center is $\frac{1}{2}MR^2$. Show that the moment of inertia about a diameter is $\frac{1}{4}MR^2$. (Use procedure similar to that of Sec. 9.9*c*.)

***8.** By direct calculation find the moment of inertia of a homogeneous right circular cone of base radius R about its axis of revolution.

***9.** Find the moment of inertia of a thin spherical shell of radius R about a diameter.

10. The mass per unit area μ of a circular lamina is directly proportional to the distance from the center ($\mu = kr$). Calculate the moment of inertia of the lamina about a transverse axis through its center; about an axis through a diameter.

11. Calculate the moment of inertia of a thin uniform rod about a perpendicular axis at the midpoint.

***12.** Use the theorem of parallel axes to calculate the moment of inertia of a cylinder of radius R and length l about a transverse axis through its center (see Prob. 7).

***13.** Two masses, 50 gm and 30 gm, are fixed to the end of a string that passes over a frictionless pulley of radius 10 cm. Under the action of gravity the 50-gm mass acquires a downward acceleration of 200 cm/sec². Calculate the moment of inertia of the pulley and the tension in the string on each side of the pulley.

14. A vertical disk is free to rotate on a horizontal axle, of radius 1 in., as shown in Fig. 9.16. A string is wrapped tightly around the axle. When a 5-lb weight is attached to the string and allowed to descend, the disk completes its first revolution in 2 sec. Find the moment of inertia of the disk and axle assembly.

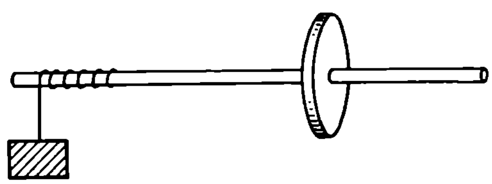


FIG. 9.16. Problem 14.

CHAPTER 10

DYNAMICS OF A RIGID BODY—Continued

10.1. Work and Energy in Rotational Motion.[1] Figure 10.1 represents a section of a rigid body that is free to rotate about an axis through O, the axis extending at right angles to the paper. Suppose that the ith particle is subject to a force $\mathbf{F}_i$. The work done by $\mathbf{F}_i$ when the particle is given displacements dx_i, dy_i is, in accordance with Eq. (8.11), given by

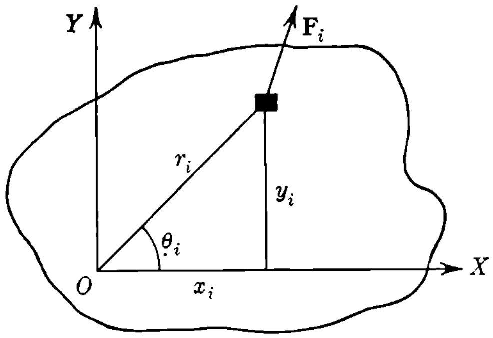


FIG. 10.1. Work in rotational motion.

$$dW_i = F_{ix}dx_i + F_{iy}dy_i$$

But $x_i = r_i \cos \theta_i$, $y_i = r_i \sin \theta_i$, whence

$$dx_i = -r_i \sin \theta_i d\theta_i$$
$$dy_i = r_i \cos \theta_i d\theta_i$$

When these relations are substituted into the expression for dW_i, we obtain

$$dW_i = (-F_{ix}y_i + F_{iy}x_i)d\theta_i \qquad (10.1)$$

But according to the definition of torque [Eq. (3.5)] the expression in parenthesis is L_i, the torque acting on the ith particle. Hence

$$dW_i = L_i d\theta_i$$

If the body is *rigid* and rotates about O, the increment in θ is the same for one particle as it is for any other, although, of course, the $\mathbf{F}_i$, the x_i, y_i, and θ_i are different. We may therefore omit the subscript on $d\theta_i$ and write $dW_i = L_i d\theta$.

There is one such equation for each particle of the body, since i may take on the values 1, 2, . . . up to the total number of particles. When all these equations are added, there results

$$\boxed{dW = L\, d\theta} \qquad (10.2)$$

Here dW is the work done when the whole body is given an infinitesimal displacement $d\theta$ under a torque L about a fixed axis. Note the similarity of this equation with Eq. (8.8), $dW = F\, dx$.

[1] Review Sec. 8.3.

Since power is defined as dW/dt in general, the formula for power spent in rotational motion is clearly

$$P = L\frac{d\theta}{dt} = L\omega$$

This should be compared with Eq. (8.27).

If the rigid body revolves from an initial position characterized by θ_1 to a final position θ_2 (cf. Fig. 10.2) (the reference line OP is drawn from

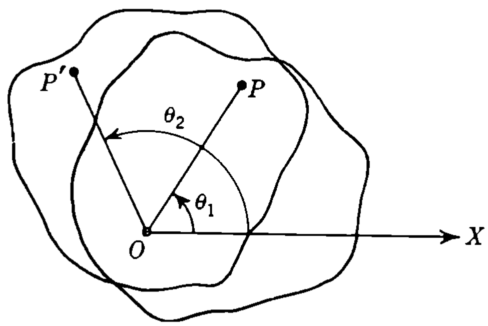


FIG. 10.2. Work in rotational motion.

a point on the axis to any fixed reference point P in the body), the work done upon it is

$$W = \int_{\theta_1}^{\theta_2} L\,d\theta$$

But if we use Eq. (9.13) and recall that $\alpha = d\omega/dt$, while $d\theta = \omega\,dt$, we find

$$W = \int_{\theta_1}^{\theta_2} I\frac{d\omega}{dt}\,\omega\,dt = \int_{\omega_1}^{\omega_2} I\omega\,d\omega = \tfrac{1}{2}I(\omega_2^2 - \omega_1^2) \qquad (10.3)$$

provided ω_1 and ω_2 are the angular velocities at θ_1 and θ_2, respectively. Thus, if we take the kinetic energy to be

$$E_{\text{kin}} = \tfrac{1}{2}I\omega^2 \qquad (10.4)$$

Eq. (10.3) tells us that the work done during the displacement from θ_1 to θ_2 is equal to the increase in kinetic energy, just as was found to be the case for the linear motion of a particle, where $E_{\text{kin}} = \frac{1}{2}mv^2$.

10.2. Kinetic Energy of a Body Rotating about a Fixed Axis. There is another way of proving Eq. (10.4). The kinetic energy of the ith particle of a rigid body is $\frac{1}{2}m_iv_i^2$. If the body is rotating with angular velocity ω about a fixed axis whose perpendicular distance from m_i is r_i, then $v_i = r_i\omega$. Hence

$$E_{\text{kin}} = \sum_i \tfrac{1}{2}m_iv_i^2 = \sum_i \tfrac{1}{2}m_ir_i^2\omega^2 = \tfrac{1}{2}\left(\sum_i m_ir_i^2\right)\omega^2 = \tfrac{1}{2}I\omega^2$$

From the result one can draw an interesting conclusion. Suppose that the body in Fig. 10.3 revolves momentarily about an axis through O. According to the theorem of parallel axes (Sec. 9.10) the moment of inertia about O is

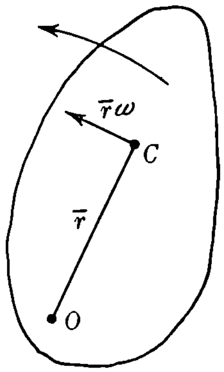


FIG. 10.3. Decomposition of E_{kin} into rotational and translational parts.

$$I = I_c + M\bar{r}^2$$

so that $$E_{\text{kin}} = \tfrac{1}{2}I_c\omega^2 + \tfrac{1}{2}M\bar{r}^2\omega^2$$

But $\bar{r}\omega$ is the speed v_c with which the center of mass, C, moves about O. Hence

$$E_{\text{kin}} = \tfrac{1}{2}I_c\omega^2 + \tfrac{1}{2}Mv_c^2 \tag{10.5}$$

In words, the kinetic energy of a rigid body is the sum of two parts: (1) the kinetic energy of *rotation* about an instantaneous axis through the center of mass, $\tfrac{1}{2}I_c\omega^2$; (2) the kinetic energy of *linear* motion of the center of mass with the entire mass of the body assumed at the center of mass, $\tfrac{1}{2}Mv_c^2$.

***10.3. The Rolling Wheel.** The motion of a rolling wheel presents some curious and interesting features. There are two ways in which it may be described.

1. One can say that the wheel rotates with an angular velocity ω_c about a *moving* axis through C, while C progresses with a linear velocity v_c (cf. Fig. 10.4).

2. One can also say that the wheel rotates with an angular velocity ω_p about an axis through the point P, *an axis which is instantaneously at rest.* Both points of view are correct.

As the point of contact moves from P to P', the wheel turns through an angle θ and C advances a distance s. From the figure, $s = R\theta$. Upon differentiating we find

$$v_c = \frac{ds}{dt} = R\frac{d\theta}{dt} = R\omega_c$$

and the acceleration of C is

$$a_c = R\frac{d\omega_c}{dt} = R\alpha$$

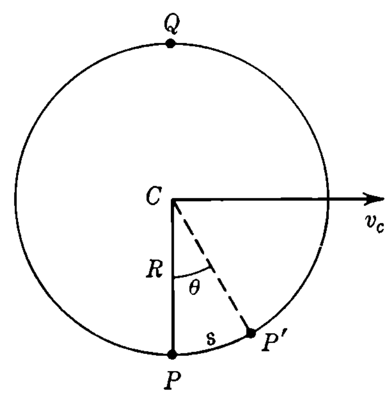


FIG. 10.4. The rolling wheel.

What is the relation between ω_c and ω_p? It may seem surprising that they are equal. But we have just shown that $v_c = R\omega_c$; on the other hand C has a velocity $R\omega_p$ as a result of its rotation about P. If these two velocities are to be equal,

$$\omega_c = \omega_p$$

Let us calculate the velocity of the point Q, using both methods. According to (1) we find v_Q by adding to v_c the velocity of rotation of Q about C, namely, $R\omega$, and this gives $2v_c$. According to (2), $v_Q = 2R\omega_p$, and this also is $2v_c$. The point P will be seen to have zero velocity by both methods.

***10.4. The Rolling Cylinder, Accelerated.** The motion of a cylinder down a rough incline (Fig. 10.5) is a composition of the following: an acceleration of the center of mass C along the plane and a simultaneous rotation of the cylinder about C. We have indicated in Sec. 9.3 that C moves as though all the external force were applied at C. Now the sum of the forces *perpendicular* to the plane is obviously zero, for there is no acceleration in that direction. Along the plane two forces are active, the component of gravity and the frictional retardation **F** produced by the rough plane. The magnitude of **F** is not yet known. Nevertheless, if the linear acceleration of C is a_c and the mass of the cylinder M,

$$Ma_c = Mg \sin\theta - F \qquad (10.6)$$

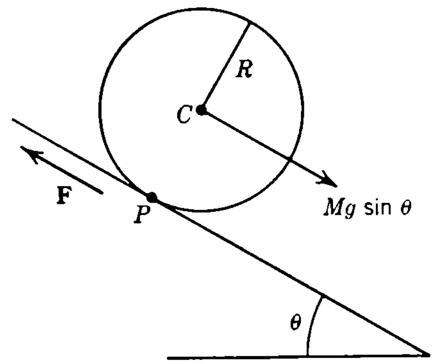


FIG. 10.5. Rolling cylinder.

We also recall that the body will rotate about C as if all external torques were applied about C. But the force $Mg \sin\theta$ has no torque about C, whereas the torque due to F is FR (disregarding the sign). Therefore

$$FR = I\alpha \qquad (10.7)$$

If the cylinder does not slip along the plane, $a_c = R\alpha$, and Eq. (10.7) gives

$$F = \frac{Ia_c}{R^2}$$

For a cylinder, $I = \frac{1}{2}MR^2$; hence

$$F = \tfrac{1}{2}Ma_c$$

When this is introduced in Eq. (10.6), we find

$$a_c = \tfrac{2}{3}g \sin\theta \qquad (10.8)$$

The acceleration is only two-thirds of what it would be in frictionless sliding. The retarding force F has the value $\frac{1}{3}Mg \sin\theta$. It is supplied by friction between the cylinder and the plane. If the coefficient of friction is μ, then the maximum force of friction is given by $\mu F_\perp$, which is $\mu Mg \cos\theta$. The force F cannot exceed this value; that is, $\frac{1}{3}Mg \sin\theta < \mu Mg \cos\theta$, and hence $\mu > \frac{1}{3}\tan\theta$. When this inequality fails, the cylinder will slip.

Our problem could have been solved more easily by applying energy considerations. When the cylinder rolls a distance s, its potential energy decreases by

$$Mgh = Mgs \sin\theta$$

Provided that it started from rest, it will have acquired kinetic energy of amount

$$\tfrac{1}{2}(\tfrac{1}{2}MR^2)\omega^2 + \tfrac{1}{2}Mv_c^2$$

according to Eq. (10.5). But since $R\omega = v_c$, the kinetic energy is $\frac{3}{4}Mv_c^2$. The principle of conservation of mechanical energy then requires that

$$Mgs \sin \theta = \tfrac{3}{4}Mv_c^2$$

or $v_c^2 = \frac{4}{3}gs \sin \theta$. From our study of uniformly accelerated motion we see that this must equal $2a_cs$ (cf. Sec. 5.5). Hence we find

$$a_c = \tfrac{2}{3}g \sin \theta$$

in agreement with Eq. (10.8).

To apply the principle of conservation of mechanical energy in a problem like the present, which involves friction, may at first seem inconsistent. It is to be observed, however, that the force F, though present, *does no work;* for the point P, which this force attacks, is momentarily at rest. This provides justification for assuming that no energy is dissipated in the motion.

10.5. Angular Momentum. In a typical lecture demonstration a man stands at the center of a turntable (piano stool), with his arms extended outward and a weight in each hand. He is given an angular velocity about a vertical axis. Then, while rotating at a constant speed, he draws in his arms, and his angular velocity increases surprisingly.

In explanation of these facts we invoke the fundamental law of rotational motion [Eq. (9.13)], which we now write in the more general form

$$L = \frac{d}{dt}(I\omega) \tag{10.9}$$

When I is constant, this is equivalent to Eq. (9.13). When I changes, Eq. (10.9) is correct and Eq. (9.13) is not applicable.

When $L = 0$, integration yields

$$I\omega = \text{const} \tag{10.10}$$

The quantity $I\omega$ is called *angular momentum*, in analogy to the linear momentum mv. What we have shown is this:

In the absence of external torques, the angular momentum of a rotating system remains unchanged.

This statement is called the principle of conservation of angular momentum.

When a torque is present, Eq. (10.9) reads, after integration,

$$\int_{t_1}^{t_2} L\, dt = (I\omega)_2 - (I\omega)_1$$

The quantity on the left-hand side is sometimes termed "angular impulse" (cf. Sec. 8.12); and the equation reads: The angular impulse equals the change in angular momentum

In the experiment involving the man on the turntable there is certainly no external torque applied to the system while the man draws in his arms. Hence Eq. (10.10) must be obeyed. When I is reduced in value, ω must increase correspondingly.

If the unweighted man has a moment of inertia of 100 lb ft^2 and the weights (10 lb each) are originally 3 ft from the axis of rotation, the total I before reduction was $(100 + 20 \times 3^2)$ lb ft^2 = 280 lb ft^2. Afterward, when the weights are pulled in to a distance of, let us say, 1 ft from the axis, I is $(100 + 20 \times 1)$ lb ft^2 = 120 lb ft^2. Therefore ω must increase by a factor $^{280}/_{120}$ in this example. The weight of the man's arms has here been neglected.

The reader may have wondered what happens to the energy in this process. For clearly, if $I_1\omega_1 = I_2\omega_2$, then $\frac{1}{2}I_1\omega_1^2$ cannot be equal to $\frac{1}{2}I_2\omega_2^2$. Energy is therefore *not* conserved. The reason is that, as the man draws in the weights, the centripetal force which he applies does work, so that the energy is *greater* when the rotation is more rapid.

10.6. Rotational and Translational Motion; Analogies. The general parallelism between the laws of motion, as they apply to "translation" (nonrotary motion) and to rotation about a fixed axis, cannot have escaped the reader. In this section we present a review and a systematic exhibition of the similarities between the two types of motion. The similarities concern the *quantities* used to describe the motions and the *laws* that govern them.

Corresponding Quantities

Translation	*Rotation*
Displacement x	Angular displacement θ
Speed, $dx/dt = v$	Angular speed, $d\theta/dt = \omega$
Acceleration, $dv/dt = a$	Angular acceleration, $d\omega/dt = \alpha$
Mass m	Moment of inertia, $I = \sum_i m_i r_i^2$
Force F	Torque, $L = Fr$
Linear momentum mv	Angular momentum $I\omega$
Linear impulse $= \int F\,dt$	Angular impulse $= \int L\,dt$

General Laws

Translation	*Rotation*
$F = ma$	$L = I\alpha$
$dW = F\,dx$	$dW = L\,d\theta$
$E_{\text{kin}} = \frac{1}{2}mv^2$	$E_{\text{kin}} = \frac{1}{2}I\omega^2$
Linear impulse $= \Delta(mv)$	Angular impulse $= \Delta(I\omega)$

There is also a thoroughgoing similarity between *special* laws for the two types of motion. Here are a few examples:

Special Laws

Uniform Motion

Translation	*Rotation*
$a = 0,\ v = \text{const}$	$\alpha = 0,\ \omega = \text{const}$
$s = s_0 + vt$	$\theta = \theta_0 + \omega t$

Uniformly Accelerated Motion

Translation	*Rotation*
$F = \text{const},\ a = \text{const}$	$L = \text{const},\ \alpha = \text{const}$
$v = v_0 + at$	$\omega = \omega_0 + \alpha t$
$s = s_0 + v_0 t + \frac{1}{2}at^2$	$\theta = \theta_0 + \omega_0 t + \frac{1}{2}\alpha t^2$
$v^2 - v_0^2 = 2as$	$\omega^2 - \omega_0^2 = 2\alpha\theta$

***10.7. Gyroscopic Motion.** The motion of an ordinary flywheel is simple because the axis of rotation does not change. If the wheel is so suspended that the axle can change its orientation in space, for example by means of gimbal rings, new effects can be observed that are not easily understandable in terms of the laws studied thus far, for these laws assumed the existence of a fixed axis of rotation. These effects are nevertheless common and very important.

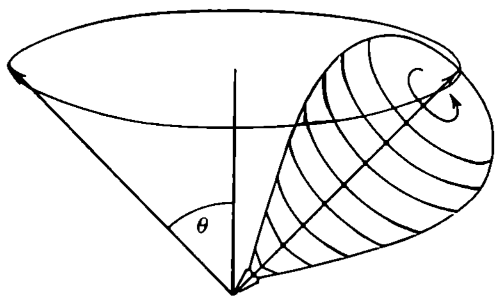


FIG. 10.6. Spinning top.

A spinning top wobbles; its axis of rotation itself revolves on the surface of a cone, as indicated in Fig. 10.6. This wobbling motion is called *precession*. It is the steady motion of the arrowhead in Fig. 10.6 upon the circular path drawn. Sometimes—in fact, usually—this motion is accompanied by a periodic change in the angle θ, which causes the spin axis to bob up and down while precessing. The latter rhythmic oscillation is called *nutation*. In the following, nutation will be disregarded.

The earth's axis of rotation performs a precession in space. When an airplane, whose propeller rotates clockwise, makes a right turn, it noses down. On making a left turn it noses up. If a pilot proceeds to dive, the airplane's course changes to the left and the right rudder must be used to stabilize the airplane. When a naval vessel turns too rapidly, its revolving machinery may be damaged because of its tendency to precess. The housings of bearings are sometimes torn for this reason.

On the other hand a rapidly spinning wheel in a universal mounting has a tendency to maintain its axis of rotation when its position is changed. This is sometimes called *gyroscopic stability*, or *rigidity of the gyroscopic axis*. It is utilized in the construction of gyrocompasses, artificial horizons, and direction indicators. Guns are rifled in order to give the shell a spin that will maintain its alignment with respect to the trajectory.

One of the best known examples of precession occurs when the rider of a bicycle turns his front wheel. If he turns to the right, the wheel tends to throw him off to the left, and vice versa. This effect is much stronger than a similar centrifugal effect. To understand it we must first learn how to represent angular velocities and torques as *vectors*.

★10.8. Angular Quantities as Vectors. In Chap. 3, torques about a specified axis were divided into positive (counterclockwise) and negative (clockwise) torques. This is a useful distinction so long as the axis is fixed in space. When this is not true, the vector concept must be associated with torques. The following rule has proved successful:

A *torque* is represented by a vector (arrow) having a magnitude proportional to the magnitude of the torque. The vector points along the axis of rotation in that direction in which a right-handed screw would advance under the action of the torque (cf. Fig. 10.7).

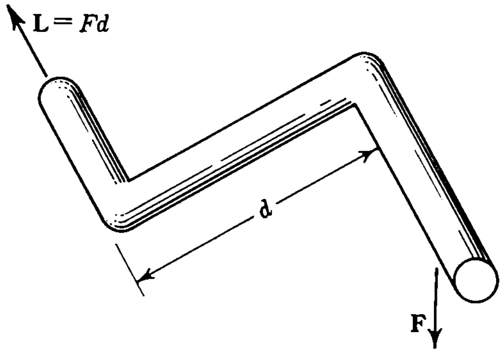


Fig. 10.7. Torque as vector.

In a similar way an *angular velocity* is represented by a vector, of magnitude proportional to ω and pointing in the direction in which a right-handed screw would advance when turned in the sense of ω. Thus the angular velocity of the earth is a vector pointing north along the earth's axis of rotation. This rule for associating a vector with an angular quantity is often called the "right-handed screw rule."

Angular momenta, too, are represented by vectors in accordance with this rule.

The student doubtless wonders whether this picturization of angular quantities as vectors is a mere convention or whether there is something in their nature which makes them truly vectors in a sense in which mass, for example, is *not* a vector. The answer, which cannot be fully explained without higher mathematics, is that these quantities are truly vectors and that the rule is not merely a convention. One proof of this may be seen in the circumstance that it yields positive information, as will be shown in the next section. On the other hand not every angular quantity is a vector; an angle, for example, is not.

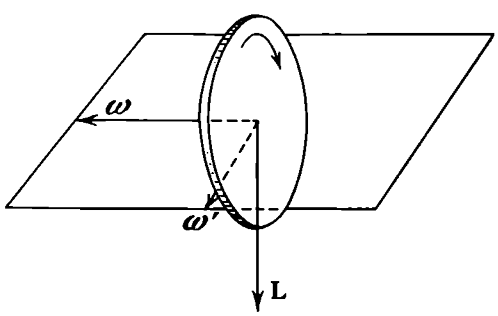


Fig. 10.8. Relation of spin, torque, and precession.

★10.9. Law of Precession. We now return to the case of the spinning bicycle wheel that is being turned to the right by a torque on the handle bars. The situation is depicted schematically in Fig. 10.8. The wheel has an angular velocity given by the vector $\boldsymbol{\omega}$ to the left; it is subjected to a downward torque **L**. Notice, however, that the vectors do not denote the same physical quantities, nor do they have the same physical dimensions.

We know from Sec. 10.5 that

$$L = \frac{d}{dt}(I\omega)$$

and this is in fact a vector relation, which we shall now write in the form

$$\mathbf{L}\,dt = d(I\boldsymbol{\omega})$$

The right-hand side is the increment in angular momentum that occurs in time dt when a torque $\mathbf{L}$ is applied to the wheel. The quantity $\mathbf{L}\,dt$ has indeed the physical dimension of angular momentum, as the student will readily verify. In Fig. 10.9 are drawn, as solid lines, the vectors $I\boldsymbol{\omega}$ and $\mathbf{L}\,dt$. Their addition yields the dotted vector whose length is equal to that of $I\boldsymbol{\omega}$ because $\mathbf{L}\,dt$ has an infinitesimal length but whose direction differs by $d\phi$ from that of $I\boldsymbol{\omega}$. We may say that the original angular momentum $I\boldsymbol{\omega}$ has turned through an angle $d\phi$ as a result of $\mathbf{L}$. Hence we have *precession* about an axis coming out of the paper, with an angular velocity

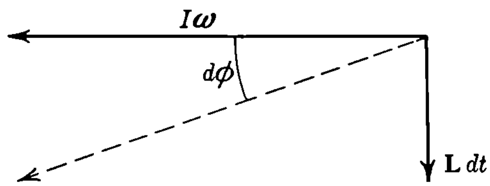


FIG. 10.9. Precession.[1]

$$\omega' = \frac{d\phi}{dt} \tag{10.11}$$

This velocity is drawn in Fig. 10.8.

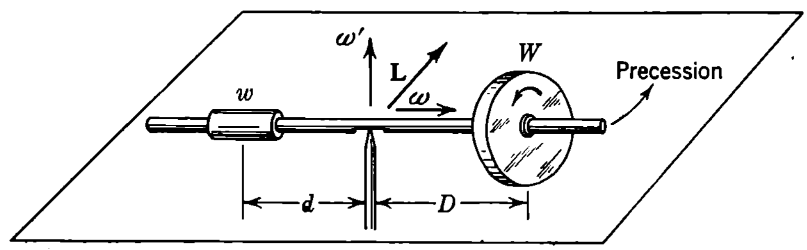


FIG. 10.10. Precession.

But Fig. 10.9 shows that $L\,dt/I\omega = \tan(d\phi) = d\phi$, the last because the tangent of a small angle, like the sine, is equal to the angle itself. Hence we find, by transposing,

$$\boxed{\frac{d\phi}{dt} = \frac{L}{I\omega}} \tag{10.12}$$

In words, the angular velocity of precession equals the torque producing it divided by the spin angular momentum; its direction is at right angles to both spin and torque vectors. The best way to remember the sense, or direction, of the precession is to note that, while precessing, the *spin axis chases the torque axis*.

[1] When this figure is compared with Fig. 10.8, it should be noted that it lies in a *vertical* plane.

Figure 10.10, showing a wheel balanced by a counterweight, provides an interesting demonstration of Eq. (10.12). If the rotating disk has an overweight, the torque is as indicated and the precession will be about a vertical axis. By increasing or decreasing the overweight the precession can be made to go fast or slow, and Eq. (10.12) can be checked quantitatively when it is noted that L equals $WD - wd$ (cf Fig. 10.10) if w and W are the weights of load and wheel. By pushing the weight w far enough to the left to reverse the torque the precession is also seen to reverse.

PROBLEMS

1. A mass m is revolving at a distance r_1 about an axis with an angular velocity ω_1. Suddenly the distance from the axis is increased to r_2. (*a*) Show that the angular velocity now becomes $(r_1/r_2)^2\omega_1$. (*b*) Show that the kinetic energy is decreased by the amount $\frac{1}{2}m\omega_1^2 r_1^2(1 - r_1^2/r_2^2)$. (*c*) Show that this is precisely the amount of work done against the centripetal force during the extension from r_1 to r_2.

2. A thin 3-ft rod weighing 8 lb may rotate freely in a vertical plane about an axis through one end. The rod is held in a horizontal position and then allowed to swing down under the action of gravity. What is the angular velocity of the rod at the instant it passes through a vertical position?

★3. A sphere of mass 2 kg and radius 10 cm rolls down a 25° inclined plane. (*a*) Find the linear acceleration of the sphere. (*b*) Find the angular acceleration of the sphere. (*c*) Determine the value of the coefficient of friction necessary to prevent the sphere from slipping.

★4. A yo-yo weighing 30 gm has a moment of inertia of 120 gm cm² about its spin axis. If it is allowed to descend on a string a vertical distance of 50 cm starting from rest, determine its final velocity of spin. Assume the smallest radius of the yo-yo to be 1 cm.

5. Assume that the total power output of all man-made machines is of the order of 1 billion (10^9) kilowatts. If the earth's rotation could be gradually stopped, how long could it supply the power needed by man? (Mass of the earth is 6×10^{24} kg, its radius is 6,400 km; assume it to be of uniform density.)

6. A cylindrical grindstone weighs 500 lb and has a radius of 2 ft. (*a*) If it starts from rest, what constant torque must be applied to give it a speed of 200 rpm in 30 sec? (*b*) What is the angular acceleration? (*c*) How much work is done on the grindstone during the 30 sec? (*d*) How many revolutions does the stone make during the 30 sec?

★7. A sphere rolls without slipping on a vertical semicircular track (see Fig. 10.11). The sphere is started from rest at point A. Show that the normal component of the force exerted by the sphere on the track when it passes B is $(17/7)Mg$.

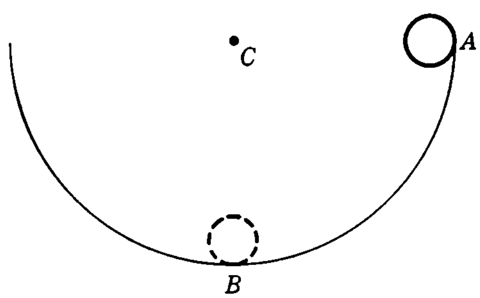


FIG. 10.11. Problem 7.

8. A torque $L = 10^3 \times \theta$ dyne cm acts upon the axis of a flywheel of moment of inertia 10 gm cm^2 while it turns from $\theta = 0$ to $\theta = 2\pi$. Find the increase in the kinetic energy of the flywheel. If its initial angular velocity was 10π radians/sec, find its final angular velocity.

9. A solid 100-lb cylinder, 2 ft in radius, rolls along a horizontal surface with a velocity of 3 ft/sec. What is its translational kinetic energy? What is its rotational kinetic energy? To what height would the cylinder roll on a 30° inclined plane?

10. An automobile engine rotating at a speed of 1,500 rpm is developing 60 hp. What torque does it deliver?

11. How far will a cylinder 10 cm in diameter roll up a 30° inclined plane if its velocity at the bottom is 20 cm/sec? (Use principle of conservation of energy.)

12. A thin rod, weighing 4 lb and 2 ft long, is revolving in a vertical plane about an axis through one end. Its angular velocity as it moves through the horizontal position is 2 radians/sec. To what height does the center of mass of the rod rise above the axis of rotation?

13. A uniform cylinder weighing 50 lb and 2 ft in radius is free to rotate about a horizontal axis. A mass of 2 lb is attached to the end of a rope, which has been wrapped around the cylinder. If the mass is allowed to fall, rotating the cylinder, find (*a*) the acceleration of the mass, (*b*) the tension in the rope, and (*c*) the angular acceleration of the cylinder.

14. A circular hoop of radius 3 ft, hanging on a nail, is turned until the line through its center and the nail makes an angle of 30° with the horizontal and is then released. Calculate the initial angular acceleration of the hoop.

15. The drive shaft of an electric motor is 2 in. in diameter. A belt passing over the drive shaft is used to transfer energy from the motor to another mechanical device. The tension in the slack side of the belt is 30 lb and in the tight side 150 lb. If the motor makes 1,500 rpm, what horsepower does it deliver?

***16.** A hoop of mass 5 lb and radius 2 ft rolls down a 30° inclined plane from a height of 10 ft. Calculate the linear acceleration of the hoop, as well as its total energy at the bottom of the plane.

CHAPTER 11
MOTIONS UNDER INVERSE-SQUARE LAWS OF FORCE

11.1. Gravitation. The search for an explanation of *why* bodies fall is as old as man himself. The early answers, half factual and half mythological, do not satisfy the modern scientist; yet it is also true that the answer that he himself accepts, the answer given by Newton, is regarded as incomplete by many scientists and philosophers. For Newton's explanation states precisely *how* bodies fall, not *why* they fall. But to modern science the "how" is all-important. It is expressed by the *law of universal gravitation.*

During the sixteenth and seventeenth centuries much speculation arose concerning the forces that hold together the bodies of the planetary system. In the year 1664 the students at Cambridge, England, were discussing the latest theories dealing with these forces. Among them was a twenty-two-year-old "scholar of the college," named Isaac Newton, from Woolsthorpe in Lincolnshire. The next year the plague broke out, school was temporarily suspended, and the students sent home. It was during this enforced period of idleness that Newton began to wonder—presumably as he was lying in the orchard of Woolsthorpe and saw an apple fall to earth—whether the same force of gravitation, which attracts the apple, might not perchance reach the moon and cause its deviation from a straight-line path.

Not content with idle speculation and having the good fortune to know mathematics, the young scholar set to work on the problem of what the force must be if it is to cause both the apple's fall and the motion of the moon in its orbit. He found that it must be inversely proportional to the square of the distance between attracting bodies. In modern terminology and without further attention to the very interesting historical sequence of discoveries that led to its complete formulation we state the law of universal gravitation as follows:

The force between any two mass particles, of masses m_1 and m_2 and separated by a distance r, is an *attraction* that acts along the line joining the particles and has the magnitude

$$F = G\frac{m_1 m_2}{r^2} \tag{11.1}$$

The constant G is called the constant of universal gravitation. Its value was roughly determined by Newton; its modern value is

$$G = 6.670 \times 10^{-8} \text{ dyne cm}^2/\text{gm}^2 \tag{11.2}$$

We note that Eq. (11.1) expresses the force between mass *particles* of infinitesimal size. If the force between extended objects is wanted—as, for instance, that between bodies A and B of Fig. 11.1—each of them must be regarded as decomposed into particles and the interaction between all particles must then be computed by integration. It is *not* correct to apply Eq. (11.1) and take for r the distance between the centers of mass of A and B. Newton was well aware of this point; and there is some evidence to indicate that one of the motives which led him to invent the calculus was the desire to be able to perform this integration!

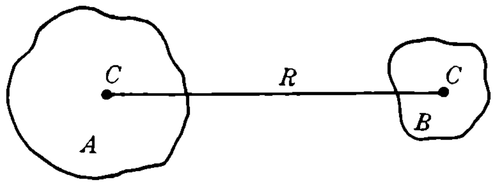


FIG. 11.1. Force between A and B is *not* $G\frac{M_A M_B}{R^2}$.

11.2. Determination of G. The gravitational force between bodies of ordinary size is extremely small (a fraction of a dyne); hence only an extraordinarily sensitive experiment can produce an accurate value of G. Such an experiment was performed by Henry Cavendish in 1798, using a balance of the type now named after him.

Figure 11.2 shows the arrangement in horizontal section. Two small, heavy (platinum) spherical masses m are placed on the ends of a light rod. At C this system is suspended from a fine vertical fiber. When the system is in equilibrium, two large lead spheres M are introduced. The attraction between m and M causes the rod to rotate slightly toward the large masses. The twist in the fiber is registered by a moving light beam, reflected from a mirror that is carried on the fiber.

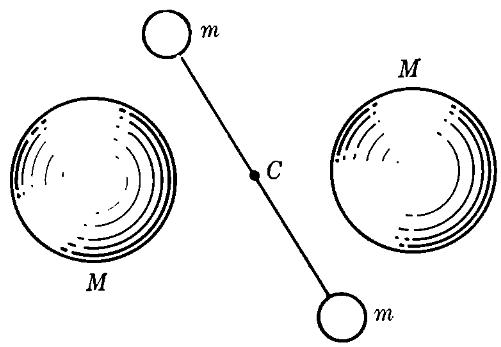


FIG. 11.2. Cavendish experiment.

The value given in Eq. (11.2) was thus found. Note that G cannot be a pure number. The left-hand side of Eq. (11.1) has the physical dimensions $[MLT^{-2}]$, the right-hand side $[M^2L^{-2}]$, except for G. Hence G must be the ratio of these, namely, $[M^{-1}L^3T^{-2}]$. Notice that the units given in Eq. (11.2) are in agreement with this result. If masses are measured in pounds and distances in feet, the numerical value of G is different.

The attracting objects in Fig. 11.2 are, in fact, extended, and we should expect that Eq. (11.1) does not apply to them directly. It will be shown, however, that uniform *spheres* play a unique role inasmuch as they act gravitationally as though all their mass were concentrated at their centers. If two spherical boulders, each weighing 1,000 kg (about 1 ton) lie with their centers 1 m (about 3 ft) apart, they attract each other with a force of

$$\frac{6.67 \times 10^{-8} \times 10^6 \times 10^6}{100^2} \text{ dynes} = 6.67 \text{ dynes}$$

or about 7 mg.

By means of his experiment Cavendish is sometimes said to have "weighed the earth." What is meant is that he determined the earth's mass, for the earth has no weight. If both G and g (acceleration of gravity) are known, the mass of the earth M_e can be determined from the equation

$$G\frac{M_e m}{R^2} = mg \tag{11.3}$$

which says that the force on a mass m at the earth's surface, that is R cm from the earth's center, is mg. Hence

$$M_e = \frac{g}{G}R^2 = \frac{980}{6.67 \times 10^{-8}}(6.37 \times 10^8)^2 \text{ gm} = 5.97 \times 10^{27} \text{ gm or } 6.6 \times 10^{21} \text{ tons}$$

Finally let us compute the change in g that occurs as we proceed outward from the earth's surface. For this purpose it is most convenient to take the logarithm of Eq. (11.1) and then to differentiate with respect to r,

$$\ln F = \ln G + \ln m_1 + \ln m_2 - 2 \ln r$$

$$\frac{dF}{F} = -2\frac{dr}{r}$$

since G, m_1, and m_2 are constants. Thus the relative increase in F is twice the relative decrease in r. In going up 10 miles from the earth's surface, r changes from approximately 4,000 miles to 4,010 miles, or by $1/400$. Hence g must change by $1/200$, or from about 980 cm/sec^2 to 975 cm/sec^2. We see, therefore, that g is very nearly constant near the earth's surface despite the action of an inverse-square law. It does vary, of course, with altitude and with latitude, as was discussed in Sec. 7.7.

11.3. Gravitational Effect of a Spherical Mass. Since most applications of the law of universal gravitation are made to spherical astronomical bodies, it becomes necessary to solve a mathematical problem of great fundamental importance: What is the force between a large spherical mass and a mass particle?

Consider the spherical shell, an equatorial section of which is drawn in Fig. 11.3. Its thickness t is small compared with its diameter. We wish to find its gravitational force on an external mass point m'.

A small mass of the shell situated at A attracts m' with a force $\mathbf{F}_1$. Another equal mass at B exerts the force $\mathbf{F}_2$. The combined effect of these two small masses is given by the vector sum of $\mathbf{F}_1$ and $\mathbf{F}_2$. In adding them we notice that the vertical force components annul each other, while the horizontal components $F_1 \cos \alpha$ and $F_2 \cos \alpha$ are equal and must be added algebraically. In this way, all transverse forces on m' cancel in pairs, a mass in the upper hemisphere always annulling the downward component of an equal and symmetrically situated mass in the lower hemisphere of the shell. To find the total effect of the shell we need consider only horizontal components.

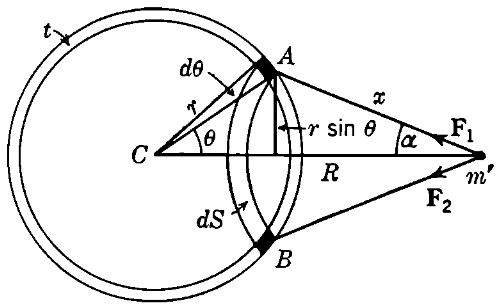


FIG. 11.3. Attraction of a spherical shell.

As an element of mass of the shell we take the circular strip dS. Its length is $2\pi r \sin \theta$, its width $r\,d\theta$, and its thickness t. Hence it has a volume

$$2\pi t r^2 \sin \theta \, d\theta$$

and if the density is ρ gm/cm³, the mass within the strip is

$$dm = 2\pi t \rho r^2 \sin \theta \, d\theta$$

The force exerted by dm on m' is

$$dF = G \frac{dm\, m'}{x^2} \cos \alpha = 2\pi G t \rho m' r^2 \frac{\sin \theta \, d\theta}{x^2} \cos \alpha \tag{11.4}$$

There now appear the three variables θ, x, and α. Before integrating it is necessary to express two of them in terms of the third. We shall take x as our variable of integration.

The figure shows that

$$\cos \alpha = \frac{R - r \cos \theta}{x} \tag{11.5}$$

Also

$$x^2 = R^2 + r^2 - 2Rr \cos \theta \tag{11.6}$$

and hence

$$r \cos \theta = \frac{R^2 + r^2 - x^2}{2R} \tag{11.7}$$

If we differentiate Eq. (11.6), we obtain

$$2x\,dx = 2Rr \sin \theta \, d\theta$$

or

$$\sin \theta \, d\theta = \frac{x}{Rr}\, dx \tag{11.8}$$

Now put Eq. (11.7) into (11.5), and then insert Eqs. (11.5) and (11.8) into (11.4). The result is

$$dF = \frac{\pi G t \rho m' r}{R^2} \frac{R^2 - r^2 + x^2}{x^2}\, dx \tag{11.9}$$

To sweep over the entire shell, x must range from $x_{\min} = R - r$ to $x_{\max} = R + r$.

Now

$$\int_{R-r}^{R+r} \left(\frac{R^2 - r^2}{x^2} + 1\right) dx = 4r$$

Hence

$$F = \int_{R-r}^{R+r} dF = G\frac{(4\pi r^2 t\rho)m'}{R^2} \tag{11.10}$$

The quantity in parenthesis is simply M, the total mass of the shell.

We have proved, therefore, that *a spherical shell attracts an external mass point as if all its mass were concentrated at its center C.*

A body like the earth, moon, or sun may (except for departures from sphericity) be regarded as composed of a large number of concentric shells of different densities. The result just stated must therefore hold for these astronomical objects, also.

11.4. A Mass inside a Spherical Shell. It is a curious fact that the mass m', when placed *inside* the spherical shell as shown in Fig. 11.4, experiences *no force whatever.*

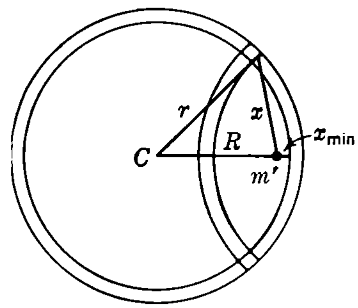


FIG. 11.4. A mass inside a spherical shell.

To prove this we integrate just as in the preceding section. But we notice that R is now smaller than r and that the minimum value of x is now $r - R$ instead of $R - r$. As before, $x_{max} = R + r$.

In integrating Eq. (11.9) the limits are now $r - R$ and $R + r$. But

$$\int_{r-R}^{R+r} \left(\frac{R^2 - r^2}{x^2} + 1\right) dx = 0$$

Hence

$$F = 0$$

If a channel were dug through the earth along a diameter, the gravitational force on a particle would become smaller the deeper we go, for the portions of matter external to the position of the particle exert no forces. The acceleration of gravity is smaller in deep mines than on the earth's surface.

11.5. Facts about the Solar System. The most important regularities in the motion of planets were discovered by Johann Kepler (1571–1630) and are contained in Kepler's three *laws of planetary motion.* They read

1. All planets move in elliptical orbits having the sun at one focus.
2. A line drawn from the sun to the planet, though changing its length, sweeps out equal areas in equal times.
3. The cube of a planet's mean distance from the sun is proportional to the square of its period of revolution.

Newton's greatest achievement was to derive all three of these from his law of gravitation [Eq. (11.1)]. The complete proof is usually given in second- and third-year physics courses and will not be produced in this book. Only the third law will be derived (cf. Sec. 11.6).

The orbits of the planets about the sun and the orbit of the moon are nearly circular. In the following discussion they may be taken as circles with good approximation. Numerical data concerning the solar system, which may be used in problems, are collected in Table 11.1.

Table 11.1. *Data Relating to the Solar System*
(In cgs system)

	Earth	Sun	Moon
Masses.......	$M_e = 5.97 \times 10^{27}$ gm	$M_s = 1.99 \times 10^{33}$ gm	$M_m = 7.34 \times 10^{25}$ gm; $\frac{M_s}{M_e} = 332{,}000$; $\frac{M_m}{M_e} = 0.0123$
Radii.........	$R_e = 6.37 \times 10^8$ cm	$R_s = 6.97 \times 10^{10}$ cm	$R_m = 1.75 \times 10^8$ cm
Distances from earth.......	....	$R_{es} = 1.50 \times 10^{13}$ cm	$R_{em} = 3.86 \times 10^{10}$ cm

11.6. Planetary Motion. When a planet of mass m moves with speed v in a circular orbit of radius r, it has a centripetal acceleration (cf. Sec. 5.9) v^2/r. The centripetal force holding it in its orbit is the attraction by the sun, which, according to Eq. (11.1), has the magnitude $(GM_sm)/r^2$. Hence by Newton's second law the motion must satisfy the equation

$$\frac{GM_sm}{r^2} = \frac{mv^2}{r} \tag{11.11}$$

Here m, the mass of the planet, may be canceled. This interesting circumstance permits us to say that the mass of a planet has no effect on its motion and also that it cannot be found from an analysis of its motion.

Equation (11.11) may be reduced by canceling m and r. It can further be changed by expressing v in terms of the frequency of revolution, f, of the planet,

$$v = 2\pi rf \tag{11.12}$$

or in terms of the period of revolution, P,

$$v = \frac{2\pi r}{P} \tag{11.13}$$

Thus Eq. (11.11) takes the form

$$GM_s = 4\pi^2r^3f^2 = \frac{4\pi^2r^3}{P^2} \tag{11.14}$$

This is the basic equation of planetary motion. It can serve a threefold purpose.

1. When r and P are known, the mass of the sun can be calculated. Taking the motion of the earth for an example, we have (see Table 11.1).

$$r = R_{es} = 1.50 \times 10^{13} \text{ cm} \qquad P = 365 \text{ days} = 3.15 \times 10^{7} \text{ sec}$$

Hence $$M_s = \frac{4 \times 9.86 \times 3.38 \times 10^{39}}{6.67 \times 10^{-8} \times 10 \times 10^{14}} \text{ gm} = 2.0 \times 10^{33} \text{ gm}$$

and this is in good agreement with the value of M_s in Table 11.1. In a similar way Eq. (11.14) can be used to determine the mass of the *earth* when r and P relate to the *moon* (see Prob. 13).

2. When M_s and P are known, r may be calculated. This is one of the best methods for determining the distances of the various planets from the sun, since P can be found easily from astronomical observations.

3. Knowing M_s and r, one can predict P. In the future it may become possible to construct an artificial satellite of our earth. If we want this to be a given distance r from the earth, Eq. (11.14), with M_s taken to be M_e, will determine its period of revolution.

Finally we note that Eq. (11.14) is tantamount to Kepler's third law. For it may be written

$$P^2 = \left(\frac{4\pi^2}{GM_s}\right) r^3$$

the quantity in parentheses being constant for all planets.

11.7. Work Done in a Gravitational Field. The region of space surrounding a massive body, like the sun or the earth, in which particles are attracted in accordance with the inverse-square law of force is called a *gravitational field*. When a particle moves or is moved from one point to another in this field, work is done by or against the force of gravitation. Let us find out how much. We divide our problem into three parts. First we calculate the *work* done *by gravity* when a particle of mass m moves from B to A in the field of a spherical body of mass M (cf. Fig. 11.5), the path being along a radius. Second we calculate the work done when m moves from C to A along an arbitrary path, and third the work done in carrying m from infinity to the point A.

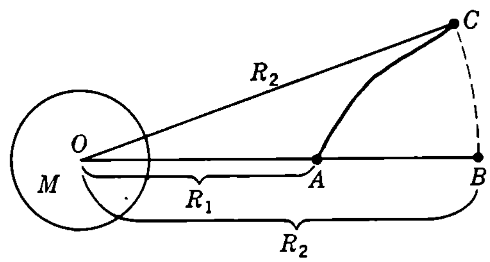


FIG. 1.15. Work in a gravitational field.

1. Since the force at any point between A and B is $G(Mm/r^2)$, and work is defined as $dW = F\,dr$, we have

$$W = \int_{R_2}^{R_1} \left(-\frac{c}{r^2}\right) dr = \frac{c}{R_1} - \frac{c}{R_2} \tag{11.15}$$

if we use the abbreviation

$$c = GMm \tag{11.16}$$

The minus sign under the integral must be written because the force, being attractive, is directed along the *negative* r axis.

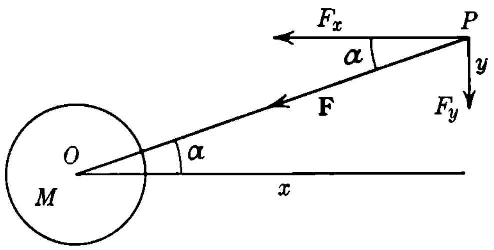


FIG. 11.6. Resolution of forces.

2. In carrying the mass from C to A the path does not lie in one dimension, and we must use the more general definition[1] of work [cf. Eq. (8.11)].

$$dW = F_x dx + F_y dy$$

Now if at any point, such as P in Fig. 11.6, the force F is directed radially toward O,

$$F_x = -F\cos\alpha \qquad F_y = -F\sin\alpha$$

But $\cos\alpha = x/r \qquad \sin\alpha = y/r$

Therefore

$$dW = -F\frac{x}{r}\,dx - F\frac{y}{r}\,dy$$

$$= -\frac{F}{r}(x\,dx + y\,dy) \tag{11.17}$$

This expression may be simplified when we remember that

$$x^2 + y^2 = r^2$$

Differentiating this relation we obtain

$$x\,dx + y\,dy = r\,dr$$

and this changes Eq. (11.17) to

$$dW = -F\,dr$$

Therefore

$$W = \int_{r_C}^{r_A} (-F)dr = \int_{r_C}^{r_A} \left(-\frac{c}{r^2}\right) dr = \frac{c}{r_A} - \frac{c}{r_C} \tag{11.17a}$$

In Fig. 11.5, we have made r_c, the distance of C from the center of the attracting mass, equal to R_2. Our result [Eq. (11.17*a*)] is then seen to be $(c/R_1) - (c/R_2)$, the same as Eq. (11.15).

The noteworthy fact is that we were able to carry out the integration without specifying the exact path which the particle took in going from

[1] This was proved in a starred section but will here be taken for granted.

C to A: according to Eq. (11.17a) the work depends only on the initial and on the final distance of the particle from O.

This remarkable result is occasioned by the mathematical circumstance that dW, as expressed by Eq. (11.17), may be reduced to $-F\,dr$. The physical meaning of it is this:

The work done in carrying a mass from one point to another in an inverse-square field of force is independent of the path.

The foregoing method shows in fact that this is true for any *central* field of force.

3. To find the work done by the force of gravity when the particle is carried from infinity to A we need only make r_C infinite in Eq. (11.17a), or R_2 in Eq. (11.15). Thus

$$W = \frac{c}{R_1} \tag{11.18}$$

In spite of the infinite length of the path, the work done is finite.

The preceding proof is somewhat analytic and may be replaced by the following simple argument. If we decompose the path CA in Fig. 11.5 into small steplike portions, drawn alternately along the radius and perpendicular to it, no work is done along the perpendicular segments, since along them the force is perpendicular to the displacement. But the work done along the radial parts of the steplike path adds up to the work done in going from B to A.

11.8. Energy in a Gravitational Field. Potential energy was defined in Sec. 8.8 as the work done *against* a force in displacing a particle to a given position. But what are we to take as the initial, or reference, position of the particle? In the motion of a spring we took the neutral (force-free) position; in dealing with the *constant* force of gravity we took the earth's surface as our plane of reference. This could still be done, but it proves inconvenient when the variation in the gravitational force is respected. It is customary, therefore, to say that the *potential energy is zero when the particle is an infinite distance from an attracting center.* This is indeed the force-free position.

Assuming this, the potential energy is simply the *negative* of W as given by Eq. (11.18), since W was the work done *by* the force. Thus

$$U = -\frac{c}{r} = -\frac{GMm}{r} \tag{11.18a}$$

provided that we write r for the distance of any point under consideration from O (Fig. 11.5).

There is hardly another situation in physics where the law of conservation of mechanical energy is so well obeyed and so fruitful in its applications as it is in gravitational fields. The law here reads

$$\frac{1}{2}mv^2 - \frac{GMm}{r} = E = \text{const} \tag{11.19}$$

From it one can calculate the increase in velocity occurring as a particle moves from a distance R_2 to a smaller distance R_1, both being measured from the attracting origin. If the velocities at R_1 and R_2 are v_1 and v_2, we have

$$\frac{1}{2}mv_1^2 - \frac{GMm}{R_1} = \frac{1}{2}mv_2^2 - \frac{GMm}{R_2}$$

and hence

$$v_1^2 - v_2^2 = 2GM\left(\frac{1}{R_1} - \frac{1}{R_2}\right) \tag{11.20}$$

It will be remembered that M is the mass of the attracting body, not the moving one. For a body falling near the surface of the earth Eq. (11.20) reduces to a familiar form. In that case R_1 and R_2 are not very different, and both are approximately equal to the radius of the earth. If the parentheses on the right be written $(R_2 - R_1)/R_1R_2$ and this is approximated by d/R^2, d being the distance traversed and R the earth's radius, the result takes on the familiar form

$$v_1^2 - v_2^2 = \frac{2GM}{R^2}d = 2gd$$

in view of Eq. (11.3).

What is the velocity of a particle that has fallen to the earth from infinity? The answer is implicit in Eq. (11.20). If the particle was at rest at ∞, then $v_2 = 0$, $R_2 = \infty$, and $R_1 = R_e$. Hence Eq. (11.20) yields

$$v_1^2 = \frac{2GM_e}{R_e}$$

Let us call this value v_∞. The particle reaches the earth with a finite speed; its numerical value is

$$v_\infty = \left(\frac{2 \times 6.67 \times 10^{-8} \times 5.97 \times 10^{27}}{6.37 \times 10^8}\right)^{1/2} \text{cm/sec}$$
$$= 1.12 \times 10^6 \text{ cm/sec} = 6.96 \text{ miles/sec}$$

Conversely, if a particle is to have a zero velocity at infinity, it must have the radial velocity v_∞ at the earth's surface. A rocket that is to leave the earth, never to return, must have a velocity of at least 7 miles/sec. This figure is not very accurate, for air resistance increases the value. It is called the theoretical *speed of escape.* A missile leaving the earth with a greater velocity will arrive at infinity with a finite speed.

The molecules of our atmosphere move very rapidly. Fortunately, the mean velocities of O_2 and N_2 are below the speed of escape. Hydrogen molecules, being lighter, have speeds that are greater, on an average, than v_∞. Hence this gas, which must have been present in the atmosphere a long time ago, has now disappeared. On the other hand it has not escaped from the sun because of the larger mass of that body. Helium, which is somewhat heavier than H_2, is still present in small quantities in the outer atmosphere but is gradually escaping.

We have seen that the potential energy of a body of mass m in a gravitational field is given by $U = -c/r$ or $-GMm/r$. Closely related to U is a quantity known as the gravitational *potential;* it is the potential energy per unit of mass U/m. Let us call it V; then

$$V = -\frac{c}{mr} = -\frac{GM}{r}.$$

11.9. The Bohr Theory of the Hydrogen Atom. An atom is a miniature solar system. The center, or *nucleus,* is relatively heavy and stationary, while light particles called *electrons* circulate about it very much like planets about a sun. The distances between sun and planets, which are vast in the solar system, are only about 10^{-8} cm in the case of an atom; the motions are much too tiny to be seen. However, they are extremely rapid, as we shall show.

In an atom the force holding the electrons in their orbits is *not* the gravitational force, which is much too small to effect this result. It is an electrical force arising from the fact that the atomic nucleus carries a positive and each electron a negative charge. The significant point, however, is that this electrical force is also an "inverse-square" attraction, varying with the distance in accordance with the law

$$F = \frac{c}{R^2}$$

where c is a constant.

The hydrogen atom is the lightest and also the simplest of all atoms. Its nucleus consists of one unit of positive charge, usually denoted by $+e$, and it contains but one planetary electron, whose charge is $-e$. Later (cf. Sec. 24.5) we shall show that under these conditions the constant in the force law is

$$c = e^2$$

The mass of the electron will be called m. In a circular orbit the centripetal force mv^2/r is supplied by the electrical attraction, so that

$$\frac{mv^2}{r} = \frac{e^2}{r^2} \tag{11.21}$$

But this result is not very useful, for the following reason. It is possible, to be sure, to measure e and m experimentally (cf. Sec. 25.12); hence these two quantities are known, but, in addition to these, Eq. (11.21) contains *two* unknowns, v and r, neither of which can be directly determined. It seems, therefore, that our theory is useless. The ordinary laws of mechanics have nothing to say about the velocity of the electron, nor about its distance from the nucleus.

An ingenious discovery by Niels Bohr (1885–), made in 1913 and called the *Bohr postulate*, completely altered this ineffectual situation. His brilliant guess, though entirely implausible when first inspected, leads to results that are found correct on being tested—leads, in fact, to the modern quantum theory of the atom. It asserts that *the angular momentum of the electron in a circular orbit is always equal to an integral number* (n) *of certain quanta of angular momentum.* A quantum of angular momentum has the value $h/2\pi$, h being the famous *Planck constant* (6.610×10^{-27} erg sec). Hence

$$I\omega = nh/2\pi \tag{11.22}$$

Here I is the moment of inertia of the electron, ω its angular velocity, n the integer called a *quantum number.* Clearly, $I = mr^2$, $\omega = v/r$. Thus

$$mrv = nh/2\pi \tag{11.23}$$

Between the two equations (11.21) and (11.23) we can now eliminate either r or v and solve for the other. We thus obtain

$$r = n^2 \frac{h^2}{4\pi^2 me^2} \tag{11.24}$$

$$v = \frac{1}{n} \frac{2\pi e^2}{h} \tag{11.25}$$

When values are inserted, the quantity $h^2/4\pi^2me^2$ is seen to have the value 5.28×10^{-9} cm. The electron can therefore move on "quantized" circular orbits, the smallest of which has a radius of that magnitude, the next 4 times, the next 9 times this value, and so on.

The quantity $2\pi e^2/h$ is found to be 2.19×10^8 cm/sec, a rather enormous speed. Ordinarily the hydrogen atom is in the state for which $n = 1$, and its electron circulates with that speed.

According to Eq. (11.19) the total energy of our electron is

$$E = \tfrac{1}{2}mv^2 - e^2/r \tag{11.26}$$

since the quantity GMm, the former c, must now be replaced by e^2. If we substitute Eqs. (11.24) and (11.25) into (11.26), we find the simple result

$$E = -\frac{2\pi^2 me^4}{n^2 h^2} \tag{11.27}$$

The quantity $2\pi^2 me^4/h^2$ has the value 2.15×10^{-11} erg. If we call it E_0, we may write

$$E = -\frac{E_0}{n^2} \tag{11.28}$$

The energy of the hydrogen atom can be $-E_0$, or $-\frac{1}{4}E_0$, or $-\frac{1}{9}E_0$, and so forth. This is a result of very great importance. As we shall show in Chap. 49, the light emitted by gaseous hydrogen allows a determination of these "discrete" energy states, and Eq. (11.28) turns out to be verified with great precision.

PROBLEMS

1. A thin, straight rod of infinite extent has a mass λ per unit length. Calculate the gravitational force between the rod and a small mass m, which is a perpendicular distance b from the rod.

2. Assume that the density of the earth is everywhere constant and that a small hole has been drilled from the earth's surface to the center. What will be the acceleration of gravity at a point in this hole halfway between the surface and the center?

3. How much work is required to raise a mass of 10 kg from the surface of the earth to a point 2,000 km above the surface?

4. A thin rod is of length l and mass per unit length λ. A spherical mass m is placed in line with the rod and at a distance d from one end. Find the gravitational force between the rod and the mass.

5. Using the necessary data from Table 11.1 find the weight of a 1-lb mass on the surface of the moon.

6. Two spherical masses m and M are a distance d apart. Where must a small mass be placed in order to be in equilibrium under the attractions of m and M?

7. Two spheres weighing 1 kg apiece are suspended on strings 1 m long and 5 cm apart at the top. What angle will the strings make with the vertical?

8. The planet Jupiter is 74,000 km in radius. It is completely encircled by a satellite once every 16.7 days. The radius of the orbit of the satellite is 27 times the radius of the planet. Calculate the mass of Jupiter.

9. It is desired to make artificial satellites of the earth that shall cruise (a) eastward, (b) westward over the equator, completing one round trip per day as judged by an observer on the equator. Find the distances from the earth's surface at which these satellites must move.

10. It is desired to make an artificial satellite of the earth, a structure that shall be continuously overhead in a certain locality on the equator. Find the distance from the earth's surface at which this satellite must move.

11. A mass m is shot upward at right angles to the surface of the earth with a velocity of 10 km/sec. How far out will it go before turning back? (Neglect air resistance.)

12. Find the acceleration of gravity on the surface of the sun, using the necessary data in Table 11.1.

13. Using the value of R_{cm} given in Table 11.1, and Eq. (11.14), calculate the mass of the earth. Take the period of the moon to be 27 days 8 hr.

14. Find the value of G in the fps system and in the mks system.

15. Convert the data in Table 11.1 into *tons* and *miles.*

***16.** Two uniform rods, each of length 1 m and weight 1 kg, are arranged as in Fig. 11.7. The distance between their centers is 2 m. Find the gravitational force between them. Compare this answer with the result of wrongly assuming that the rods interact as though all their mass were concentrated at their centers.

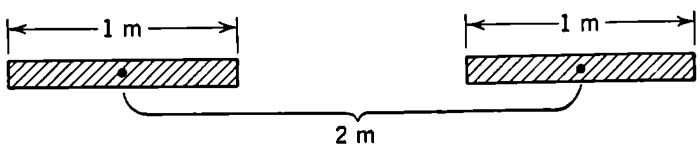


FIG. 11.7. Problem 16.

CHAPTER 12

OSCILLATIONS

12.1. Simple Harmonic Motion. Motions that repeat themselves after a period of time are said to be *periodic*. Examples of periodic motion are readily available—the swinging pendulum, the vibrations or "oscillations" of elastic bodies, the rotating wheel, the moving piston, the crankshaft, the tree swaying in the wind, the heartbeat, the motion of planets, and many others. In all these instances the *period* of the motion is the time interval after which it repeats itself, the time of a single rhythm.

By far the simplest of all periodic motions are those performed by elastic bodies. Their study forms the basis for an understanding of *all* periodic motions and will concern us here. The instantaneous displacements of the parts of an elastic body, when plotted against the time, yield—as will be shown—sine or cosine curves. Because these are called *harmonic* curves, the motion itself is called *harmonic* motion. In general, harmonic motion may be a composite one in which several periods are present simultaneously. In the presence of a single period the motion is called *simple harmonic motion*, henceforth abbreviated SHM.

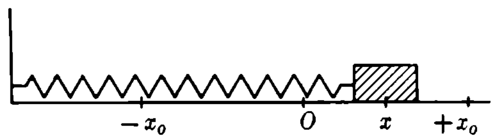


FIG. 12.1. Simple harmonic motion.

A typical arrangement for producing this ideal type of motion is a spring carrying a mass at its end and resting on a smooth horizontal table, as in Fig. 12.1. When at rest, the mass lies at its neutral point O, which we take as origin of coordinates. If the mass is displaced to x_0 and then released, it will continue to oscillate between $-x_0$ and $+x_0$. The distance x_0, which is half the total excursion of the mass, is called the *amplitude* of the SHM. The *period* P is the time taken by the mass to move from x_0 to $-x_0$ and back to x_0, or the time for one complete excursion. We know from Chap. 4 [cf. Eq. (4.1)] that the force acting on m at x is

$$F = -kx \qquad (12.1)$$

the minus sign indicating that the force is to the left when x is positive, to the right when x is negative. The constant k, called the "coefficient of stiffness," represents the force per unit extension and may be measured in dynes/cm. Indeed every elastic body, in obeying Hooke's law, calls into being a force that is proportional to the displacement x, and this accounts for the wide occurrence of SHM.

12.2. Mathematical Representation of SHM. To find the motion that results from a force given by Eq. (12.1) we must write and solve Newton's second law,

$$m\frac{d^2x}{dt^2} = -kx \tag{12.2}$$

To simplify matters we use an abbreviation

$$\frac{k}{m} = \omega^2 \tag{12.3}$$

which is suggested by the fact that the physical dimension of $\sqrt{k/m}$ is indeed that of an angular velocity. At present Eq. (12.3) is nothing but a convenient substitution, though we shall learn later that it has much physical meaning. Equation (12.2) then becomes

$$\frac{d^2x}{dt^2} = -\omega^2 x \tag{12.4}$$

This is sometimes said to be the "differential equation for SHM"; every x satisfying it describes SHM. Its physical meaning is the following:

The acceleration is *proportional* and *opposed in direction* to the displacement.

If the student studies the motion of the mass in Fig. 12.1, he will find this statement to be true for all values of x.

The mathematics of solving Eq. (12.4) is rather typical of the way in which differential equations are dealt with. There is no royal road, and the steps must simply be remembered. Equation (12.4) cannot be integrated directly. But we may put $v = dx/dt$, so that $d^2x/dt^2 = v\,dv/dx$. It then reads

$$v\frac{dv}{dx} = -\omega^2 x$$

or

$$v\,dv = -\omega^2 x\,dx$$

On integration this becomes

$$\frac{1}{2}v^2 = \frac{-\omega^2}{2}x^2 + C \tag{12.5}$$

This is a rather interesting result. Multiplication by m, transposition, and use of Eq. (12.3) give

$$\tfrac{1}{2}mv^2 + \tfrac{1}{2}kx^2 = \text{const}$$

which is nothing other than the law of conservation of energy [Eq. (8.22)].

But let us return to Eq. (12.5) and write it in the form

$$v = \frac{dx}{dt} = \sqrt{2C - \omega^2 x^2} = \omega\sqrt{A^2 - x^2}$$

where we have put $2C/\omega^2$ equal to A^2, another constant. Hence

$$\frac{dx}{\sqrt{A^2 - x^2}} = \omega\, dt$$

Integral tables show that

$$\int \frac{dx}{\sqrt{A^2 - x^2}} = \sin^{-1}\frac{x}{A}$$

Hence we find on integrating

$$\sin^{-1}\frac{x}{A} = \omega t + \delta$$

δ being a constant of integration. On taking the sine of both sides there results the final answer

$$x = A \sin(\omega t + \delta) \tag{12.6}$$

The quantities A and δ are constants of integration and are therefore arbitrary.

12.3. Properties of SHM. When Eq. (12.6) is plotted, one obtains the curve of Fig. 12.2. The name *harmonic* motion is obviously justified. The quantity A is the maximum value of x; it is the amplitude and equals x_o in Fig. 12.1.

$$A = x_o$$

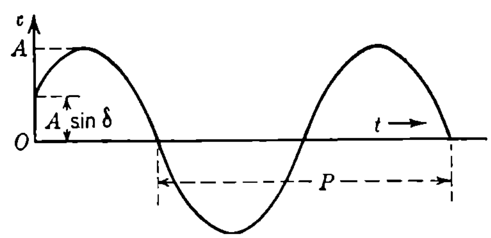


FIG. 12.2. Graph of SHM.

The intercept of the sine curve on the x axis at O is $A \sin \delta$. The quantity $(\omega t + \delta)$ is called the *phase* of the SHM, and δ is the *phase constant.* The phase increases as time goes on. The period P is the time interval marked off in Fig. 12.2; it is the time in which the phase increases by 2π radians. It is found by putting

$$[\omega(t + P) + \delta] - (\omega t + \delta) = 2\pi$$

Therefore

$$P = \frac{2\pi}{\omega} \tag{12.7}$$

This looks very much like the relation between the period and the angular velocity in uniform angular motion [cf. Eq. (5.24)], but so far that is merely an accident. On using Eq. (12.3) we have

$$\boxed{P = 2\pi \sqrt{\frac{m}{k}}} \tag{12.8}$$

a very useful result since it allows at once the calculation of the period of the motion when mass and stiffness of the spring are given.

The frequency f is $1/P$; hence

$$f = \frac{\omega}{2\pi} = \frac{1}{2\pi}\sqrt{\frac{k}{m}} \tag{12.9}$$

Here again, ω appears as though it were some sort of angular velocity; yet there is nothing rotating in the motion of the spring.

One final point concerning the mathematical form of SHM, Eq. (12.6). This could be expanded to read

$$x = A \cos \delta \sin \omega t + A \sin \delta \cos \omega t$$

Here $A \cos \delta$ and $A \sin \delta$ are just new constants that might as well be called a and b. Clearly, then, our SHM could also be represented by the formula

$$x = a \sin \omega t + b \cos \omega t \tag{12.10}$$

and, in special cases, either a or b may be zero.

12.4. Worked Examples. *a.* A spring is arranged as in Fig. 12.1. It carries a mass of 2 oz, and a force of 2 lb stretches it 16 in. Find its frequency of oscillation, its period, its amplitude, and its phase constant.

The equations in the preceding sections, being obtained with the use of Newton's second law in its absolute form, hold in *absolute units* only. In our example

$$k = \frac{2 \text{ lb-force}}{16 \text{ in.}} = \frac{64 \text{ poundals}}{4/3 \text{ ft}} = 48 \text{ poundals/ft}$$

$$m = 1/8 \text{ lb}$$

From Eq. (12.8), $$P = 2\pi\sqrt{\frac{1/8 \text{ lb}}{48 \text{ lb/sec}^2}} = 0.32 \text{ sec}$$

$$f = 1/P = 3.13 \text{ sec}^{-1}$$

The amplitude and phase constant cannot be determined from the data thus far available. They are found from the *initial* conditions of the motion. For instance, if the mass was released at $t = 0$ with a displacement of 6 in., we find that $A = 6$ in. and $\delta = \pi/2$.

b. If the spring of Example *a* is hung *vertically* and loaded with a mass of 2 oz, what will be its period of oscillation?

Assume the spring to be so light as to have a negligible mass itself. The mass of 2 oz pulls it down 1 in. Hence the extension of the spring in its neutral position is 1 in. greater than in Example *a*. However, k has still the previous value, and so has m. Period and frequency are therefore the same.

c. When the prong of a tuning fork is drawn aside 0.2 in. and then released, it snaps back with an acceleration of 50,000 in./sec². Find the frequency of the fork.

We are told that $d^2x/dt^2 = -50{,}000$ in./sec² when $x = 0.2$ in. Using Eq. (12.4),

$$\omega^2 = \frac{-d^2x/dt^2}{x} = \frac{50{,}000 \text{ in./sec}^2}{0.2 \text{ in.}} = 250{,}000 \text{ sec}^{-2}$$

$$\omega = 500 \text{ sec}^{-1}$$

From Eq. (12.9), therefore, $f = 500/2\pi$ sec⁻¹.

d. A body is known to describe SHM. Its amplitude is 1 ft; its velocity while passing through the neutral point is 10 ft/sec. What is its frequency?

If $x = x_o \sin(\omega t + \delta)$, differentiation gives us

$$v = x_o\omega \cos(\omega t + \delta)$$

The maximum speed, possessed when the body passes through the neutral point, is clearly

$$v_{\max} = x_0\omega$$

Hence
$$\omega = \frac{v_{\max}}{\text{amplitude}} = \frac{10 \text{ ft/sec}}{1 \text{ ft}} = 10 \text{ sec}^{-1}$$

and
$$f = \omega/2\pi = 10/2\pi \text{ sec}^{-1}$$

12.5. The Mystery of ω. What does the fictitious angular velocity ω have to do with the to-and-fro motion of the mass in Fig. 12.1? Suppose we do make a point P' go around a circle (Fig. 12.3) with an angular velocity ω, and we watch the x component of r. Since $x = r\cos\theta$ and $\theta = \omega t$, we see that

$$x = r\cos\omega t$$

and comparison with Eq. (12.6) shows that x describes SHM [with $\delta = \pi/2$, since $\sin(\omega t + \pi/2) = \cos\omega t$]. Its amplitude is r, the radius of the circle.

We may look upon P as the shadow of P' cast by a distant light source upon the diameter of the circle. In mathematical language, P is the projection of P'. What we have shown is this:

The projection of a point moving in uniform circular motion describes SHM.

It is sometimes convenient to associate with a SHM that circular motion of which it is the projection. The (fictitious) circle on which it takes place is called the *reference circle*, and the circular motion the *reference motion*. The quantity ω that appeared throughout our analysis is the angular velocity of the reference motion. This affords a nice mental picture of what ω means.

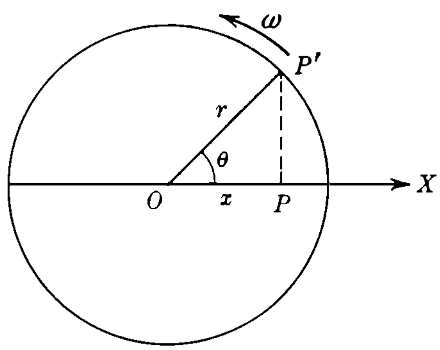


FIG. 12.3. Reference motion.

The same result can be obtained in another way. The point P' has an actual acceleration v^2/r toward O. Its projection along X is $(-v^2/r)\cos\theta$, or $(-v^2/r^2)r\cos\theta$, which equals $-\omega^2 x$. Hence the acceleration of x is $-\omega^2 x$. In view of Eq. (12.4) this shows that P describes SHM.

12.6. The Simple Pendulum. A simple pendulum is a small bob suspended by an inextensible, weightless cord. The bob moves along the arc of a circle, in periodic fashion to be sure; whether this motion is SHM remains to be investigated.

In the momentary position shown in Fig. 12.4 two forces act on the bob—gravity, mg, and the tension along the cord, T. The component of mg along s is $-mg\sin\theta$; that along the cord is $mg\cos\theta$. But since there is no acceleration along the cord, the forces T and $mg\cos\theta$ must cancel each other; hence

$$T = mg\cos\theta$$

This leaves as the resultant of T and mg a force

$$F = -mg\sin\theta$$

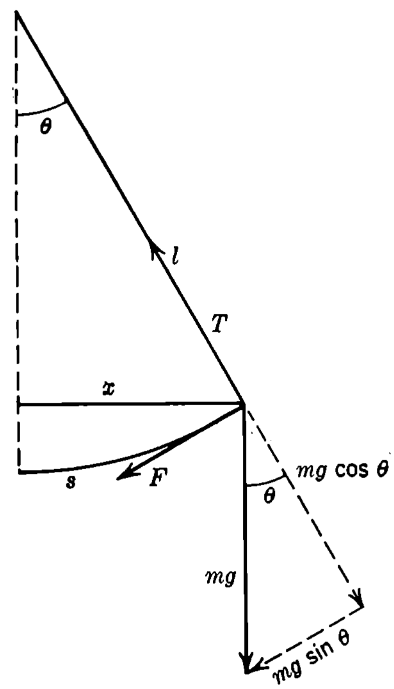


FIG. 12.4. Simple pendulum.

directed tangentially along the arc. According to Newton's second law,

$$m\frac{d^2s}{dt^2} = -mg\sin\theta$$

But

$$\sin\theta = x/l$$

so that $$\frac{d^2s}{dt^2} = -\frac{g}{l}x$$

If x were equal to s, this equation would imply SHM, for it would then say that the acceleration is (negatively) proportional to the displacement. But Fig. 12.4 shows that x and s are, in general, different. Only for small displacements are they approximately equal. For small displacements, then,

$$\frac{d^2x}{dt^2} = -\frac{g}{l}x$$

Under these conditions the motion is SHM, and comparison with Eq. (12.4) shows that

$$\omega = \sqrt{g/l}$$

whence $$f = (1/2\pi)\sqrt{g/l}$$

and $$P = 2\pi\sqrt{l/g} \tag{12.11}$$

Further investigation shows that the motion of a simple pendulum is essentially SHM so long as the maximum value of θ is smaller than 15°. Equation (12.11) is then correct to within 0.5 per cent.

The fact that P (for small oscillations) is independent of the amplitude of motion was recognized by Galileo. Because of this a pendulum clock keeps correct time regardless of variations in the width of its swings.

Equation (12.11) is useful for the determination of g. If we measure l and find P by counting swings for a certain time, g can be computed.

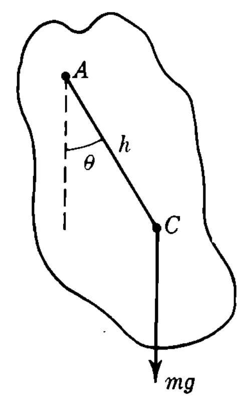


FIG. 12.5. Physical pendulum.

***12.7. The Physical Pendulum.**[1] Any irregular body suspended as in Fig. 12.5 and swinging about a fixed axis through A is called a "physical pendulum." If h is the distance between A and the center of mass C, the torque about A is $-mgh \sin\theta$. In view of Eq. (9.13),

$$I\frac{d^2\theta}{dt^2} = -mgh\sin\theta$$

where I is the moment of inertia of the physical pendulum about A. Again for *small* angles θ, $\sin\theta$ is nearly equal to θ, and

$$\frac{d^2\theta}{dt^2} = -\frac{mgh}{I}\theta$$

This represents SHM provided that we take as our displacement, not x, but the

[1] Review Sec. 9.8.

angle θ. Here $\omega^2 = mgh/I$, so that

$$P = 2\pi \sqrt{\frac{I}{mgh}} \tag{12.12}$$

12.8. Angular SHM. Let a solid disk be hung from a wire as shown in Fig. 12.6. When the disk is rotated through an angle θ, the wire gives rise to a *restoring torque* of magnitude $-\tau\theta$, the minus sign indicating that the torque is in the sense opposite to θ. The quantity τ is called the "torsion constant" of the wire. The motion is described by

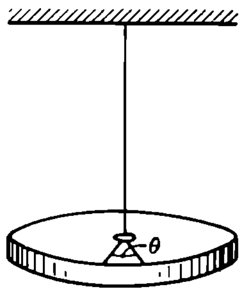


FIG. 12.6. Torsion pendulum.

$$I\frac{d^2\theta}{dt^2} = -\tau\theta$$

I being the moment of inertia of the disk about its transverse axis, or $\frac{1}{2}MR^2$. The motion is SHM, in this instance not only for small amplitudes but for the entire range of θ in which Hooke's law, $L = -\tau\theta$, is obeyed. The period of oscillation is clearly

$$P = 2\pi \sqrt{\frac{I}{\tau}} \tag{12.13}$$

This formula provides a method for measuring the torsion constant.

12.9. Summary on SHM. Any motion that satisfies the *defining* equation

$$\frac{d^2u}{dt^2} = -\omega^2 u \tag{12.14}$$

is SHM. The displacement variable u in this equation may be x, s, θ, or any other quantity specifying the position of an object, and ω^2 is simply a *positive* constant; its being positive is ensured by taking it as a square.

The frequency of the motion is $f = \omega/2\pi$, the period $P = 2\pi/\omega$. The solution of Eq. (12.14) is

$$u = A \sin (\omega t + \delta)$$

A and δ are arbitrary constants, any values of which are compatible with Eq. (12.14). Physically A is the amplitude of the motion, δ its phase at $t = 0$.

From this result it is easy to see that the total energy of the oscillator,

$$E = E_{\text{kin}} + E_{\text{pot}} = \frac{1}{2} m \left(\frac{du}{dt}\right)^2 + \frac{1}{2} ku^2 = \frac{1}{2} m\omega^2 A^2 \cos^2 (\omega t + \delta) + \frac{1}{2} kA^2 \sin^2 (\omega t + \delta) = \tfrac{1}{2} m\omega^2 A^2$$

is a constant throughout the motion of the oscillator.

★**12.10. Damped Oscillations.** In reality, SHM does not go on forever. The prongs of a tuning fork vibrate with steadily decreasing amplitude; the swing of a pendulum becomes smaller and smaller. The cause of this diminution of amplitude is not far to seek: it lies in the presence of frictional forces, both within the metal of the tuning fork (internal friction) and in the retarding action of the medium (air) upon the motion.

What is actually observed is the motion plotted in Fig. 12.7, where x is the ordinate, time t the abscissa. If the motion were SHM, its amplitude would always come up to the horizontal line A_0 (dotted oscillation); in reality the amplitude diminishes and the motion is confined between the converging curves $+A$

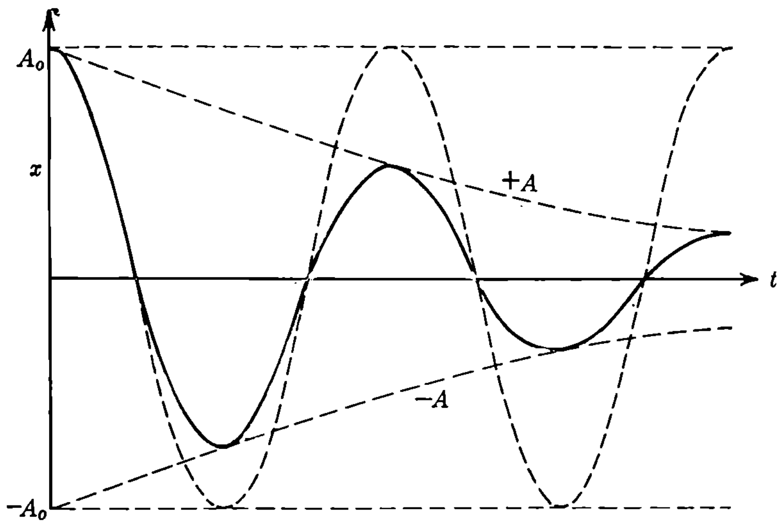


FIG. 12.7. Damped oscillatory motions.

and $-A$, its frequency remaining the same as time goes on. If we were to represent this motion mathematically, we should write

$$x = A \cos \omega t$$

where A is a continually diminishing function of unchanging sign, not a constant. The simplest function of this kind is A_0e^{-bt}, where A_0 and b are any positive constants. We might suppose on reasonable grounds, therefore, that the mathematical expression for damped oscillations is

$$x = A_0e^{-bt} \cos \omega t \tag{12.15}$$

This we shall verify.

To do so it is necessary to start with Newton's second law. There are now *two* forces acting on the mass m; first the restoring force $-kx$, which alone would give rise to SHM; second a retarding force due to friction, whose exact form must now be determined. Frictional forces usually increase as the velocity of the moving objects increases; in fluids they are in fact directly proportional to v for moderate velocities. Much experimental work has been done to make sure how the force of fluid friction depends on v, since this is a matter of very great concern in aerodynamics and in exterior ballistics. The early researches of the Russian Mayevski with rotating projectiles have shown that the force of air resistance upon elongated missiles (the shape is important in this connection) can be represented by the formula

$$F_{\text{ret}} = R_n v^n$$

where the exponent n and the corresponding constant R_n have different values in different velocity ranges. At ordinary velocities $n = 1$. However, $n = 2$ from the lowest velocity used in gunnery up to 790 ft/sec; $n = 3$ from 790 ft/sec to 970 ft/sec, and so on. The values of the corresponding R coefficients do not interest us here.

Returning then to our problem, we have found that at ordinary velocities we can write for the frictional retarding force $-R_1(dx/dt)$, the minus sign indicating that the force opposes the direction of x. Henceforth we omit the subscript 1 from R. The differential equation of motion for damped harmonic motion thus becomes

$$m\frac{d^2x}{dt^2} = -kx - R\frac{dx}{dt} \tag{12.16}$$

***12.11. Solution of Equation for Damped SHM.** To see whether Eq. (12.15) is a solution of Eq. (12.16) we substitute it in. The solution of differential equations in physics is frequently done in this way: a reasonable guess is made as to the form of the solution, and this guess is then either substantiated or disproved by substitution. If the guess is a good one, this method will also lead to a knowledge of the constants assumed to occur in the solution—in our case, A_0, b, and ω.

Proceeding in this way we find from (12.15) that

$$\frac{dx}{dt} = A_0e^{-bt}(-\omega \sin \omega t - b \cos \omega t)$$

$$\frac{d^2x}{dt^2} = A_0e^{-bt}[(b^2 - \omega^2) \cos \omega t + 2b\omega \sin \omega t]$$

On putting these results, as well as Eq. (12.15), into Eq. (12.16), we obtain, after canceling A_0e^{-bt},

$$[m(b^2 - \omega^2) - Rb + k] \cos \omega t - (R\omega - 2mb\omega) \sin \omega t = 0$$

Since this equation must be true at every instant, the coefficients of $\cos \omega t$ and of $\sin \omega t$ must individually be zero. Hence, since $R\omega = 2mb\omega$,

$$b = \frac{R}{2m} \tag{12.17}$$

Putting this value for b in the equation

$$m(b^2 - \omega^2) - Rb + k = 0$$

we find

$$\omega = \sqrt{\frac{k}{m} - \frac{R^2}{4m^2}} \tag{12.18}$$

The interpretation of Eqs. (12.17) and (12.18) is rather interesting and confirms fully our physical intuition. The greater b, the more rapidly will the curves A and $-A$ of Fig. 12.7 approach the zero axis, and Eq. (12.17) tells us that b is proportional to R, the frictional force per unit velocity. If R were zero, the motion would be SHM.

In that case ω, too, would have the value derived in Sec. 12.2, namely, $\sqrt{k/m}$. But if friction is present, the frequency is *smaller* and the period is *longer*. Friction slows down the motion, as might be expected. If the force of friction is too great, the motion will not be periodic at all. This happens when $\omega = 0$, that is, when $R = 2\sqrt{km}$. For this value of R the motion is said to be *dead beat*, or *critically damped*. For it and all larger values of R, x simply falls to zero in the manner of Fig. 12.8, without oscillating at all.

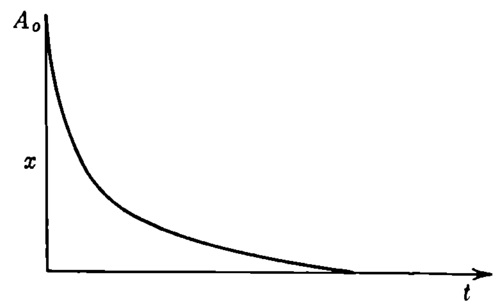


FIG. 12.8.

Finally it is to be noted that this method does not determine A_0. This, too, is significant, for it means that Newton's laws are indifferent to this constant. It may have any value and is only a measure of the initial displacement—which may indeed possess any magnitude we wish to give it.

12.12. Forced Oscillations. Thus far attention has been confined to the *natural* oscillations of an elastic system, *i.e.*, the oscillations that occur when it is displaced and then released. Under such circumstances every elastic body (tuning fork, pendulum, all metal structures) vibrates with a *natural frequency* that is wholly determined by the elastic constants and the vibrating mass. For a mass m attached to a spring of stiffness k in the presence of a friction force Rv this natural frequency is

$$\omega = \sqrt{\frac{k}{m} - \frac{R^2}{4m^2}}$$

in accordance with Eq. (12.18). In the absence of friction it is given by

$$\omega_0 = \sqrt{\frac{k}{m}}$$

A totally different situation arises when the mass is subjected to an oscillatory external force, as in the crude example of a mass attached to a spring and also to the moving end of a shuttling rod (Fig. 12.9). To be sure, we shall continue to work with this model for the sake of definiteness, but one could easily cite numerous less artificial examples. There is the bridge, which vibrates under the rhythm of marching soldiers; the prong of a tuning fork, which is exposed to the regular impulses of a sound wave; the housing of a motor, which may be subjected to the periodic beatings of an irregularity on the shaft. The oscillations that ensue are called *forced* oscillations; they proceed with the frequency of the *external force* and not with the natural frequency of the elastic body. But the response of the vibrating system depends very much on whether the externally applied frequency is nearly equal to or very different from the natural frequency of the system.

The problem here under consideration is a very general one. Its solution is also of interest in the study of alternating currents and in acoustics and will be used again in Chap. 35. Because of its generality we shall develop this problem in some detail and ask the student to endeavor to understand fully the mathematics involved.

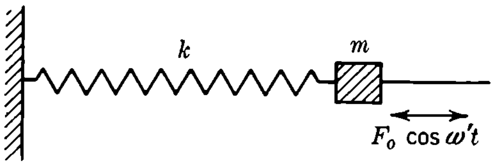


FIG. 12.9.

Let the frequency of the external force be ω', in general different from ω. The external force is then $F_0 \cos \omega' t$, and the total force, including damping, will be

$$m \frac{d^2x}{dt^2} = -R \frac{dx}{dt} - kx + F_0 \cos \omega' t \tag{12.19}$$

In solving this differential equation we allow ourselves again to be guided by what we already know about the motion.

***12.13. Solution of the Equation for Forced Oscillations.** Our knowledge amounts to this: The mass will execute SHM of constant amplitude and with the frequency of the force. There are complications—so-called *transient* motions when the force is first applied—but these soon die out and the mass attains a steady oscillation. We shall ignore these transient phenomena and confine our attention to the steady state, which usually results after a few initial vibrations. This steady state, though characterized by the same frequency ω' as the force, will *not* necessarily *be in phase* with it, nor can we say anything about the constant amplitude of its motion. What we have just stated reads, when expressed mathematically (we might also have chosen a cosine function!),

$$x = A \sin (\omega' t - \alpha) \tag{12.20}$$

From this, we obtain

$$\begin{aligned} x &= A(\sin \omega' t \cos \alpha - \cos \omega' t \sin \alpha) \\ \frac{dx}{dt} &= \omega' A(\cos \omega' t \cos \alpha + \sin \omega' t \sin \alpha) \\ \frac{d^2x}{dt^2} &= -\omega'^2 A(\sin \omega' t \cos \alpha - \cos \omega' t \sin \alpha) \end{aligned}$$

where we have expanded the sine and cosine in accordance with Eqs. (1.1) and (1.2). When these results are substituted into Eq. (12.19), we find, on collecting terms,

$$[(k - m\omega'^2)\cos\alpha + R\omega'\sin\alpha]\sin\omega't + [(m\omega'^2 - k)\sin\alpha + R\omega'\cos\alpha - (F_0/A)]\cos\omega't = 0$$

Again, this equation must be true at all times. At $t = \pi/2\omega'$, $\cos\omega't = 0$ and $\sin\omega't = 1$. Hence the first bracket must vanish. From it we get

$$\frac{\sin\alpha}{\cos\alpha} = \tan\alpha = \frac{m\omega'^2 - k}{R\omega'} \qquad (12.21)$$

At $t = 0$, $\sin\omega't = 0$ and $\cos\omega't = 1$. Hence the second bracket must vanish. It yields

$$F_0/A = (m\omega'^2 - k)\sin\alpha + R\omega'\cos\alpha \qquad (12.22)$$

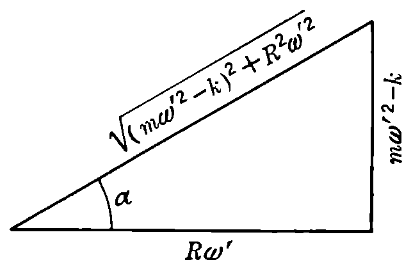


FIG. 12.10. Definition of α.

The meaning of the phase angle α is best demonstrated by means of the right triangle, Fig. 12.10. From it we also see that

$$\sin\alpha = \frac{m\omega'^2 - k}{D} \qquad \cos\alpha = \frac{R\omega'}{D}$$

provided that we denote by D the quantity

$$D = \sqrt{(m\omega'^2 - k)^2 + R^2\omega'^2} \qquad (12.23)$$

Equation (12.22) then takes the form

$$\frac{F_0}{A} = \frac{(m\omega'^2 - k)^2 + R^2\omega'^2}{D} = D$$

Hence

$$A = \frac{F_0}{D}$$

When all these results are put back into Eq. (12.20), we have, as our final result for the displacement,

$$x = \frac{F_0}{D}\sin(\omega't - \alpha) \qquad (12.24)$$

The quantity D is given by Eq. (12.23), α by Fig. 12.10.

The phase angle α depends on the values of R, ω', k, and m, that is, on the entire physical condition of the system. But, for any external frequency, it is zero when R is very large. This means that force and displacement are *out of phase* by 90° since the force is proportional to $\cos\omega't$, the displacement to $\sin\omega't$. On the other hand, if friction is absent so that $R = 0$, α is either $-\pi/2$ or $+\pi/2$. The first is true for frequencies below resonance, the latter above resonance. Hence x is proportional to $\sin(\omega't \pm \pi/2) = \pm\cos\omega't$; it is therefore in phase or out of phase by 180° with the driving force.

***12.14. Impedance, Resonance.** In the problem of forced oscillations the *velocity* of vibration is usually of greater interest than the displacement. (In a-c work it is the *current* that corresponds to v.) It is obtained at once by differentiating Eq. (12.24).

$$v = \frac{F_0\omega'}{D} \cos(\omega' t - \alpha)$$

The quantity D/ω' is called the *mechanical impedance* of the system and is usually denoted by Z. From Eq. (12.23),

$$Z = \sqrt{R^2 + \left(m\omega' - \frac{k}{\omega'}\right)^2} \qquad (12.25)$$

In view of this,

$$v = \frac{F_0}{Z} \cos(\omega' t - \alpha) \qquad (12.26)$$

We remember that α was 0 for large R. In that case the displacement was out of phase with the force; we now see that the velocity is then in phase with the force. But when R is zero, velocity of motion and force will be exactly out of phase by $\pi/2$.

The largest value of v during any period of motion is F_0/Z and is called the *velocity amplitude*. We shall consider how it varies with the impressed frequency ω'. For this purpose it is necessary to investigate the variation of Z with ω'.

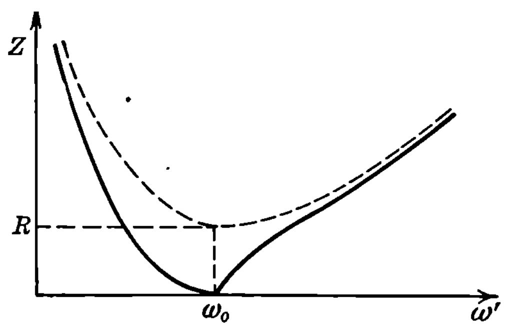


FIG. 12.11. Variation of impedance with impressed frequency.

1. Assume that $R = 0$; no friction. In that case $Z = \infty$ for $\omega' = 0$ and for $\omega' = \infty$. It takes on a zero value when $k/\omega' = m\omega'$, that is, when $\omega' = \sqrt{k/m}$ (see Fig. 12.11, solid curve). This, however, means ω' equals ω_0, the *natural frequency* of the system. When this equality holds, the velocity becomes infinite and what is called the "resonance catastrophe" occurs. It is indeed a dangerous condition, and the designer of moving machinery and the airplane engineer have to be very careful never to allow driving frequencies that are close to natural frequencies of structural parts. The phenomenon of resonance in general is the excitation of large amplitudes of motion or velocity by means of small periodic forces applied at the proper frequency. It is said that bridges may collapse under the feet of marching soldiers unless their step is broken. An experience familiar to most readers will be the appearance of rattles in an automobile at definite driving speeds.

2. Assume that R is finite; friction is present. Even if friction is present, resonance occurs. But the resonance catastrophe is now reduced to limited proportions; v has a maximum but remains finite. The value of ω' corresponding

to this maximum is the one that minimizes Z and is still given by $\omega' = \omega_0$, as differentiation of Z and equating to zero will show. Hence resonance does not take place at the natural frequency of the free motion, which is given by Eq. (12.18) when friction is present. It happens when the impressed frequency equals the natural frequency of the *undamped* motion, given by $\sqrt{k/m}$. The

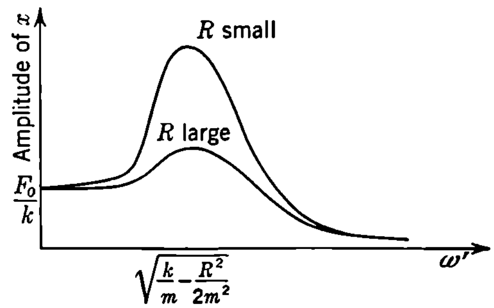


FIG. 12.12. Amplitude of forced oscillations as function of impressed frequency.

dotted curve in Fig. 12.11 represents Z as a function of ω' when friction is present.

At resonance $Z = R$; and Fig. 12.10 shows that $\alpha = 0$. Hence

$$v_{\text{res}} = \frac{F_0}{R} \cos \omega' t$$

In Fig. 12.12 is plotted the *displacement* amplitude $A = F_0/D$ as a function of ω'. The smaller R, the steeper will be the maximum at resonance. The maxi-

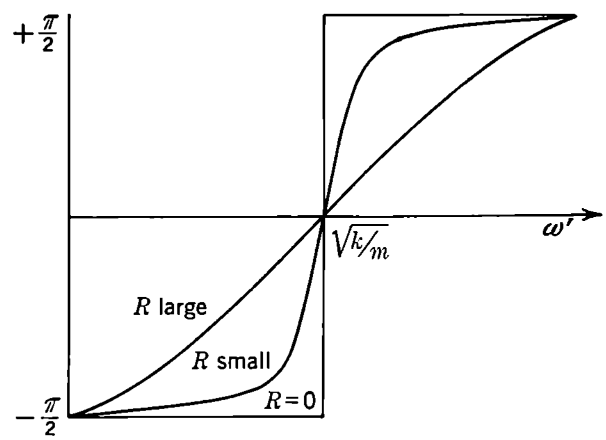


FIG. 12.13. Phase angle as a function of impressed frequency.

mum of A does not come at ω_0, nor at ω; the minimum of D occurs at

$$\sqrt{\frac{k}{m} - \frac{R^2}{2m^2}}$$

Fig. 12.13 shows how the phase angle α varies as ω' goes from 0 to ∞. It is zero at resonance, where $\omega' = \omega_0$.

PROBLEMS

1. Differentiate Eq. (12.6) with respect to t to get v; differentiate once more to get the acceleration a. Take $A = 1$ unit, $\omega = \pi$ sec^{-1}, $\delta = \pi/4$, and plot x, v, and a against t. Observe that $a = -\omega^2 x$ at every t.

2. Compute v from Eq. (12.6) by differentiation, and show that the total energy, that is, $\frac{1}{2}mv^2 + \frac{1}{2}kx^2$, is constant.

3. Prove that the potential energy of a particle in SHM at the extreme point of the motion equals the kinetic energy at its passage through the neutral point.

★4. Show that the equation $u = Ae^{\pm i\omega t}$ satisfies the differential equation for SHM ($i = \sqrt{-1}$).

5. A mass of 10 gm executes SHM on the end of a spring completing 2 oscillations every second. The amplitude of the motion is 8 cm. (*a*) What is the velocity of the mass as it passes through the neutral point? (*b*) What is the acceleration of the mass at the extremity of its motion? (*c*) Find the stiffness k of the spring. (*d*) What is the total energy of the vibrating mass?

6. A mass of 4 lb is free to execute SHM as shown in Fig. 12.1. When $t = 1$ sec, the mass is projected at $x = 0$ with a velocity $v = +20$ ft/sec. The stiffness of the spring is 64 poundals/ft. Find the equation of motion of the mass, *i.e.*, determine A and δ for Eq. (12.6).

7. Show that the period of a mass executing linear SHM is given by

$$P = 2\pi\sqrt{-x/a},$$

where x is the displacement of the mass at any time and a is the corresponding acceleration.

8. An unknown mass, when hung on a light spiral spring, stretches it 20 cm. Determine the period of oscillation of the mass if it is displaced from its neutral position and then released.

9. A particle of mass m, free to move in the xy plane, is subject to a force whose components are $F_x = -kx$, $F_y = -ky$, where k is a constant. The particle is released when $t = 0$ at the point (2,3). Show that its subsequent motion is SHM along the line $2y = 3x$.

10. A horizontal surface moves up and down in SHM with an amplitude of 3 cm. Find the maximum frequency that this motion may have if a mass resting on the surface is to remain continuously in contact with it.

11. At a certain location on the earth a simple pendulum, 1 m in length, makes 100 complete oscillations in 205 sec. What is the value of the acceleration of gravity at this point?

12. What must be the length of a simple pendulum, if it is to beat seconds (*i.e.*, execute 1 complete oscillation in 2 sec) at a point on the earth where the acceleration of gravity is 979 cm/sec^2?

13. A circular hoop of radius 2 ft and mass 7 lb is suspended on a horizontal nail. Find the frequency with which it oscillates if displaced slightly from its neutral position. What would be the length of a simple pendulum with the same frequency?

14. A solid sphere, 15 cm in radius and weighing 2 kg, is suspended on a wire as shown in Fig 12.14. The torque required to twist the wire is 5×10^4 dyne cm/radian. If the sphere is twisted slightly from its neutral position, what will be its period of angular oscillation?

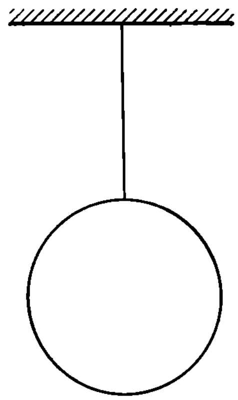


FIG. 12.14. Problem 14.

15. A rectangular bar, 1 by 3 by 10 cm and weighing 50 gm, is suspended horizontally at its mid-point by a vertical wire. When twisted slightly and then released the bar oscillates with a period of 0.5 sec. Find the torsion constant of the wire.

16. A mass of 8 gm executes SHM along the x axis under the action of a restoring force equal to 10^4 dynes/cm. Assuming the motion to be undamped, find the frequency of oscillation of the mass. Now assume a damping force of 10^2 dynes per (cm/sec), and again find the frequency. What value of R would give critically damped motion for this system?

***17.** A mass undergoes damped SHM given by the equation $y = A_0e^{-bt} \sin \omega t$. If $A_0 = 10$ cm, $\omega = \pi\ \text{sec}^{-1}$, and $b = 0.5\ \text{sec}^{-1}$, plot y against t from $t = 0$ to $t = 10$.

***18.** A mass of 2 gm, free to execute linear SHM, is subject to a restoring force of -10^3 dyne/cm and a damping force -10^2 dyne per (cm/sec). If a driving force $F = 10^4 \cos (10\pi t)$ dynes is applied, calculate the amplitude of the motion and the phase angle α.

***19.** Calculate the mechanical impedance of the mass in Prob. 18. Also, find its resonance frequency.

20. The path of a particle moving in the xy plane is given by the following parametric equations: $x = A \cos \omega t$; $y = A \sin \omega t$. Show that the motion is in a circle of radius A with its center at the origin.

Now assume that a mass m is subject to a force whose components are $F_x = -kx$ and $F_y = -ky$. Let the mass be set moving at the point $(A,0)$, with a velocity $\sqrt{(k/m)}\, A$ in the positive y direction. Show that the motion described above results. (Consider the x and y components of the motion independently.)

CHAPTER 13
HYDROSTATICS

13.1. Nature and Properties of Fluids. Hydrostatics is the branch of physics that deals with nonsolid substances at rest (Greek *hydor*, water; *statikos*, causing to stand). Nonsolids are often called *fluids* (Latin *fluere*, to flow); they include liquids, both viscous and nonviscous, as well as gases, even though the popular use of the term excludes the latter. Liquids and gases have so many things in common that they require in many respects the same scientific treatment.

Like solids, fluids consist of molecules. Democritus, a philosopher of ancient Greece, thought fluid molecules were smooth while those of solids were rough and equipped with interlocking mechanisms. Today we know that the same molecules can, under different physical conditions, compose a solid or a liquid or a gas.

Molecules attract one another by forces that are different for the constituents of different substances. In a certain range of pressures and temperatures these forces are able to hold the molecules in definite positions, permitting them only enough freedom to vibrate about these places of equilibrium. The molecules then form a solid, which is by far the simplest kind of aggregate to understand. As the temperature rises (and for a corresponding pressure change), the energy of vibration of the solid molecules increases until the forces can no longer hold them in definite positions; the solid melts or sublimes (turns into vapor). Usually it melts. The molecules are now in a state of chaos; they move freely and yet have latent bonds that maintain a certain average distance between them. Hence the liquid has a fixed volume.

Just why liquids should exist, just why molecules should "hang onto" one another in a manner to preserve a constant volume, instead of moving apart as they do in a gas, is a puzzle frequently overlooked by the beginner. In some respects the existence of liquids is still a mystery of nature, and a completely satisfactory theory of liquids has not yet been evolved. Nevertheless their simpler aspects are easily described.

In some instances the forces are such as to cause the motions of molecules in a liquid to proceed with difficulty. The liquid is then called *viscous*. A truly nonviscous liquid, of which there is no actual example, is called an *ideal* liquid. It has recently been discovered that helium becomes an almost ideal liquid at extremely low temperatures.

The nature of a gas is much simpler to understand. The molecules have gained enough freedom to move wherever they like. They collide, rebound from the walls containing them, and fill the whole volume that is at their disposal.

From the point of view of deformations, fluids differ markedly from solids: they do not resist a shearing stress. When they flow, a finite shear is produced by a zero stress. Hence we may say that their shear modulus is zero.

13.2. Density. Density is mass per unit volume. It is measured in pounds per cubic inch, grams per cubic centimeter, and so forth. The density of liquids is nearly uniform; it does not change from one point to another. In a large body of gas such as our atmosphere, this is not true.

Density is sometimes expressed in another way by stating how many times as dense a substance is as water. This ratio, the density of the substance in question divided by the density of water, is called *specific gravity*. It is a pure number and has no units. Since the density of water is 1 gm/cm³, the specific gravity of all substances is equal, *numerically*, to its density in gm/cm³. Notice that

$$1 \text{ gm/cm}^3 = \frac{1/454 \text{ lb}}{(1/30.5)^3 \text{ ft}^3} = 62.4 \text{ lb/ft}^3$$

Table 13.1. *Densities of Various Substances*

Substance	Density, gm/cm³	Substance	Density, gm/cm³
Aluminum	2.7	Iron	
Brass	8.1–8.7	Cast	7.1–7.7
Copper	8.9	Steel	7.8
Glass, common	2.4–2.6	Lead	11.3
Gold	19.3	Platinum	21.4
Ice (0°C)	0.9167	Silver	10.5
		Tin	7.3
		Mercury	13.6

13.3. Pressure. Fluids exert forces on all objects immersed in them. To understand why this should be we consider the motion of the molecules. Much more will be said about molecular motion later; here the following facts are of importance: The motion is *random* in a double sense: there are just as many molecules moving in one direction as in any other, and the velocity differs in unpredictable fashion from one molecule to the next. Naturally, the molecules collide frequently with one another and with the walls of the vessel containing them, and during impact both direction and speed are usually altered.

Now suppose an external object to be immersed in the liquid. It will be bombarded by molecules from all sides and in all directions. The larger its surface, the greater will be the number of molecules that rebound from it. This molecular bombardment produces a *force* on the surface of the body. Clearly, its properties must be these:

1. *The force is perpendicular to all immersed surfaces.* This is because every surface suffers a great number of impacts from all directions, and the tangential effects of these impacts cancel each other.

2. *The force is the same whether the surface is inclined at one angle or at any other.* This again is due to the fact that molecules move in all directions.

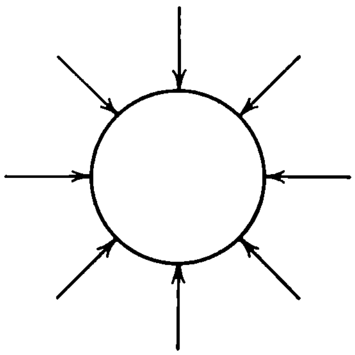


FIG. 13.1. Force on a sphere.

As a result of this the forces on the six sides of a cube are equal; the force on a sphere (cf. Fig. 13.1) is uniform all around and merely tends to compress it; the force on the left-hand side of the object in Fig. 13.2 is greater than that on the right-hand side.

Simplest account is rendered of these facts if we focus attention on *perpendicular force per unit area*, rather than on the force itself. The name for "normal force per unit area" is *pressure;* the symbol we shall use for it is p,

$$p = \frac{F_\perp}{A} \tag{13.1}$$

Pressure is indiscriminate with regard to direction and is therefore not usually regarded as a vector. Henceforth it will be understood that the force involved in p is normal to A, and the subscript $\perp$ on F will not be written.

In general p, while it is the same in all directions at a given point in a fluid, may vary from one point to another. In that case, Eq. (13.1) must be replaced by the more general

$$p = \frac{dF}{dA} \tag{13.2}$$

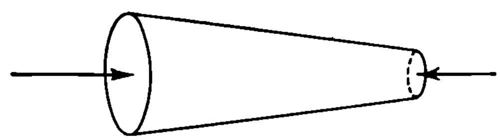


FIG. 13.2. The same pressure produces different forces.

Obvious pressure units are the *dyne per square centimeter* and *the pound per square inch.* One million dynes per square centimeter are called a *bar;* 10^3 dyne/cm^2 a *millibar.*

13.4. Pressure and Density. To go on with our study, it is well for the present to dismiss from consideration the molecular aspects of pressure, which served to explain the aforementioned facts, and to think of pressure in its more ordinary sense as a force per unit area. It may be traced to two physical causes. (1) The lower portions of a fluid must somehow carry the higher ones, and this burden makes itself felt as an added pressure below. (2) The surface of the fluid itself may be subject to pressure, and this pressure transmits itself to the interior of the fluid. A case in point is the pressure of the atmosphere, which acts on all bodies of water exposed to the atmosphere.

We shall first ignore the second cause and determine how p varies from point to point in a fluid as a result of its own weight. Typical examples of this variation are found in the atmosphere, where p decreases with height, and in the ocean, where it increases with the depth. But the rate of variation is quite different in these two instances; the difference is due to the fact that water is nearly incompressible, while air is a compressible fluid.

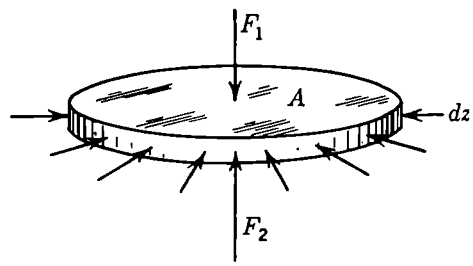


Fig. 13.3. A fluid in equilibrium.

There is a general relation between p and the density, ρ. From it one can derive by integration specific formulas for p as a function of height or depth. We first find this general relation.

The flat cylinder shown in Fig. 13.3 has a height dz, and the area of its top and bottom faces is A. It is meant to represent a mathematical surface inside a fluid of density ρ. The fluid inside it is in equilibrium under the forces exerted upon the cylinder by the surrounding fluid.

Since there is no lateral acceleration, all lateral forces (small arrows in Fig. 13.3) must have a zero resultant. This can be true only if the pressure is the same everywhere around the cylinder. Hence we may conclude that the pressure is the same within the horizontal segment occupied by the cylinder and that p can be only a function of z, the vertical coordinate.

But the vertical forces on the cylinder must also add up to zero. They are the force of gravity mg downward (not drawn in Fig. 13.3), and the "pressure forces" F_1 and F_2. But $m = \rho A\,dz$. Therefore

$$F_1 + \rho g A\,dz = F_2 \tag{13.3}$$

If now we take the pressure acting upon the upper face to be p, that on the lower face will be $p + dp$. Then

$$F_1 = Ap, \qquad F_2 = A(p + dp)$$

On introducing this in Eq. (13.3), canceling, and rearranging we find

$$\frac{dp}{dz} = g\rho$$

for the rate of change of pressure with distance. If we follow custom and take the positive direction of z upward, a minus sign must be introduced in the preceding equation, for p increases as z decreases. Hence we write

$$\frac{dp}{dz} = -g\rho \tag{13.4}$$

To integrate this equation, ρ has to be known as a function either of p or of z. This requires further study.

13.5. Pressure in Liquids. Liquids are very nearly incompressible (see the values of their bulk moduli in Table 4.1) and therefore have an approximately constant density. Integrating Eq. (13.4) for constant ρ we obtain

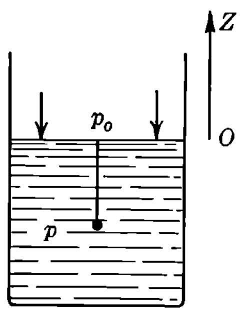


FIG. 13.4. Pressure in a liquid.

$$p = -g\rho z + \text{const}$$

The origin is conveniently chosen somewhere in the surface of the liquid. The constant is then the value of p at the surface, which will be called p_0. (Cf. Fig. 13.4.) Our conclusion may therefore be stated:

$$p = p_0 - g\rho z \tag{13.5}$$

Ordinarily p_0 is the pressure due to the atmosphere.

Without specifically asking the question how external pressures are transmitted through a liquid, we have found the answer to it as well. For the result shows that p_0 is simply added on, at every point of the liquid, to $-g\rho z$, which is the pressure due to the liquid's own weight. This fairly obvious fact was first discovered by Pascal (1623–1662) and is known as his principle. Stated *in extenso* it says: *External pressure is transmitted equally to all parts of a fluid.* An important application of it will be given in Sec. 13.8.

13.6. Manometers. Instruments for measuring pressure are called *manometers* (Greek *manos*, rare, vacuous). The simplest type is a tube, open at both ends, and partly filled with a liquid, usually mercury or water (Fig. 13.5). Assume the left opening to be connected with a space in which the pressure is p; the other end is exposed to atmospheric pressure p_0. Since the points A and B have the same vertical height, they are at the same pressure. Hence $p = p_0 + \rho g h$. The distance h is usually measured on a scale S placed alongside the tube.

Figure 13.5 shows the instrument measuring a pressure greater than p_0. If $p < p_0$, the height of the liquid is higher in the left arm than in the right and h in the formula is a negative quantity.

The usual *mercury barometer* (Greek *baros*, weight [of the air]) is a closed-tube manometer, shown in Fig. 13.6. A vertical glass tube, closed at one end, is first completely filled with mercury. Then, while the open end is held closed by one finger, this end is dipped into a pool of mercury and then opened. The mercury column will drop, leaving a vacuum at the top to a height h above the level in the pool. Clearly, if ρ is the density of the mercury,

$$p_0 - h\rho g = 0$$

or

$$p_0 = h\rho g$$

The height h is found to be approximately 76 cm, and it varies in a well-known manner with the weather. The density of mercury is 13.60 gm/cm³ at 32°F; hence the average atmospheric pressure is

$$p_0 = 76 \text{ cm} \times 13.60 \text{ gm/cm}^3 \times 980 \text{ cm/sec}^2 = 1.013 \times 10^6 \text{ dynes/cm}^2$$

or approximately 1 *bar*. The Weather Bureau expresses variations in atmospheric pressure in millibars.

The "centimeter of mercury" is a pressure unit frequently used in physics. It is the pressure exerted by a column of mercury 1 cm. high and equals one-seventy-sixth of an "atmosphere" (another common unit), or 13.60×980 dynes/cm² = 13.3 millibars. The common *gravitational* unit of pressure is the lb/in.² Gravitational units are especially useful in measuring pressures, for the pressure, in lb/in.², is simply the weight of a vertical fluid column 1 in.² in cross section above the point where the pressure is sought. Also, the formula for pressure in gravitational units is simply

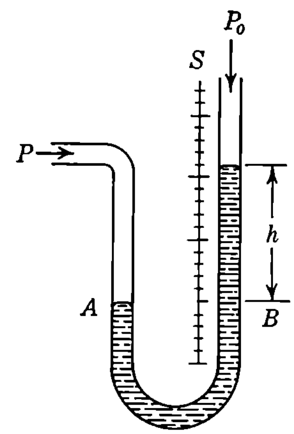


FIG. 13.5. Open-tube manometer.

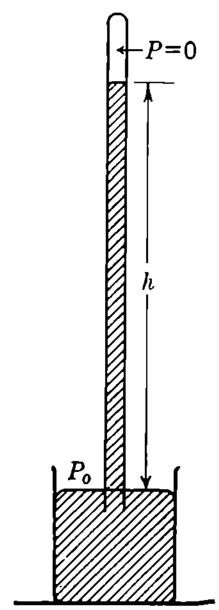


FIG. 13.6. Mercury barometer.

$$p = h\rho$$

instead of $p = h\rho g$.

The *aneroid* barometer (Greek *a*, not, + *neros*, wet, *i.e.*, dry) is a favorite instrument for home use. It is an exhausted cylindrical metal box, the top of which is slightly deformed when the atmospheric pressure changes. A mechanism magnifies the deformations and translates them into deflections of a pointer, which can be read on a scale. To increase the amount of the motion, the box is corrugated.

To measure high pressures the Bourdon pressure gauge is often used. It consists of a bent metal tube of oblong cross section, one end of which is open and exposed to pressure, the other being closed (Fig. 13.7). Pressure causes the tube to straighten out as shown by the dotted outline, and this motion is conveyed to a pointer on a scale. Such gauges are usually calibrated to read pressures *above atmospheric pressure.*

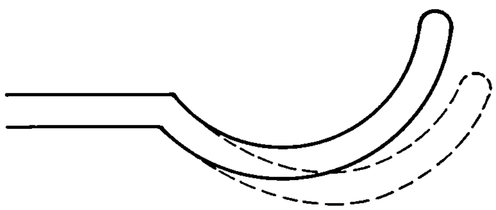


FIG. 13.7. Bourdon gauge.

The highest pressures thus far produced and measured are of the order of 10,000 atm.

13.7. Hydraulic Press. Figure 13.8 illustrates the action of a hydraulic press. Two cylinders, one of cross section a, the other A, are hydraulically connected. If a force F_a is exerted on the smaller piston, a pressure $p = F_a/a$ is produced and transmitted throughout the liquid. The upward force on the larger piston is then pA; hence

$$F_A = \frac{A}{a} F_a$$

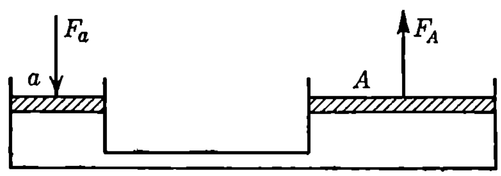


FIG. 13.8. Hydraulic press.

Hydraulic brakes operate in the same way; a is the area of the plunger, and A that of the brake shoe. To reduce evaporation losses a liquid of low vapor pressure is chosen.

13.8. Archimedes' Principle. The circumstances surrounding the discovery of Archimedes' (287–212 B.C.) principle, while somewhat legendary, are memorable indeed; Archimedes' joy over his discovery, his leaving the bath, presumably unclad, and exclaiming *heureka* (I have found it) form one of the more delightful scientific anecdotes. Yet the principle itself is almost obvious.

Consider any closed mathematical surface drawn inside a fluid at rest. The fluid within is in equilibrium. Hence the forces exerted upon it by the fluid portions external to the surface must support it; *i.e.*, the resultant of all external forces must be equal to the *weight* of the fluid within the surface.

But these external forces are in no way affected by what is contained within the surface. If the latter were filled with another substance, or a solid body, the same forces would act on it. They will support it and cause it to float when the density of the solid is smaller than that of the fluid; otherwise the solid will sink, but its weight is diminished by the weight of the liquid that it displaces. Hence Archimedes' principle: *A body immersed in a fluid is buoyed up by a force equal to the weight of the fluid that it displaces.*

An important extension of the principle, not recognized by its Greek discoverer, has to do with the exact location of the buoyant force. Since the latter equilibrates the force of gravity on the fluid filling the cavity before its displacement, the *force of buoyancy acts through the center of gravity of the displaced fluid.* This point is the geometrical center of the immersed portion of the body and is called the *center of buoyancy.*

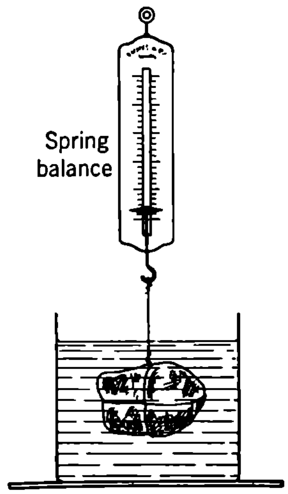


FIG. 13.9. Determining densities.

13.9. Uses of Archimedes' Principle. Archimedes discovered, along with his principle, a great number of ways in which it could be employed to advantage (look up "Hiero's crown" in any history of science). Later the Arabs (A.D. 500–1000), who were traders in precious stones and therefore required methods for determining the density of gems, perfected these procedures.

To find the specific gravity of an irregular object it is first weighed in air and found to have a weight w. It is then lowered into a container of water (cf. Fig. 13.9) and is found to have an apparent weight w'. The specific gravity of the object is then given by

$$\text{Sp gr} = \frac{w}{w - w'}$$

where $w - w'$, the "loss of weight in water," is equal by Archimedes' principle to the weight of water displaced.

If the body floats on the water, the same operations may be performed with an added weight below the body; and weighing takes place first with the weight alone immersed in the water, then with both body and weight submerged.

To determine the density of a liquid one takes a body of mass m and determines its apparent weight in a liquid of known density ρ_0 such as water. Let its apparent weight be m_0g. Next the apparent weight is determined after submersion in the liquid of unknown density and found to be $m'g$. We then write

$$(m - m_0)g = \rho_0 Vg$$

and

$$(m - m')g = \rho Vg$$

On dividing one equation by the other we see that

$$\frac{\rho}{\rho_0} = \frac{m - m'}{m - m_0}$$

The volume of the body, V, drops out and need not be determined.

A simpler though less accurate way to find the density of liquids is to measure the depth of immersion of a floating body (hydrometer principle). The cylinder depicted in Fig. 13.10 sinks to a depth a in a liquid of known density ρ_0. In the liquid of unknown density it sinks to a depth x. Let both a and x be measured from the base of the cylinder, and let the weight of the cylinder be mg, its area of cross section A. Then

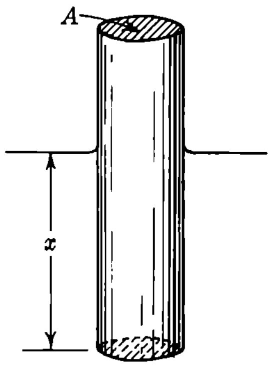


FIG. 13.10. The hydrometer.

$$ma = Aa\rho_0 g = Ax\rho g$$

and

$$\rho/\rho_0 = a/x$$

The hydrometer used in determining the density of the radiator fluid in automobiles works on this principle.

Gases, too, exert buoyant forces, as is clear from the fact that balloons rise. The density of the air, 0.00123 gm/cm^3, is so small that in the measurements discussed above it may be neglected.

***13.10. The Atmosphere.** Air is not incompressible, and it therefore is not proper to set ρ in Eq. (13.4) equal to a constant when reference is made to the atmosphere. The density is in general a function of the pressure itself, though the precise variation of ρ with p is not easy to specify and depends upon the way in which the temperature changes as we proceed upward in the atmosphere. A rough idea of the actual situation can be formed if we assume the temperature to be constant, which is, of course, not true. In that case, ρ is proportional to p, and one may write

$$\frac{\rho}{\rho_0} = \frac{p}{p_0}$$

if ρ_0 and p_0 are the known values of density and pressure at sea level. On inserting this relation in Eq. (13.4) we obtain

$$\frac{dp}{dz} = -g\rho_0 \frac{p}{p_0}$$

or

$$\frac{dp}{p} = -\frac{g\rho_0}{p_0}\,dz$$

When this is integrated, there results

$$\ln p = -\frac{g\rho_0}{p_0} z + \text{const}$$

The constant is $\ln p_0$, as may be seen by letting z be zero.

Hence
$$p = p_0 e^{-g(\rho_0/p_0)z}$$
Now
$$g\frac{\rho_0}{p_0} = \frac{980 \text{ cm/sec}^2 \times 0.00123 \text{ gm/cm}^3}{1.01 \times 10^6 \text{ gm cm}^{-1} \text{ sec}^{-2}} = 1.16 \times 10^{-6} \text{ cm}^{-1} = 0.116 \text{ km}^{-1}$$
Therefore
$$P = P_0 e^{-0.116z}$$

provided that z is measured in kilometers. According to Eq. (13.6) the pressure decreases by about 11 per cent, *i.e.*, from 76 cm to 68 cm of mercury, as we go 1 km aloft. This is not far from true. Table 13.2 shows some measured values of pressure and temperature in the earth's atmosphere.

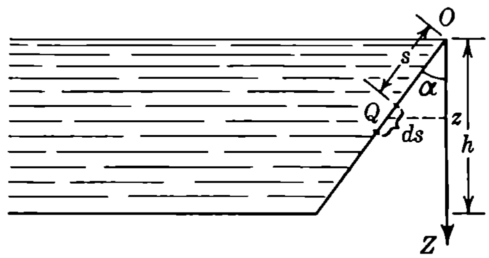


FIG. 13.11. Force on a dam.

***13.11. Force of Water on a Dam.** A typical engineering problem is to calculate the force of water on a dam. Assume the side of the dam to be inclined at an angle α with the vertical, as in Fig. 13.11, which shows a section of the dam. Take the origin in the upper surface of the water (of density ρ)

Table 13.2. *Variation of Pressure and Temperature at Different Heights in the Atmosphere*

Pressure, mm Hg	Height above sea level, km	Temp, °C
762	0	14.8
677	1	13.6
600	2	6.3
530	3	1.0
468	4	− 4.3
412	5	− 9.9
206	10	−44.9
177	11	−50.1*
152	12	−52.9
130	13	−52.8
116	14	−52.4
96	15	−52.0
82	20	−49.0

* Note near constancy of temperature in stratosphere, which begins here.

and measure z downward. The wall has a width l at right angles to the plane of the paper, and the distance of the point Q from O is s. Then the force dF on a strip of width ds and length l at Q is $\rho gzl\,ds$. But $s = z/\cos\alpha$; hence

$$dF = \frac{\rho gl}{\cos\alpha}\, z\, dz$$

We integrate from $z = 0$ to $z = h$ and find

$$F = \frac{\rho gl}{\cos\alpha}\int_0^h z\,dz = \frac{\rho gl}{\cos\alpha}\,\frac{h^2}{2}$$

But ρgh is p_b, the pressure at the bottom; $(hl)/\cos\alpha = A$, the area of the dam. The force on the dam is therefore

$$F = \tfrac{1}{2}p_bA \tag{13.7}$$

Imagine now a canal of triangular cross section, like that shown in Fig. 13.12, and of length l. The weight of the water in the canal is

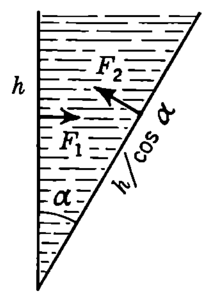


FIG. 13.12. Hydrostatic paradox.

$$W = \frac{h\tan\alpha\, h}{2}\, l\rho g = \frac{1}{2}\, p_b hl\tan\alpha$$

since the horizontal side of the triangle is $h\tan\alpha$. In terms of W the force F on the slanting wall, given by Eq. (13.7), becomes

$$F = \frac{1}{2}\cdot\frac{2W}{hl\tan\alpha}A = \frac{W}{\sin\alpha}$$

This means that the force on the dam is actually *greater* than the weight of the water which produces it. This is sometimes called the "hydrostatic paradox."

However, there is nothing unreasonable about this situation. For let us look again at Fig. 13.12. According to Newton's third law the forces acting on the body of water are equal and opposite to those exerted by the water on the walls. The water is therefore subject to three forces,

$$F_1 = \frac{1}{2}\, p_b hl \qquad F_2 = \frac{1}{2}\, p_b\frac{hl}{\cos\alpha} \qquad W = \frac{1}{2}\, p_b hl\tan\alpha \tag{13.8}$$

The last of these acts vertically downward and is not indicated in the figure. If the body is in equilibrium, the sum of all horizontal components must vanish, and hence

$$F_1 = F_2\cos\alpha$$

This relation is satisfied by Eq. (13.8). Resolving along the vertical we must have

$$W = F_2\sin\alpha$$

and this is also true. The fact that F_2 is greater than W is occasioned by the presence of another horizontal force, F_1.

★13.12. Flotation. It is not enough to build a ship that floats; the naval architect must also be able to predict in what position it will float, and float stably. The ship drawn in Fig. 13.13, when slightly listed, returns to its normal position; the plank of Fig. 13.14, after a slight displacement, departs farther from the vertical and floats on its larger face. Let us see why this should be.

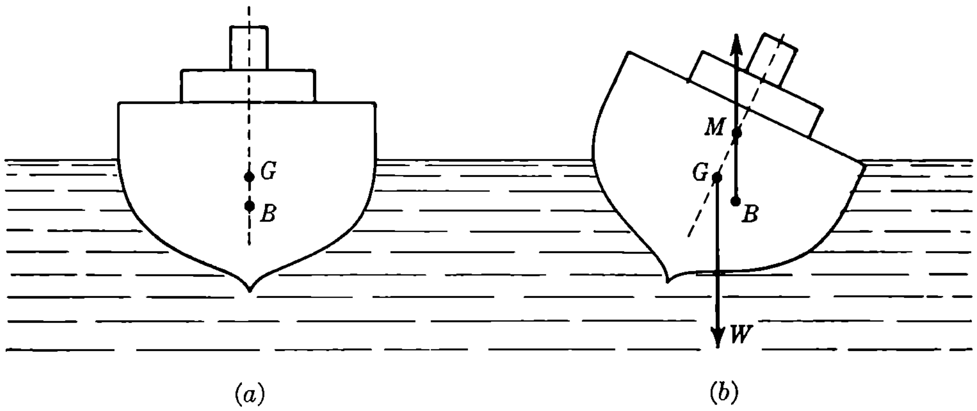


FIG. 13.13. Stable flotation.

In its normal position, two forces act on the ship (Fig. 13.13*a*)—gravity downward at the center of gravity *G*, buoyancy upward through the center of buoyancy *B* (center of gravity of the displaced water). They are equal and opposite: a ship's weight is equal to its water displacement.

Now let the ship be tipped as in Fig. 13.13*b*. The point *B* is *not* a fixed point in the body of the ship; it moves to a new position. The two forces now form a torque, and this torque restores the original position. The torque on the plank in Fig. 13.13*b* does the opposite: it urges the body farther from its original position.

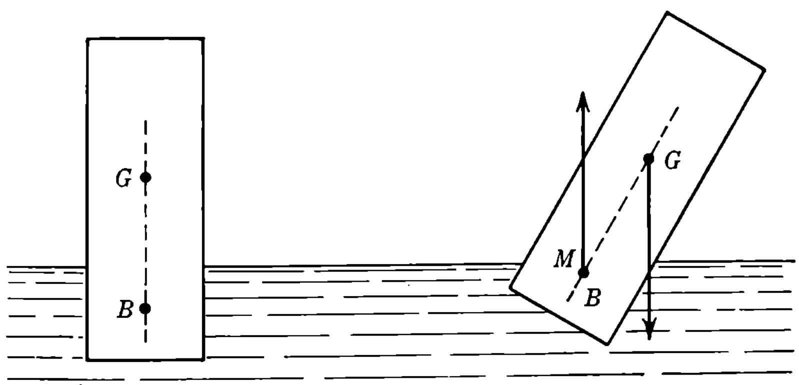


FIG. 13.14. Unstable flotation.

Whether a body floats stably depends on the location of a certain point in the body, called the *metacenter*. It is a fixed point through which the force of buoyancy acts in both the normal and the tipped position. In Fig. 13.13*a* the force acts along the dotted line, in Fig. 13.13*b* along the vertical arrow through *B*. These intersect at the point called *M*, and this is the metacenter.

Flotation is stable when M lies above G, unstable when M lies below G (which is the case in Fig. 13.14).

The distance between M and G is called the "metacentric height." If it is large, the boat is stable but stiff; if small, the boat rides gently but is unsafe. Shipbuilders strike a compromise between these extremes.

More generally the metacenter of a floating object is the center of curvature of the curve on which B moves as the object is tipped slightly from its normal position. A body may have several stable positions for flotation and hence several metacenters.

PROBLEMS

1. Find the pressure, both in lb/in.^{2} and in dynes/cm^2, 500 ft below the surface of the ocean. (Specific gravity of sea water is 1.03.)

★2. Compute the pressure in the atmosphere 25,000 ft above sea level.

3. What is the force on a diver's outfit, 100 ft below the surface of a lake, if its surface is 40 ft^2?

4. A swimming pool is 50 by 100 ft. The bottom slopes from a depth of 4 ft to 12 ft from one end to the other. Supposing the pool to be completely filled, find (*a*) the normal force on the bottom; (*b*) the vertical force on the bottom; (*c*) the force on the four sides.

★5. Calculate the moment of force exerted by the water on the vertical wall at the deep end of the pool in Prob. 4, (*a*) about the upper edge; (*b*) about the lower edge.

6. A cylindrical cask of height 5 ft and diameter 3 ft stands upright. A vertical 1-in. pipe, 50 ft long, is attached to its upper surface. What is the force on the wall of the cask when both cask and pipe are filled with water?

7. An automobile lift, operated as a "hydraulic" press employing air pressure of 100 lb/in.^{2}, has a plunger 10 in. in diameter. What weight can it lift?

8. Water pressure in the city mains is 50 lb/in.^{2} If a hydraulic elevator has a piston of diameter 1 ft and an efficiency of 75 per cent, what load can it raise?

9. An iceberg, floating in sea water (density 1.03 gm/cm^3), has $\frac{9}{10}$ of its volume below the water surface. Find the density of ice.

10. A block of wood floats in water with $\frac{2}{3}$ of its volume submerged; in oil it has $\frac{9}{10}$ of its volume submerged. Find the density of the wood and the oil.

11. A piece of cork (sp gr = 0.25) weighs 20 gm. A piece of metal weighing 100 gm is fastened to the cork, and the system is weighed in water. Its apparent weight is found to be 20 gm. What is the density of the metal?

12. A piece of metal weighs 250 gm in air, 210 gm in water, 225 gm in oil. Find the densities of metal and oil.

13. What is the buoyancy of air upon a brass weight of 1 kg?

14. Look up the story of Hiero's crown (cf. Encyclopaedia Britannica). In your opinion, what method did Archimedes employ in determining the composition of the crown?

CHAPTER 14
HYDRODYNAMICS

14.1. Streamline Flow and Turbulence. Hydrodynamics (Greek *hydor*, water; *dynamis*, force) is the study of the motion of fluids under forces; it is a field of great complexity, yet of unusual beauty in its mathematical structure. Physicists and engineers are at present devoting to it much enthusiasm and painstaking effort, for from its further development will come an understanding of the conditions under which flight at very high speeds is possible. We give here only the barest outline of the subject.

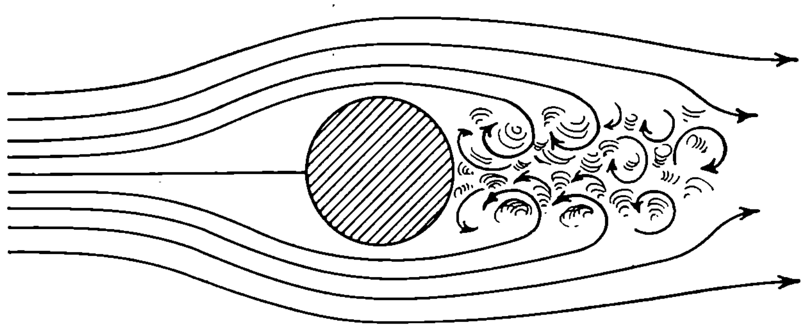


FIG. 14.1. Stream line flow and turbulent flow.

A liquid or gas can move in two fundamentally different ways. When a liquid flows slowly through a pipe, the motion of its molecules is orderly; all molecules passing a given point inside the pipe proceed with the same velocity; the motion is everywhere wholly predictable. On the other hand, when air rushes past an obstacle in a wind tunnel, a churning of the air masses behind the obstacle takes place, numerous eddies are formed, and the motion of the air molecules is unpredictable in detail. The first kind of motion is called *streamline* flow, the second *turbulence.*

Interest in turbulence arises from the disadvantages and dangers that attend its presence. Whenever a body moves through a fluid and causes turbulence in its wake, the resistance to its motion is very great. For this reason airplanes, cars, and locomotives are streamlined, *i.e.*, are given a shape that will prevent turbulence (cf. Fig. 14.1). (To discover the reason for streamlining of fountain pens and ice boxes is left as an exercise for the reader.) The sudden onset of turbulence caused by icing of wings and other circumstances has thrown many an airplane out of control. The eddies occasionally trailing the wing tips of fast airplanes are a familiar sight to many pilots; they are signs of turbulence.

Turbulence may occur even in fluids streaming through pipes. In this instance it occurs fairly regularly where the fluid moves with a velocity greater than what is called the critical velocity v_c. *Reynolds* discovered that, for a given fluid at a certain pressure and temperature,

$$v_c = \frac{C}{d} \tag{14.1}$$

where C is a constant peculiar to the fluid and d the diameter of the pipe. For water, $C = 0.0384\ \text{ft}^2/\text{sec}$. Thus turbulence occurs at higher speeds in narrow pipes than in wide.

Oil is sometimes said to calm a turbulent sea. This, while true in a sense, is a rather remarkable misstatement. When an oil film covers the water surface, the wind, tearing at the oil molecules, sets up small eddies, or vortices, in the water below the oil film and these irregular eddies prevent the formation of orderly waves and their growth, which would occur in the absence of the oil.

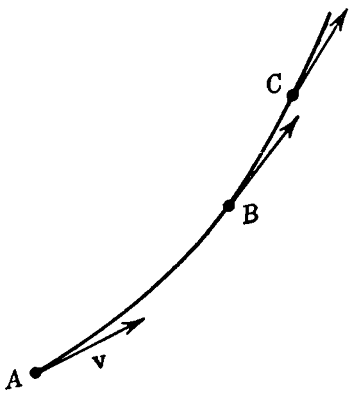


FIG. 14.2. A streamline.

Mathematically, the motion of a fluid can best be described by making a sort of three-dimensional map of the velocity of the fluid molecules at all points within the fluid. Thus in general the velocity $\mathbf{v}$ at a given point x, y, z, will be a function of the time; $\mathbf{v}$ will change as time goes on. In turbulent flow this change is erratic.

In the remainder of this chapter we shall dismiss turbulent flow from our consideration and assume that $\mathbf{v}$ at a given point is *not* a function of t; $\mathbf{v} = \mathbf{v}(x,y,z)$.

14.2. Streamline Flow. When $\mathbf{v}$ at a given point is constant in time, the fluid motion is said to be *steady*. Consider the point A (Fig. 14.2) within the fluid. Since $\mathbf{v}$ at A does not change in time, every particle arriving at A will pass on with the same speed in the same direction. The same is true about the points B and C. Therefore, if we trace out the path of the particle, as is done in Fig. 14.2, that curve will be the path of every particle arriving at A. It is called a *streamline*. A streamline is parallel to the velocity of the fluid particles at every point.

Theoretically it is possible to draw a streamline through every point in the fluid. We thus obtain a streamline pattern, as in Fig. 14.3. Let us select a bundle of streamlines, like the one indicated in the figure. This is called a *tube of flow*. The boundary of such a tube, consisting of streamlines, is always parallel to the velocity of the fluid particles. Hence *no fluid can cross the boundaries of a tube of flow.* The fluid that enters at one end must leave at the other. In steady flow the pattern of tubes pervading the fluid is stationary in time.

In Fig. 14.4 we have drawn a thin tube of flow. The velocity of the fluid within, while everywhere parallel to the tube, may change its magnitude. Suppose it to be v_1 at P, v_2 at the point Q. If A_1 and A_2 are the cross-sectional areas of the tube at these two points, the mass of fluid passing across A_1 per second is $A_1\rho_1v_1$, that passing across Q is

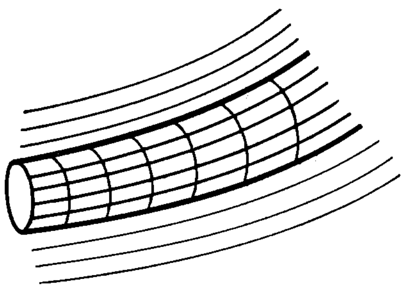


FIG. 14.3. Streamline pattern and tube of flow.

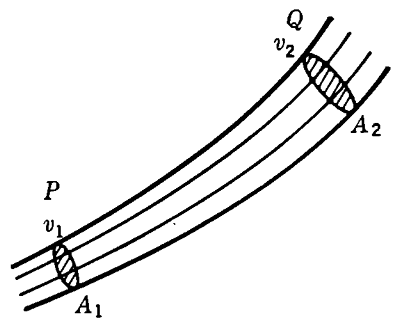


FIG. 14.4. Equation of continuity of flow.

$A_2\rho_2v_2$, provided that we write for the density of the fluid ρ_1 and ρ_2, for this may also differ from point to point in the tube. Since no fluid can leave through the walls of the tube,

$$\boxed{A_1\rho_1v_1 = A_2\rho_2v_2} \tag{14.2}$$

This very important result is called the *equation of continuity of flow.* If the fluid is incompressible, $\rho_1 = \rho_2$ and Eq. (14.2) takes the simpler form

$$A_1v_1 = A_2v_2 \tag{14.3}$$

It says: Where the area is large, the speed of flow is small, and vice versa. A river is not unlike a tube of flow: an abrupt decrease in velocity without increase in width implies increase in depth, for water is nearly incompressible.

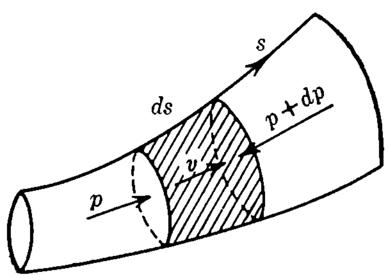


FIG. 14.5. Bernoulli's equation.

14.3. Bernoulli's Equation. One of the most useful and practical theorems of hydrostatics and aeronautics was discovered by Daniel Bernoulli in 1738. It is very easily demonstrated by applying Newton's second law to the fluid within a tube of flow.

Consider the portion of fluid in the shaded part of the tube drawn in Fig. 14.5, a part of infinitesimal length ds and of cross-sectional area A. The pressure on the left face of the shaded portion is p, that on the right is $p + dp$. If m is the mass of the fluid, Newton's law requires that

$$m \frac{dv}{dt} = -A\ dp$$

or, since

$$m = \rho A\ ds$$

ρ being the density, and

$$\frac{dv}{dt} = v \frac{dv}{ds}$$

we have

$$v\ dv + \frac{dp}{\rho} = 0$$

and, on integrating,

$$\boxed{\frac{1}{2} v^2 + \int \frac{dp}{\rho} = \text{const}} \tag{14.4}$$

This is the general form of Bernoulli's equation. To use it we must know ρ as a function of p so that we can carry out the indicated integration. For the special case of an *incompressible* fluid, ρ is constant, and Bernoulli's equation takes the form $\frac{1}{2}v^2 + p/\rho = \text{const}$. This is customarily written

$$\boxed{\tfrac{1}{2}\rho v^2 + p = H} \tag{14.5}$$

the right-hand side here being the equivalent of the former constant times ρ.

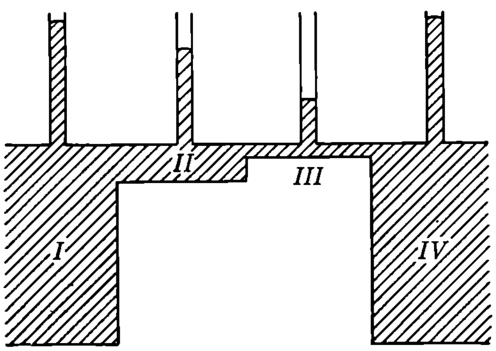


FIG. 14.6. Application of Bernoulli's equation.

14.4. Significance of Bernoulli's Equation. Each term in Eq. (14.5) has the physical dimension of a *pressure* (show this!), and H is called the *pressure head*, or total pressure. This is made up of two parts, p, the ordinary, or static, pressure; and $\frac{1}{2}\rho v^2$, the "kinetic pressure." When a liquid passes through a pipe, the pipe very nearly forms a tube of flow, and if we disregard fluid friction, Eq. (14.5) may be applied to it. Let us therefore consider Fig. 14.6. In parts I and IV the pipe is supposed to be so wide that, in view of Eq. (14.3), v is practically zero. Equation (14.5) then tells us that the manometers on I and IV read a pressure p which approximately equals H, the total head. The pressures in II and III are less because $\frac{1}{2}\rho v^2$ is greater.

When a liquid flows through the pipe of Fig. 14.7, the pressure at the constriction is greatly reduced. When the velocity in the normal part of the pipe is v and the area of the constriction is one-tenth the normal area, we have

$$\tfrac{1}{2}\rho v^2 + p_1 = \tfrac{1}{2}\rho(10v)^2 + p_2$$

or

$$p_1 - p_2 = \tfrac{1}{2}\rho \times 99v^2$$

For water, $\rho = 62.4$ lb/ft³, and with $v = 1$ ft./sec,

$$\begin{aligned} p_1 - p_2 = 31.2 \text{ lb/ft}^3 \times 99 \text{ ft}^2/\text{sec}^2 \\ &= 3{,}090 \text{ lb/ft sec}^2 = 3{,}090 \text{ poundal/ft}^2 \\ &= 96 \text{ lb/ft}^2 = \tfrac{2}{3} \text{ lb/in}^2 \end{aligned}$$

If Eq. (14.5) is applied to the motion of gases it is not a very good approximation to the truth (since gases are not at all incompressible) but it gives results that are qualitatively correct.

***14.5. Venturi Meter and Air-speed Indicator.** The reduction in pressure occurring at a throat in a pipe has a number of applications.

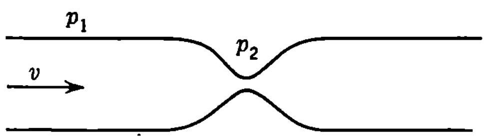


FIG. 14.7. Principle of the aspirator.

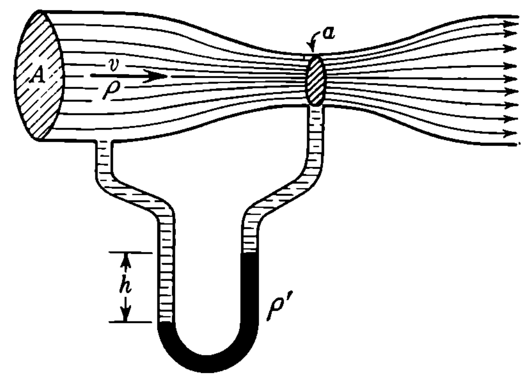


FIG. 14.8. Venturi meter.

The principle at work in these devices is always the same; the *equation of continuity* (Eq. 14.3) requires that the velocity of the fluid at a constriction increase; *Bernoulli's equation* (14.5) then shows that the pressure must fall.

The *Venturi meter* is a gauge inserted in the flow pipe for the purpose of measuring the speed of flow of a liquid. Its action may be understood from Fig. 14.8. A liquid of density ρ flows through a pipe of normal cross section A. At the throat, where the area is reduced to a, a manometer tube is attached, and the manometer liquid (mercury or water) has a density ρ'. By applying Eqs. (14.3) and (14.5) the student may show that

$$v = a\sqrt{\frac{2\rho' gh}{\rho(A^2 - a^2)}} \tag{14.6}$$

The volume of the liquid transported per second, Q, is

$$Q = vA$$

Usually a is about one-fifth of A, and for many purposes a^2 may be neglected in comparison with A^2. Approximately, therefore,

$$Q = a\sqrt{\frac{2g\rho'}{\rho}h} \tag{14.7}$$

The air-speed indicator of airplanes involves an adaptation of this principle. Two cylindrical tubes, attached near the forward edge of the wing with their

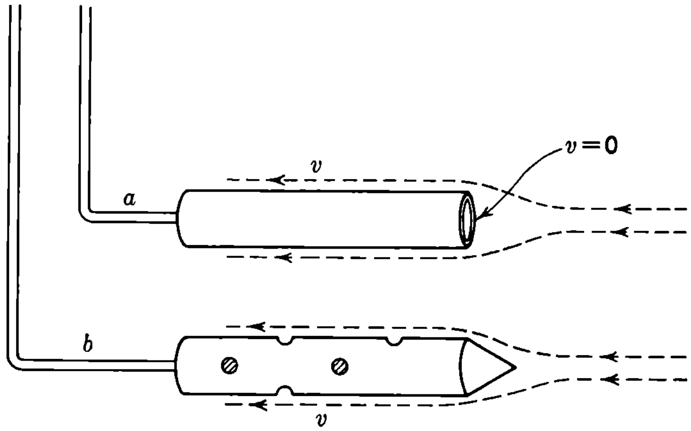


FIG. 14.9. Pitot air indicator.

axes parallel to the fuselage, are exposed to the air stream as is indicated in Fig. 14.9. The upper tube a is open at the forward end; the lower b is closed but carries holes in its wall. The air flows past these holes with a speed v that equals the air speed of the airplane. In front of tube a, however, the air is stagnant, and the velocity is nearly zero. If we call the pressures inside the tubes p_a and p_b, we have, by Bernoulli's equation,

$$p_b + \tfrac{1}{2}\rho v^2 = p_a \tag{14.8}$$

The pressure in tube a is higher than in tube b; therefore a is called the "pressure" tube, b the "static" tube. The leads from the two tubes are connected to a pressure gauge, usually in the form of an aneroid barometer. Tube b is connected to the inside, the other to the outside of the barometer box, and $p_a - p_b$ is measured. By Eq. (14.8) this is approximately (air is compressible!) equal to $\frac{1}{2}\rho v^2$. This device, called a *Pitot* tube, can be calibrated to read v directly.

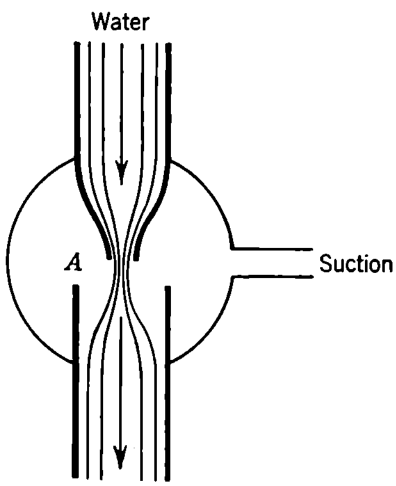


FIG. 14.10. Aspirator pump.

Though it serves a different purpose we mention in this connection the aspirator pump, which also uses the reduction in pressure at a throat. Its action is illustrated in Fig. 14.10. At the point A, where the water stream leaves the constricted nozzle, the pressure is lower than atmospheric, and suction occurs through the side tube. The pump can be used to produce pressures as low as a few millimeters of mercury.

Gasoline vapor finds its way into the manifold of an automobile engine by throat suction.

14.6. Speed of Efflux under Gravity. The speed with which a liquid flows out of a spout can be found to a fair degree of approximation with the use of Bernoulli's equation. Assume that the container drawn in Fig. 14.11 is kept filled to a height h above the spout while liquid runs out. The stream has a varying cross section; it is sometimes called a *vena contracta*. Since it is exposed to the air, the pressure acting upon it is nearly atmospheric. At the narrowest point of the stream it is almost exactly atmospheric ($p = p_0$).

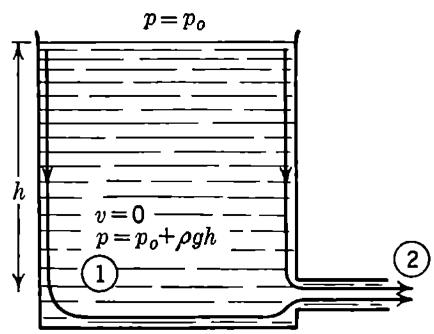


FIG. 14.11. Efflux under gravity.

A tube of flow has been drawn in Fig. 14.11; this tube comprises practically the whole vessel and the spout. We apply Bernoulli's equation to the liquid in this tube. At the point (1), where the tube is wide, v is practically zero, but $p = p_0 + \rho gh$. At the point (2) in the *vena contracta* the liquid has velocity v, but the pressure is p_0.

Hence
$$p_0 + \rho gh = p_0 + \tfrac{1}{2}\rho v^2$$

and
$$v = \sqrt{2gh} \tag{14.9}$$

This result was first found by *Torricelli* and is known as his theorem.

The speed of water flowing from the bottom of a standpipe 30 ft high would be

$$v = (2 \times 32 \text{ ft/sec}^2 \times 30 \text{ ft})^{1/2} = 43.8 \text{ ft/sec}$$

***14.7. Lift of an Airplane Wing.** Wherever, in any flow pattern, the streamlines are abnormally crowded, the velocity of flow is large. Hence Bernoulli's equation requires the pressure to be small. Figure 14.12 shows the streamlines around the wing of an airplane or a model of it (airfoil) such as is used in wind tunnels. The crowding of streamlines above causes p_2 to be greater than p_1, and the lift force results.

The crowding can be strengthened, and therefore the lift can be increased, by tilting the tail of the foil downward, thus changing the "angle of attack." To be sure, this is practicable within limits. But for every air speed there is a certain maximum angle above which streamline flow cannot be maintained. The air then breaks into myriads of small vortices past the edge of the foil, and the airplane stalls. Even at the proper angle of attack, irregularities on the wind surface (ice) can cause turbulence and stalling.

As already mentioned, the drag due to turbulence is very great. As an illustration we have drawn in Fig. 14.13 three shapes in a wind tunnel. The first is an ordinary cylinder, the second a cylinder tipped by a rounded cone, the last a "teardrop" shape. In the first instance there is much turbulence, in the second

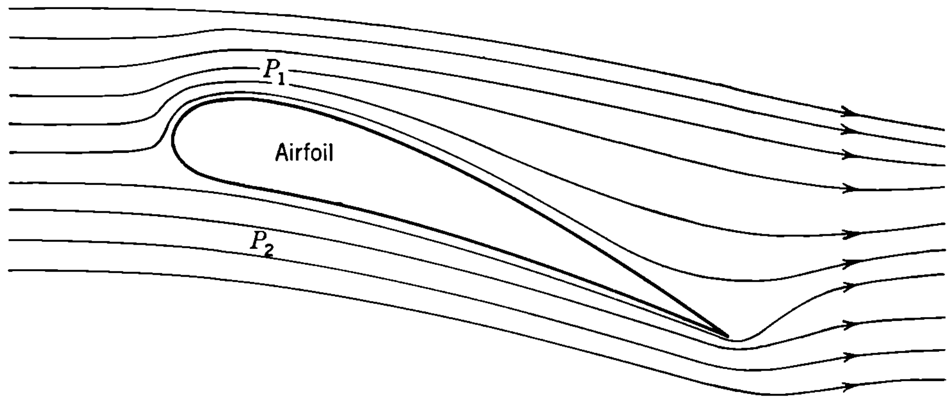


FIG. 14.12. Lift on airfoil.

some, in the last none. For a certain air speed, equal in the three cases, the drag on the objects is listed in the figure.

Rapidly falling drops have the teardrop shape. This is partly because the forces exerted by the air on the falling drop produce that shape, partly because

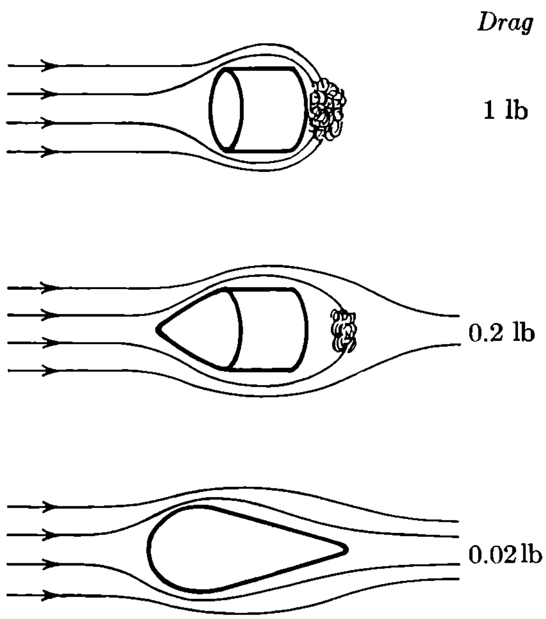


FIG. 14.13. Air resistance.

if the drop did not have this form it would not fall rapidly, or be destroyed by turbulence.

***14.8. Limiting Velocity in Air.** The student has probably wondered about the so-called "sound wall," the mysterious obstacle that makes it so extremely difficult to fly an ordinary airplane with a speed approaching that of sound. The answer is given to a fair approximation by Bernoulli's equation, which does not allow streamline flow to exist at high velocities.

Let us consider the air as incompressible and write

$$p + \tfrac{1}{2}\rho v^2 = H$$

Assume the airplane to be stationary, and let the air move past it, as in a wind tunnel. This equation must be true both for the stagnant air at the front edge of the wing and for the air rushing over the wing surface. Let the velocity of the latter—which is the velocity of the flying airplane—be v. Where the air is stagnant, the pressure is approximately atmospheric. Hence, putting $p = 10^6$ dynes/cm² and $v = 0$, we find $H = 10^6$ dynes/cm². Now the greatest v is that for which $p = 0$. Therefore

$$\tfrac{1}{2}\rho v_{max}^2 = 10^6 \text{ dynes/cm}^2$$

For air, $\rho = 1.2 \times 10^{-3}$, and

$$v_{max} = \sqrt{\frac{2 \times 10^6}{1.2 \times 10^{-3}} \frac{\text{cm}^2}{\text{sec}^2}} = 4 \times 10^4 \text{ cm/sec} = 900 \text{ miles/hr}$$

approximately. Actually, troubles start at lower velocities because the air is not incompressible. The velocity of sound is 770 miles/hr. The ratio of the air velocity to the velocity of sound is called the *Mach number*. A gas may be treated as incompressible without much error if the Mach number is less than ½.

PROBLEMS

1. Water flows through a 6-in. pipe at the rate of 20 gal/min. At one place there is a constriction of radius 2 in. What is the velocity of the water, in ft/sec, at this constriction?

2. In Prob. 1, the normal pressure within the pipe is 1.6 atmospheres. What is the pressure at the constriction?

3. A standpipe is 30 ft tall and filled completely. At half its height there is a circular hole of radius ½ in. (*a*) How much water will squirt out per minute? (*b*) Where will the stream hit the ground?

4. Explain why water flows as a continuous stream down a vertical pipe, while it breaks into drops when falling freely.

5. A river carries 1 million cubic feet of water per minute. It is dammed and drops 30 ft. What is the maximum horsepower available for a hydroelectric plant at this dam?

***6.** A Venturi water meter registers a pressure difference of 8 lb/in.² between the constriction, of internal diameter ¼ in., and the normal section of diameter 1 in. How many gallons does it pass per second?

***7.** Neglecting all viscous drag, find from Bernoulli's principle the maximum velocity with which water at atmospheric pressure can flow past an obstacle without breaking into turbulence.

***8.** What is the difference of pressure in the two arms of an air-speed indicator (Pitot) of an airplane traveling at 250 miles/hr?

***9.** Derive Formula (14.6).

CHAPTER 15
INTERMOLECULAR FORCES

15.1. Origin and Nature of Intermolecular Forces. On many occasions in this text, reference has been made to the forces that act between the molecules, and the time has come when these should be more carefully considered. All atoms and molecules contain electric charges. If all carried an excess of the same kind of charge, *i.e.*, positive or negative, there would be very strong forces of repulsion between the particles. This, however, is not the case, for the negative charge on the electrons (Sec. 24.5) is just equal and opposite to the charge on the nucleus.

From this fact one might draw the conclusion that molecules cannot interact electrically; this would be erroneous. In fact, when two molecules approach each other, the charges in each are disturbed and depart slightly from their normal positions in such fashion that the average distance between opposite charges in the two molecules is a little smaller than that between like charges. Attraction thus wins over repulsion, and intermolecular forces result. Because these delicate internal readjustments take place only when molecules are fairly close together, the forces act over short distances only: they are said to have a short range. A sphere of radius equal to the range of the forces, constructed about a given molecule as center, is called a "sphere of action."

The situation is different when the molecules come *very* close together, so close that their outer charges touch. They must then repel each other strongly, for there is no way for the molecule to rearrange itself internally so as to anul the repulsion of the closely adjacent external electrons. It is this repulsion on contact that accounts for the rigid behavior or the elastic-billiard-ball character of molecules. Were it not for its occurrence, molecules could not rebound in collisions but would move right through each other.

To describe intermolecular forces graphically it is customary to plot, not the force, but the *potential energy* of two molecules as a function of the distance, x, between their centers. The force F is related to the potential energy V by the equation

$$F = -\frac{dV}{dx} \tag{15.1}$$

It equals the negative slope of V. This allows us at once to interpret the graph of V vs. x (Fig. 15.1): Let one molecule be fixed at O. Then the other will be urged away from it where the slope of V is downward, toward it where the slope is upward. The range from O to x_0 is called the repulsive range of the force; that from x_0 outward to a point where the curve is practically horizontal and the force vanishes is the attractive range. At x_0 no force acts between the molecules, for $dV/dx = 0$. The steepness of the slope is a measure of the strength of the forces. All this may be summed up by noting that a particle "slides down a potential hill."

When molecules collide, their centers approach each other until they are approximately within a distance x' from each other. This distance is therefore to be regarded as the *diameter* of a molecule. For the simplest molecules, x' is about 2 or 3×10^{-8} cm; x_0 about 3 or 4×10^{-8} cm, and the forces practically cease at about 10^{-7} cm.

Equation (15.1) and the potential diagram of Fig. 15.1 illustrate a general principle of physics, sometimes called the *principle of minimum potential energy*. It says, somewhat vaguely, that physical systems always tend to make their potential energy a minimum. Really this is nothing new, but an immediate consequence of Eq. (15.1), which indicates that forces always act to compel a system *down* the potential hill until $dV/dx = 0$, and this point must be a minimum. The tendency for V to seek a minimum is not limited to molecules or other particles but applies to large assemblages of objects, for Eq. (15.1) can be generalized to all such situations. We shall have occasion to use this principle again in Sec. 15.3.

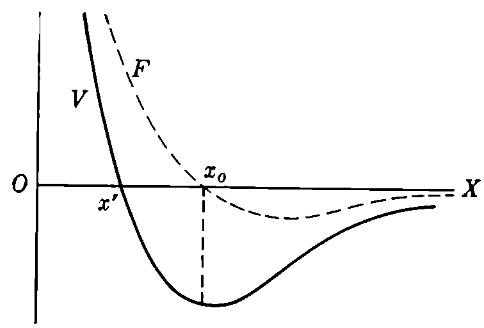


FIG. 15.1. Intermolecular potential. (Dotted curve represents force.)

Since different molecules have different internal arrangements of charges, intermolecular forces differ from one kind of molecule to another. But they always show the characteristic qualitative behavior drawn in Fig. 15.1. In Table 15.1 we have listed the force of attraction for a few simple molecules at a distance of separation of 4×10^{-8} cm. In the last column there is given the *gravitational* attraction (cf. Sec. 11.1) at the same distance. It is seen to be so small in all instances as to be utterly negligible in comparison with the intermolecular forces. Because the mass of a molecule is so small (about 10^{-23} gm) the intermolecular forces given in the table are strong enough to impart to it an instantaneous acceleration of about 10^{17} cm/sec^2. However, these enormous accelerations last for only a very short time, since one molecule moves quickly out of the range of action of the other.

***15.2. Effects of Intermolecular Forces.** *Cohesion*, or the "hanging together" of the molecules of a solid body, is due in most instances to the action of the forces described above. The fact that glass resists breaking, the hardness and rigidity of rocks, the elastic properties of solids may all be traced to their presence. By fitting together the pieces of broken glass no joint can be achieved in spite of the fact that before breaking the same molecules attracted one another strongly. This is because of the short range of the forces; it is impossible by mere mechanical fitting to bring all molecules back within their ranges of action. Union can be achieved, however, by fusing the pieces together. Very carefully polished glass plates, so-called "optical flats," when placed with their faces in contact cling together so firmly that they are exceedingly difficult to separate.

Adhesion is the sticking of one substance upon another. Powder and dust adhere to almost anything; water adheres to glass, oil to water, paint to walls. These are some of the more obvious examples illustrating the action of intermolecular forces. The lubricating quality of oil, which creeps over metals, is another useful instance. When gases adhere to solids, they are said to be *adsorbed*. A charcoal surface attracts and retains upon itself great quantities of gas molecules. At low temperatures every metal surface exposed to gas is covered with a film of gas molecules.

Table 15.1. *Comparison of Intermolecular and Gravitational Forces*

Molecule	Intermolecular force, dynes, at a separation of 4×10^{-8} cm	Gravitational force, dynes, at a separation of 4×10^{-8} cm
He	6×10^{-8}	7.3×10^{-47}
H_2	43×10^{-8}	1.8×10^{-47}
A	225×10^{-8}	7.3×10^{-45}
N_2	215×10^{-8}	3.6×10^{-45}
O_2	150×10^{-8}	4.7×10^{-45}
CO_2	560×10^{-8}	8.8×10^{-45}

The properties of a liquid cannot be explained without an understanding of intermolecular forces. By virtue of their presence a liquid is prevented from expanding into a gas. In the process of evaporation, certain molecules are taken out of the clutches of their neighbors; this requires work, and it has been found that the work needed to pull one molecule free is of the same order of magnitude as the depth of the minimum in Fig. 15.1. Again, the forces manifest themselves in the behavior of liquids at membranes, in osmosis, in the lowering of the freezing point and the raising of the boiling point of liquids by foreign admixtures. Their complex action in living organisms is gradually revealing itself to the scientist, and the new science of biophysics involves them in innumerable instances.

Three specific effects caused or greatly influenced by intermolecular forces in liquids will be dealt with in the following sections. They are *surface tension*, *capillarity*, and *viscosity*. To be sure, fluids would be viscous even in the absence of attractive forces between their molecules; but these forces modify the viscous behavior sufficiently so that it is proper for us to treat viscosity in this chapter.

In gases the effects of intermolecular forces are but minor. They introduce an additional internal pressure at high compression, when the gas molecules are close together. This, however, can be measured with great accuracy and is the experimental source of most of our present knowledge concerning these forces.

15.3. Surface Tension. Consider an imaginary surface S as dividing a liquid into two parts (Fig. 15.2). Since there are forces of attraction between the molecules of the liquid, stresses must act across this surface; the molecules on the left attract the molecules on the right. Therefore, if the liquid were actually separated into two parts with S as the bounding surface of one of the parts, *work* would have to be done.

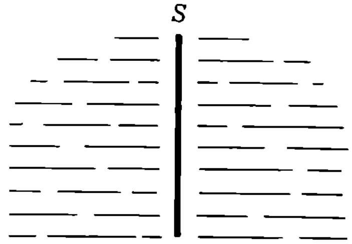


FIG. 15.2. Separating two halves of a liquid requires work.

The only result of this work is the creation of two additional surfaces of the liquid. Hence we may say that the production of a larger amount of free surface requires an increase of potential energy of the liquid. And since the liquid, when undisturbed, seeks to minimize its potential energy, it must take a shape in which its free surface is as small as possible. Its surface tends to contract, and the measurable force of contraction is called "surface tension."

As evidence of this contraction we note that a small drop of liquid, when falling in a vacuum and thus subject to no stresses except for the forces between its molecules, takes on the shape which has the smallest surface, *i.e.*, the shape of a sphere. Every distortion from sphericity involves an increase in the distance between some of the molecules, and this is resisted by the attractive forces between them.

The surface tension of a liquid has many features in common with the tension in a stretched rubber membrane, but it is quite different in one important respect. Imagine a plane liquid surface of area A and a stretched rubber membrane of the same area. If a straight line is drawn across each surface, the part of the surface on one side of the line pulls on the other with a force, F_l for the liquid and F_r for the rubber, that is proportional to the length of the line, d. In general F_r will be much greater than F_l, but otherwise the situation is the same in both examples. However, if the surface area is increased, let us say to $2A$, F_r will increase while F_l remains the same. Stretching the rubber membrane increases the tension; increasing the liquid surface leaves the surface tension unaltered.

As was noted, the force F_l is proportional to the length of the line we imagined drawn across the liquid surface. What physicists mean by surface tension is not the total force F_l but the *force per unit length*, F_l/d.

Surface tension, therefore, is not truly a force; it is a force divided by a distance. When thus defined, surface tension is constant for a given liquid at a given temperature.

To measure surface tension one may apply this definition in the following way: A wire is bent into the shape of an open rectangle (Fig. 15.3), and a thin, straight wire is placed across it. A liquid film (*e.g.*, soap film) is then formed as indicated, and the force is measured that will just keep the film from contracting. If the length of the cross wire in contact with the film is d, the surface tension is given by

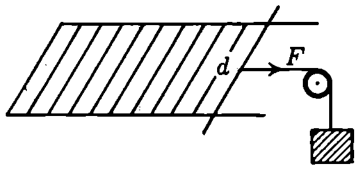


FIG. 15.3. Measuring surface tension.

$$\sigma = \frac{F}{2d} \tag{15.2}$$

the factor 2 being inserted because the film has two surfaces. This experiment has the advantage of conceptual simplicity; but for most liquids such a film cannot be formed, and the present method of measurement is not practicable. In Sec. 15.6 we describe a more useful procedure.

Surface tension is measured in dynes/cm. Some typical values are given in Table 15.2.

Table 15.2. *Values of Surface Tension for Various Substances*

Liquid	Temp, °C	σ, dynes/cm
Water	0	75.6
	30	71.4
	90	62.9
Ethyl alcohol	0	23.5
	30	20.8
	70	17.3
Mercury	18	520
He	−269	0.12
Ne	−247	5.15
O_2	−193	15.7

***15.4. Pressure within a Spherical Drop.** The pressure within a spherical drop is greater than the pressure outside. Suppose the drop were divided into two hemispheres (cf. Fig. 15.4). If the pressure inside is p and that outside is zero, the force that the upper hemisphere exerts on the shaded equatorial plane is $\pi r^2 p$. Since the drop holds together, surface tension acting across the equator must provide an equal force. Since σ is the force per unit length and since the length of the equator is $2\pi r$, we must have

$$\pi r^2 p = 2\pi r \sigma$$

Hence

$$p = \frac{2\sigma}{r} \tag{15.3}$$

If the outside pressure is not zero, p stands for the excess pressure inside the drop. The smaller the radius of the drop, the greater the pressure within.

Equation (15.3) also represents the pressure within a gas bubble enclosed in a liquid. It poses an interesting paradox with respect to the mechanism of bubble growth. For if, in the formation of a bubble, a very small cavity in the liquid had to be formed first, enormous initial pressures would be required. In water, for example, a spherical cavity of diameter equal to the size of 10 molecules (3×10^{-7} cm) calls for a pressure of $150/(1.5 \times 10^{-7})$ dynes/cm^2 = 1,000 atm, approximately. It is likely, therefore, that air bubbles form on nuclei of impurities in the water.

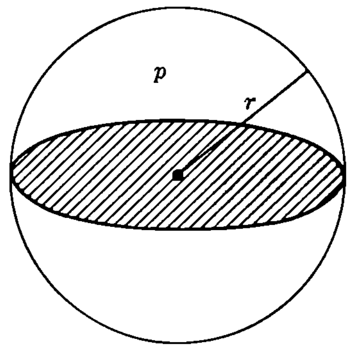


FIG. 15.4. Pressure inside spherical drop.

The excess pressure inside a soap bubble is also given by Eq. (15.3), but allowance must be made for the fact that the soap bubble has two surfaces. so that

$$p = \frac{4\sigma}{r}$$

The same result holds for an inflated balloon, provided that σ is taken to be the (variable) surface tension of its walls. If two soap bubbles are formed on the ends of a tube that connects them, the larger one will grow at the expense of the smaller until the latter disappears, for the pressure in the smaller bubble is greater.

15.5. Angle of Contact. Liquids frequently "creep up" on solids. To understand this we note that, while there are forces between the molecules of the liquid, there are also attractions between those of the solid and the liquid, and these latter may be either greater or smaller than the former. If they are greater, it is energetically favorable, in the sense of the minimum principle discussed in Sec. 15.1, for the liquid molecules to be near the solid surface or at least to strike a compromise between adhering to the solid surface and minimizing the free surface of the fluid.

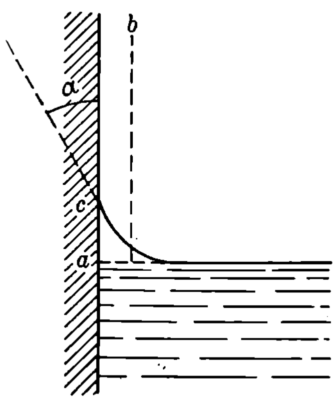


FIG. 15.5. Angle of contact.

This may be illustrated with reference to Fig. 15.5, which shows the behavior of a liquid near a solid wall. Surface tension alone would force the liquid molecules to form a minimum surface, which is horizontal (a in the figure). Attraction by the wall would cause them to creep up in a vertical film (b in the figure). The actual behavior, a compromise between these two situations, is depicted by curve c. The liquid comes up to the surface under a definite angle α, called the *angle of contact*.

Table 15.3. *Angles of Contact*

Glass-water	0°
Steel-water	90°
Glass-mercury	140°
Paraffin-water	110°
Glass-ethyl alcohol	0°

Angles of contact have been measured for many solid-liquid pairs. Their values are not very accurate or meaningful, for they vary greatly with conditions of purity. A few representative values are given in Table 15.3. Liquids that "wet" a surface have contact angles smaller than 90°. Clean water makes a zero angle with glass, whereas mercury with a contact angle of 140° forms a convex meniscus in a glass tube (Fig. 15.6).

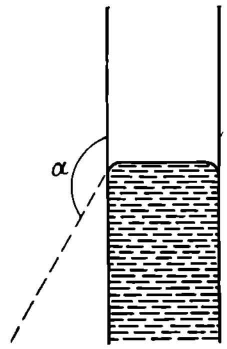


FIG. 15.6. Mercury in glass.

15.6. Capillarity. The rise of water in tubes of small bore, so-called "capillaries" (Latin *capilla,* hair), is a well-known phenomenon. Sap rises in trees partly because of capillary action; water creeps up through cracks in rocks, often causing them to crumble when the water freezes. Unless soil is plowed, water rises quickly through its pores and evaporates, causing it to dry out. Capillary action is caused by the intermolecular attraction between liquid and solid molecules on the one hand and by surface tension on the other.

The rise or depression of liquid in a tube of circular cross section will be studied by reference to Fig. 15.7, which shows a liquid that has risen in a tube to an average height h above the liquid surface. Surface tension acts along the junction of the liquid and the tube, *i.e.*, along a circle of length $2\pi r$ if the tube has an internal radius of r cm. It is directed upward at an angle α with the wall and exerts a force of $2\pi r\sigma \cos \alpha$ dynes vertically upward. This is just sufficient to balance the weight of the liquid, which is $\pi r^2 h\rho g$, provided that its density is ρ. Thus

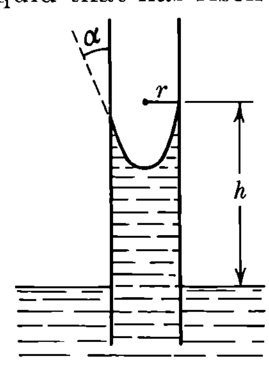


FIG. 15.7. Capillary action.

$$2\pi r\sigma \cos \alpha = \pi r^2 h\rho g$$

or

$$h = \frac{2\sigma \cos \alpha}{r\rho g} \tag{15.4}$$

For water, $\cos \alpha = 1$. For a liquid like mercury, which makes an obtuse angle of contact, $\cos \alpha$ is negative; hence h is negative, and the liquid is depressed below the level in the container.

The student can show by the same method that the liquid will rise between parallel plates to a height

$$h = \frac{2\sigma \cos \alpha}{a\rho g}$$

if a is the distance between the plates. Observations on the capillary rise of liquids and on the angle of contact provide a practical method for measuring surface tension.

***15.7. Viscosity.** In Sec. 4.3 we defined the shear modulus as the ratio of shearing stress to shearing strain,

$$M = \frac{F_{\parallel}/A}{\Delta s/\Delta y} \tag{15.5}$$

in the notation illustrated by Fig. 15.8. In Sec. 4.2 the ratio $\Delta s/\Delta y$, being small, was identified with the angle θ. If the cube drawn in the figure were a cube of liquid, the tangential force F would *not* produce a finite displacement Δs; it would, in fact, cause the horizontal layers of liquid to flow with varying velocities. If the bottom of the cube is in contact with a stationary wall, observation shows that the lowest layer would be at rest and the higher ones would have uniformly increasing velocities as shown at the right. By virtue of its viscous behavior an upper layer drags the lower ones along, but the lower follows with a somewhat smaller speed.

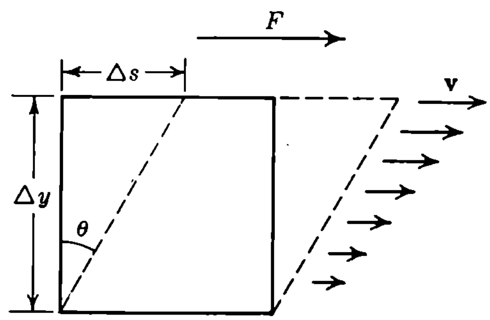


FIG. 15.8.

To define a shear modulus in accordance with Eq. (15.5) is therefore meaningless. If, however, we replace s by ds/dt in that equation, so that the coefficient measures the ratio of stress to *rate* of straining, then it becomes a finite quantity, called the *coefficient of viscosity* and denoted by η. Let us write v for ds/dt and take Δy to be infinitesimally small. The definition of η is then

$$\eta = \frac{F_{\parallel}/A}{dv/dy} \tag{15.6}$$

The cgs unit of η is the *poise*, or dyne sec/cm^2. Other units are the centipoise (10^{-2} poise) and the micropoise (10^{-6} poise). Values of η for various substances are given in Table 15.4. It will be noted that η varies with the temperature, increasing with T for gases, decreasing for liquids.

When a fluid is forced through a tube of small bore the flow is controlled entirely by its viscosity. Poiseuille showed, by a simple calculation that may be found in books on hydrodynamics, that the volume of fluid, Q, passing through a tube of length L and of internal radius r per second under a difference in pressure Δp between its ends, is given by

$$Q = \frac{\pi r^4}{8L\eta} \Delta p \tag{15.7}$$

Table 15.4. *Coefficient of Viscosity*

Substance	Temp, °C	η, dyne sec/cm²
Water	0	0.0179
	20	0.0101
	40	0.0066
	60	0.0047
Air	0	1.68×10^{-4}
	20	1.72×10^{-4}
	40	1.90×10^{-4}
	60	2.02×10^{-4}
Mercury	20	0.00155
Glycerin	20	8.3
Oil:		
SAE 10*	40	2
SAE 20	40	2.8
SAE 30	40	4

* SAE denotes a scale established by the Society of Automotive Engineers.

This provides an excellent method for determining the coefficient of viscosity for liquids. Since a gas is compressible, Poiseuille's law needs slight modifications when it is applied to gases; but it is still essentially correct.

Another important relation, which will not be derived here, is *Stokes' law*. A small drop falling in a viscous medium is retarded by a force that is proportional to its radius r and to the coefficient of viscosity of the medium, as well as to its speed v. Thus

$$F = 6\pi\eta r v \tag{15.8}$$

In view of what was learned in Sec. 7.4 [see Eq. (7.10), with $b = 6\pi\eta r$] the drop reaches a terminal velocity

$$v_t = \frac{mg}{6\pi\eta r} \tag{15.9}$$

This relation will again be used in Sec. 25.12, where we discuss the electric charge on an electron.

PROBLEMS

1. How high will water rise in a glass capillary of internal diameter 1 mm?

2. Find the depression of the mercury in a glass tube of internal diameter 0.01 in.

3. A mercury barometer consisting of a glass tube filled with mercury reads 75.43 cm at the top of its meniscus. The internal diameter of the tube is 0.5 cm. Correct this reading for capillary action.

4. What force, in addition to gravity, is required to pull a horizontal glass rod of length 5 in. slowly upward through a water surface?

5. A piece of a matchstick of length 1 cm floats on a water surface. A drop of alcohol added on one side of the stick reduces the surface tension on that side by 8 dynes/cm. In what direction will the stick move? What force will act on it and what will be its initial acceleration if its mass is 0.02 gm?

★**6.** What is the diameter of a spherical drop of water in which the pressure exceeds that on the outside by 3 atm?

7. Two vertical glass plates are placed face to face in a basin of water, protruding above the water surface. They are separated by 1 mm distance. How high will the water rise between them? Calculate the average pressure between the plates, and find the force with which the plates are pressed together. Assume the plates to have a length of 20 cm (Fig. 15.9).

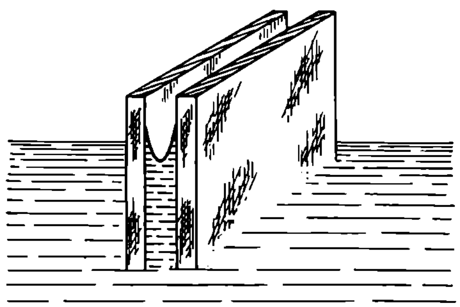


FIG. 15.9. Problem 7.

8. Two vertical glass plates are set in a dish of water in the manner of a slightly open book. They make an angle of $\frac{1}{2}°$ with each other and are in contact at one edge. Calculate the curve that the water surfaces makes between them.

★**9.** Water at room temperature drains slowly from the bottom of a container through a capillary tube of length 10 cm and internal diameter 1 mm. Within the container it stands 20 cm high. How much water flows out in 1 min?

★**10.** Find the viscous force on an oil drop of diameter 10^{-4} cm, falling with a speed of 10^{-3} cm/sec in air.

★**11.** Taking account of weight, buoyancy, and viscous retardation, calculate the limiting velocity of the oil drop of Prob. 10. (Density of oil = 0.956 gm/cc; density of air = 0.00125 gm/cm³; $\eta = 1.7 \times 10^{-4}$ dyne sec/cm².)

★**12.** Oil having a viscosity of 0.2 dyne sec/cm² is forced through a tube of length 1 m and diameter 0.2 mm under a pressure of 50 lb/in.² How much oil will pass per minute?

CHAPTER 16

TEMPERATURE AND THERMAL EXPANSION

16.1. Temperature. If we are given a number of similar bodies each of which has been heated a different amount, it may be possible by the sense of touch to arrange them in increasing order of hotness. The hotter the body, the higher its *temperature*, which is merely a number on some arbitrarily chosen scale expressing how cold or hot a body is. Although our concept of temperature is basically associated with thermal sensation, this sense is not very acute and its availability is, of course, limited at both ends of the scale. It is therefore necessary to devise more accurate ways of measuring temperature, making use of some of its many effects such as the change in volume of a substance. Similarly, forces are measured by means of some effect such as the extension of a spring. Any instrument that measures temperature is called a *thermometer*.

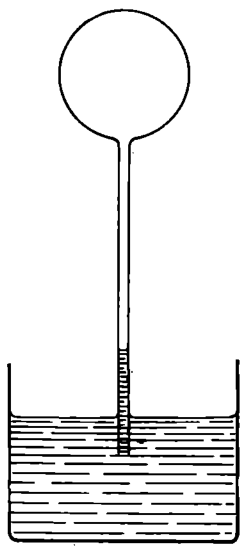


FIG. 16.1. Galileo's thermoscope.

Galileo's thermoscope, invented about the year 1600, was the first thermometer of which we have any knowledge. This instrument consisted merely of a glass bulb with a long, narrow stem dipping downward into a vessel of water (Fig. 16.1). As the temperature of the bulb rises, the air within it expands and the water level in the stem falls. If the temperature of the bulb drops, the contraction of the air causes the water to rise further in the stem. Because of the large thermal expansion of the enclosed air, this is a temperature-sensitive device, but it is useless as a practical thermometer since variations in the atmospheric pressure also affect the height of the water in the stem. This was recognized during the first half of the seventeenth century, and by 1650 there had been invented the liquid-in-glass instrument with the end of the tube sealed. This was the forerunner of the modern mercury-in-glass or alcohol-in-glass thermometer.

16.2. Thermometers. Changes in physical properties with temperature that are used by thermometers include variation of the physical dimensions of a substance, the pressure of a gas held at constant volume, the electric resistance of a wire, the thermoelectric effect, and the color of a body so hot that it is radiating visible light. Common, everyday thermometers are based either on the differential volume expansion of a liquid and its container or on the linear expansion of a metal strip, while thermometers based on the other effects have special applications such as the measurement of very high or very low temperatures.

The common liquid-in-glass thermometer consists of a thin-walled glass bulb from which a long glass capillary tube rises. Mercury or colored alcohol is the liquid usually employed. With increase in temperature these expand much more than does the glass, and the sensitivity is increased by using a capillary of very fine bore. Before sealing off the upper end of the capillary tube the space above the liquid is usually evacuated. The scale is either engraved on the glass capillary or marked on the baseboard.

In the *metallic* thermometer, an example of which is the oven thermometer, the expansion of a coiled bimetallic strip acts through a lever system to move a pointer over a scale. Because of shifts in sensitivity this instrument must be recalibrated frequently.

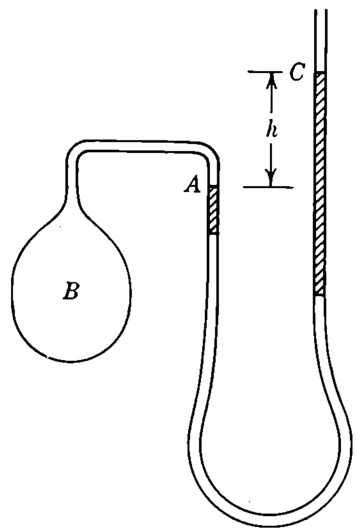


FIG. 16.2. Constant-volume gas thermometer.

The constant-volume gas thermometer is illustrated in Fig. 16.2. The pressure of the gas, usually hydrogen or helium, in the container B of constant volume is a function of the temperature only. With change in temperature of B the mercury level is adjusted to the fixed point A and the pressure determined by adding the atmospheric pressure given by a barometer to the difference in level h of the two interconnected mercury columns. There are good reasons for considering this instrument as the standard thermometer with which other more convenient forms of thermometers may be calibrated.

Since the electric resistance of a wire is a function of its temperature (Chap. 30) and is an easily measured quantity, the *resistance thermometer* is a useful instrument, particularly at high and low temperatures. Another type of electric thermometer makes use of the fact that in an electric circuit formed of two different metals a current flows if the two junctions are at different temperatures. This thermoelectric effect is described in detail in Chap. 31.

In the *thermoelectric pyrometer* the total radiant energy from a hot source falls on a thermocouple or thermopile (Chap. 31). Since the total radiation from the source is a function of its temperature (Sec. 22.6), the heating of the thermocouple junction and hence the electric current in the circuit varies with the intensity of the radiation, and the reading of the millivoltmeter in the circuit may be used as an indicator of the temperature of the source.

The *optical pyrometer* for the measurement of very high temperatures is based on the fact that, when an object is so hot that it is incandescent, its temperature may be judged from its color. This instrument consists of a telescope in whose tube are a red filter and a small electric lamp whose filament current is adjustable. The telescope is sighted on the hot object, for example a furnace, and the current in the lamp varied until its filament appears to have the same brightness as the furnace. From a calibration obtained by noting the filament currents to give a brightness match for hot sources of known temperature the meter may be marked to indicate the temperature directly.

The readings of two different kinds of thermometer will not agree exactly over any range of temperatures. But before making any comparisons we should first consider the construction of temperature scales and the general definition of temperature.

16.3. Temperature Scales. In defining a temperature scale two *fixed points* must be chosen and arbitrary values assigned to these temperatures. This determines the zero point and the size of the temperature unit or degree (°). For all scales the fixed points are taken as the temperature of a mixture of ice and pure water and the temperature at which pure water boils, both at atmospheric pressure. On the *Fahrenheit scale,* which is in common use in this country, the melting point of ice is designated 32 and the boiling point of water 212. On the *centigrade scale,* which is universally used in scientific work, these two fixed points are labeled zero and 100, respectively.

Let P represent the value of the physical property whose variation with heating is used in the construction of a thermometer. Then on the centigrade scale if we denote by P_0 the value of P at the ice point, by P_{100} its value at the steam point, and by P_t its value at any other temture t, this temperature is defined by the equation

$$t = \frac{P_t - P_0}{(P_{100} - P_0)/100} \tag{16.1}$$

This represents temperature as a linear function of the property P. On the Fahrenheit scale the corresponding definition of temperature is

$$t - 32° = \frac{P_t - P_{32}}{(P_{212} - P_{32})/180} \tag{16.2}$$

Comparison of these two temperature scales, together with a third, the *absolute scale,* on which temperatures are reckoned from the lowest temperature theoretically conceivable, is made in Fig. 16.3. This absolute scale, which is also known as the "Kelvin scale" after Lord Kelvin (William Thomson, 1824–1907), who suggested it from thermodynamical considerations, uses centigrade divisions. Since between the ice point and the steam point there are 100 centigrade divisions or 180 Fahrenheit divisions, the centigrade degree is $^{180}/_{100}$, or $^9/_5$, times larger than the Fahrenheit degree. Therefore in converting from one scale to the other we may use the relations

$$\left.\begin{aligned} C &= \tfrac{5}{9}(F - 32) \\ F &= \tfrac{9}{5}C + 32 \end{aligned}\right\} \qquad (16.3)$$

where C and F represent corresponding temperatures on the centigrade and Fahrenheit scales, respectively. For example, given that absolute zero is $-273°C$, to convert this to the Fahrenheit scale we compute from Eq. (16.3) that $F = (\tfrac{9}{5}) \times (-273) + 32 = -491 + 32 = -459°$, the value given in Fig. 16.3.

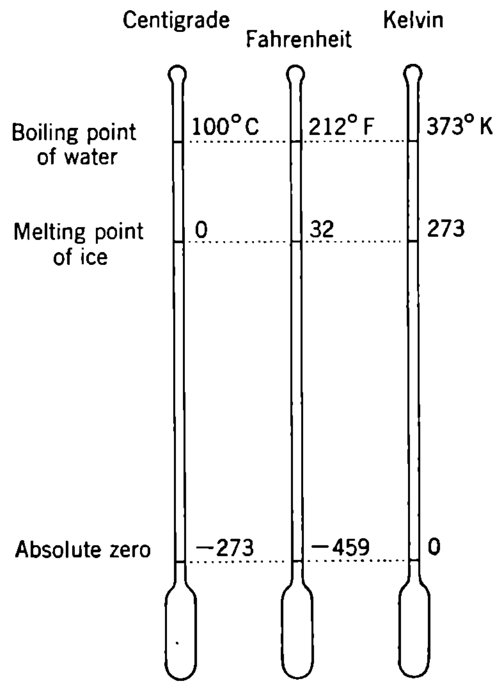


Fig. 16.3. Comparison of temperature scales.

By using Eqs. (16.1) and (16.2) any one of the various types of thermometers may be used to establish a temperature scale. They will all agree at the fixed points, but the properties P are in general different functions of the temperature, and hence the scales based on the variations of different P's will not in general agree exactly at any points other than 0 and 100°C. It is therefore necessary to have a particular property of some thermometric substance for the establishment of a standard temperature scale against which the scales of other kinds of thermometer may be compared. The standard adopted is the constant-volume gas thermometer using hydrogen or helium, for which the property P is the pressure p of the gas. With this standard gas thermometer, temperature on the centigrade scale is thus defined by the relation

$$t = \frac{p - p_0}{\tfrac{1}{100}(p_{100} - p_0)} \qquad (16.4)$$

Except for the different zero point this scale for the constant-volume hydrogen or helium thermometer differs but little from the Kelvin thermodynamic scale (Chap. 21), which is independent of the choice of property P or of the substance. A comparison of the constant-volume hydrogen thermometer with several other thermometers is given in Table 16.1. Note the close agreement between the air and hydrogen scales.

Table 16.1. *Comparison of Temperature Scales*

Const-vol hydrogen thermometer	Const-vol air thermometer	Mercury-in-glass thermometer	Platinum resistance thermometer
0°C	0°C	0°C	0°C
20	20.008	20.091	20.240
40	40.001	40.111	40.360
60	59.990	60.086	60.360
80	79.987	80.041	80.240
100	100	100	100

16.4. Linear Expansion of Solids. Very nearly all solids expand when they are heated. Steel rails are usually laid with small gaps between their ends to accommodate this expansion, and in welded streetcar tracks large thermal stresses accompany temperature changes. The expansion of a steel bridge may be allowed for by having one end slide on rollers on the abutment. If a metal rim can be placed on a wheel only when hot, its shrinkage when cooled will result in the desired tight fit. In a long steam pipe, expansion and contraction are provided for by expansion joints or by a section of the pipe bent into a U form. Many other such examples of thermal expansion might be cited.

If a rod has a length L_0 at 0°C, while L_t is its length at t°C, L being a regular function of t, it is known in mathematics that any such relationship may be represented by a series of the form

$$L_t = L_0(1 + \alpha t + \alpha' t^2 + \alpha'' t^3 + \cdots) \tag{16.5}$$

where α, α', and α'' are constants. When the measurements are extended over a considerable temperature range, all the terms in Eq. (16.5) must be retained. But for small temperature differences, say 100°C, just the first constant α is sufficient, and hence

$$L_t = L_0(1 + \alpha t) \tag{16.6}$$

The constant α, known as the *coefficient of linear expansion* of the substance, has the value

$$\alpha = \frac{L_t - L_0}{L_0 t} \tag{16.7}$$

It is to be noted that the dimensions of α are reciprocal temperature. Its value depends on the temperature scale used but not on the units with which L is measured. Some representative values of this coefficient are given in Table 16.2. The value of α for aluminum of 23×10^{-6} means that a 1-cm length of aluminum at 0°C becomes 1.000023 cm at 1°C and a 1 ft length at 0°C becomes 1.000023 ft at 1°C. A *mean* coefficient of expansion between two temperatures t_1 and t_2, defined by the equation

$$\alpha_{1,2} = \frac{L_2 - L_1}{L_1(t_2 - t_1)} \tag{16.8}$$

is often used. In a temperature range of 100° about 0°C the difference between α and $\alpha_{1,2}$ is insignificant.

Table 16.2. *Coefficients of Linear Expansion*
(Approximate values)

Substance	α (per °C)	α (per °F)
Aluminum	23×10^{-6}	13×10^{-6}
Brass	18.9×10^{-6}	10.5×10^{-6}
Copper	16.8×10^{-6}	9.4×10^{-6}
Pyrex glass	3.2×10^{-6}	1.8×10^{-6}
Ice (−10 to 0°C)	51×10^{-6}	28×10^{-6}
Invar	0.9×10^{-6}	0.5×10^{-6}
Quartz, fused	0.40×10^{-6}	0.22×10^{-6}
Steel	10.7×10^{-6}	6×10^{-6}
Tungsten	4.5×10^{-6}	2.5×10^{-6}

Inspection of Table 16.2 shows that the coefficients of expansion of all solids are very small, being especially small for invar (a nickel-steel alloy) and for fused quartz. From Eq. (16.8) a steel rail 60 ft long at −20°F increases in length 60 ft $\times 6 \times 10^{-6} \times 130 = 0.047$ ft, or 0.56 in., when the temperature rises to 110°F. Because of its small α, fused quartz would have been the best material out of which to make the 200-in. mirror of the Mt. Palomar reflecting telescope, for there would then be a minimum change in the focal length of the paraboloidal surface with change in temperature. Actually, because of technical difficulties, a borosilicate glass was used.

A bimetallic strip (Fig. 16.4) formed by riveting together a strip of brass and one of steel is used in metallic thermometers and in thermostats. Owing to the larger coefficient of expansion of brass the strip bends when heated. The movement of the pointer may be a temperature indicator, or, in the thermostat, it may make or break an electrical contact and thus control a heating unit.

In an accurate pendulum clock there must be some compensation to keep the effective pendulum length constant in spite of temperature changes. One method of accomplishing this is to construct the pendulum of a number of steel and brass bars so that the downward expansion of the steel is compensated by the upward expansion of the brass. To do this the total length of the brass must be to that of the steel inversely as their coefficients of expansion. In another method of pendulum compensation the bob includes a vessel of mercury. The expansion of the mercury raises the center of gravity, compensating the lowering of the center of gravity from the expansion of the pendulum rod.

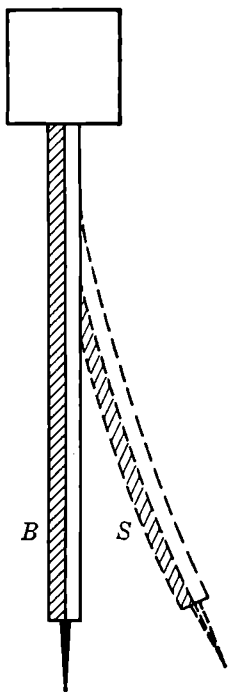


FIG. 16.4. Bimetallic strip.

The successful sealing of the metallic wires into the glass of electric lamps, radio tubes, etc., is a matter of closely matching the expansion coefficients of the metal and the glass, since otherwise cracks develop when the seal cools. Inspection of Table 16.2 shows that pyrex glass and tungsten metal have coefficients of expansion differing but little, and hence this combination is widely used today in the fabrication of vacuum apparatus.

If a metal rod is so rigidly held that it cannot expand when its temperature is raised, a compressive stress is set up in the rod. From Eq. (16.8) the fractional change in length of the rod if it could expand is

$$\frac{\Delta L}{L} = \alpha\, \Delta t \tag{16.9}$$

But the rod is not free to expand, and hence the compressive force must increase by the amount necessary to produce this same fractional change in length. Now Young's modulus is defined by

$$Y = \frac{\Delta F/A}{\Delta L/L}$$

and therefore

$$\Delta F = AY\frac{\Delta L}{L}$$

Using Eq. (16.9) for $\Delta L/L$ we have

$$\Delta F = AY\alpha\, \Delta t \tag{16.10}$$

for the increase in the compressive force needed to counteract the tendency to expand. The corresponding *thermal stress* is

$$\frac{\Delta F}{A} = Y\alpha\, \Delta t \tag{16.11}$$

16.5. Surface and Volume Expansion. Suppose a plane sheet of solid material to be homogeneous (consisting of the same substance throughout) and isotropic[1] and to have a square shape. Let the side of the square be of length L_0 at 0°C and L at t°C. Then, since $L = L_0(1 + \alpha t)$, the area $A = L^2$ of the square after expansion is related to its area $A_0 = L_0^2$ at 0° by the expression $A = A_0(1 + \alpha t)^2 = A_0(1 + 2\alpha t + \alpha^2 t^2)$. The small term $\alpha^2 t^2$ may be neglected in comparison with $2\alpha t$. Therefore

$$A = A_0(1 + 2\alpha t) = A_0(1 + \beta t) \qquad (16.12)$$

where $\beta = 2\alpha$ is the *coefficient of surface expansion.*

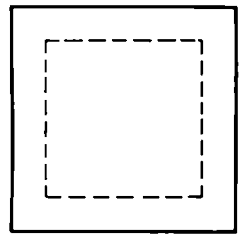


FIG. 16.5. Expansion of a frame is the same as that of a continuous plate of the same area.

If part of the area of a plate is surrounded by a dotted line as indicated in Fig. 16.5, the inscribed area will expand proportionately with the portion framing it if the whole is heated. Now if this central part had been removed before the heating, the frame would expand just as if the plate were continuous. Therefore Eq. (16.12) also gives the expansion of a hole in a surface when the latter is heated.

The expansion of a solid volume may be treated in a similar fashion. If V_0 is the volume of a cube of edge L_0 at 0°C, its volume $V = L^3$ after the temperature rise is

$$V = V_0(1 + \alpha t)^3 = V_0(1 + 3\alpha t + \cdots)$$
$$= V_0(1 + \gamma t) \qquad (16.13)$$

Since the square and cube of the small quantity αt may be neglected, the *coefficient of volume expansion* γ equals 3α. The value of γ for any solid may thus be simply computed from the linear coefficient, which is the only one usually tabulated.

A cavity in a solid volume expands at the same rate as if it were filled with the solid. Thus, a steel tank of capacity 1,000 gal at 0°C will have a volume of $1{,}000(1 + 32.1 \times 10^{-6} \times 100)$ or 1,003.21 gal at 100°C.

16.6. Expansion of Liquids. The performance of a liquid-in-glass thermometer is evidence that liquids such as mercury or alcohol have larger volume coefficients of expansion than has the glass. In general the expansion of a liquid may be represented by the same approximate relation as for solids,

$$V = V_0(1 + \gamma t) \qquad (16.13)$$

but one usually observes only the differential expansion of the liquid and its container.

[1] A mixture of flour and sugar is not homogeneous but isotropic; wood is homogeneous but not isotropic (Greek *isos*, same, + *trope*, turning; *i.e.*, having the same properties in all directions).

This container usually consists of a glass bulb with a capillary stem (Fig. 16.6) so that small changes in volume become evident. Remembering that the volume of the bulb expands as if it were solid glass, let

V_0 = vol of bulb = vol of liquid at 0°C
V_g = vol of bulb at t°C
V_l = vol of liquid at t°C
γ_g = vol coefficient of the glass
γ_l = vol coefficient of the liquid

Then,

$$V_l - V_g = V_0[(1 + \gamma_l t) - (1 + \gamma_g t)] = V_0(\gamma_l - \gamma_g)t \quad (16.14)$$

This is the *apparent expansion*, the difference $\gamma_l - \gamma_g$ being known as the *apparent coefficient of expansion* of the liquid in glass. If the volume coefficient γ_g of the glass is known, observation of the apparent expansion, V_0 and t yields the value of the coefficient γ_l for the liquid.

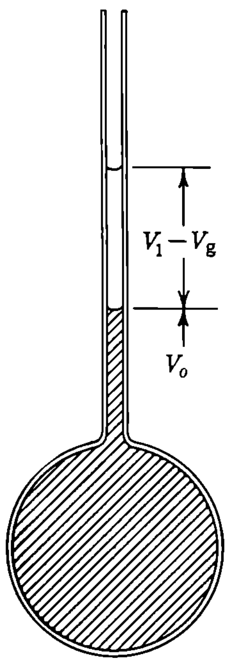


FIG. 16.6. Differential expansion of a liquid and its container.

As a volume of liquid expands, its mass remaining constant, the density of the liquid decreases. Substituting m/ρ and m/ρ_0 for V and V_0, respectively, we may write Eq. (16.13) in terms of density

$$\rho = \frac{\rho_0}{1 + \gamma t} \quad (16.15)$$

By the method of Dulong and Petit (Prob. 14) the expansion coefficient of a liquid may be obtained without reference to the expansion of the container. This method has proved of value in determining accurately the expansion coefficient of mercury, which is a substance much used in physical experimentation. If the value of γ for mercury is known, this liquid may then be used for obtaining the coefficients of expansion of solids by some differential method. In Table 16.3 are given the volume coefficients of expansion for several liquids.

16.7. Expansion of Water. Water has the anomaly of having a maximum density at 4°C at atmospheric pressure. Above 4°C it expands with increasing temperature, while below 4°C it contracts as the temperature rises. This anomalous expansion of water is plotted in Fig. 16.7. Because of this behavior, when a body of quiet water cools from contact with cold air above the surface, the water that has cooled to 4°C sinks to the bottom. It is therefore the surface water that first drops to 0°C and begins to freeze. Were it not for this property of water, our northern waters would freeze from the bottom upward and become choked with

Table 16.3. *Expansion Coefficients of Liquids*

Substance	γ (per °C)	γ (per °F)
Alcohol (ethyl)	110×10^{-5}	61×10^{-5}
Ether	164×10^{-5}	91×10^{-5}
Glycerin	51×10^{-5}	28×10^{-5}
Mercury	18.2×10^{-5}	10.1×10^{-5}
Sulfuric acid	56×10^{-5}	31×10^{-5}
Water (20°C)	20.7×10^{-5}	11.5×10^{-5}

ice. Salt water has its maximum density at a temperature several degrees lower than this value. This anomalous expansion of water is undoubtedly produced by the association of the water molecules into groups or liquid crystals near the freezing point, but as the temperature gradually rises above 4°C these groups break up into single molecules.

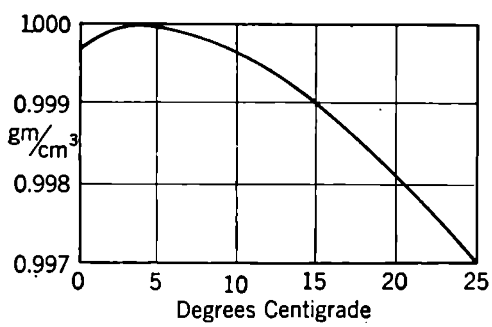


Fig. 16.7. The peculiar expansion of water.

16.8. Mercury Thermometers. Mercury is not the ideal thermometric substance, but because of their compactness and simplicity mercury thermometers are widely used. Since mercury freezes at −38.8°C, its use for low temperatures is limited. For measurement of lower temperatures, similar thermometers using alcohol (freezing point −115°C) or pentane (freezing point −131°C) are used. Mercury boils at 356.9°C at atmospheric pressure, but thermometers are made of strong, high-melting-point borosilicate glass with the space above the mercury filled with an inert gas such as argon under about 20 atm pressure to prevent the mercury from boiling. Temperatures to 550°C may be read with such thermometers.

Thermometers are either calibrated for use with the bulb plus capillary stem at the temperature read or for a specified depth of immersion. If the bulb alone is at a high temperature, the stem remaining cool, the thermometer will read several degrees too low at the highest temperatures.

Clinical thermometers have a constriction in the capillary near the bulb. In expanding the mercury pushes through this constriction; but after the maximum point is reached and the mercury is allowed to contract, the mercury column breaks at the constriction because of surface tension. The mercury column then remains at its maximum reading until vigorous shaking causes it to reunite with the mercury in the bulb.

In another form of *maximum* thermometer, which is mounted at a slight inclination with the horizontal, a small iron piece is pushed along before the mercury column and left when the mercury contracts. The lower end of the iron piece then indicates the highest point reached by the mercury column. In the accompanying *minimum* thermometer, also mounted nearly horizontal, alcohol is used. A suitably shaped light index is dragged down by the meniscus of the alcohol surface, but when the alcohol subsequently expands it flows past the index, which thus indicates the lowest point reached by the liquid meniscus in any period of varying temperatures.

PROBLEMS

1. Convert body temperature of 98.6°F and room temperature of 68°F to the centigrade scale.

2. Write an equation by which any temperature in degrees centigrade may be changed to degrees Fahrenheit. At what temperature do the Fahrenheit and centigrade scales read the same?

3. Liquid oxygen has a normal boiling point of −183°C; gold melts at 1063°C. Convert these temperatures to the Fahrenheit scale.

4. An iron steam pipe ($\alpha = 10 \times 10^{-6}$ deg C^{-1}) is 150 ft long at 0°C. What will be its increase in length when heated to 100°C?

5. Calculate the increase in length of 300 ft of copper wire when its temperature changes from 10 to 35°C.

6. A clock has a brass pendulum, which we assume to be a simple pendulum and which makes one complete vibration in exactly 2 sec at 20°C. What will be the gain or loss in time per day if the temperature of the clock drops to 0°C?

7. A steel surveyor's tape correct at 65°F is used at 90°F. What is the percentage error in lengths measured with this tape at the higher temperature?

8. A steel ring 4 in. in diameter at 20°C is to be shrunk on to a steel shaft 4.002 in. in diameter at 20°C. To what temperature must the ring be heated to just slip onto the shaft?

9. A brass plate at 70°F has in it a circular hole 10 cm in diameter. What will be the area of this hole at 200°F?

10. A pyrex glass flask has a volume of exactly 1 liter at 15°C. If it is completely filled with mercury and then raised to 75°C, how much mercury will spill out?

11. A bar of aluminum 2 by 3 by 40 cm in size is immersed in water at 4°C. What will be the change in the buoyancy of the water on the bar if the temperature of the water rises to 25°C?

12. A copper bar of cross section 0.5 in.^{2} is heated from 20 to 100°C. What force would be necessary to prevent it from expanding?

13. If steel rails are welded together and held so that no expansion is possible, what is the increase in the stress in the rails when the temperature rises from 50 to 100°F?

14. Two vertical columns filled with a liquid are interconnected at their lower ends by a horizontal capillary tube. One of the columns is surrounded by a jacket containing ice and water, while a high-temperature bath encloses the other. If the height and density of the liquid in the two columns are denoted by h_0, ρ_0 and h_t, ρ_t respectively, show that the volume coefficient γ of the liquid is given by $\gamma = (h_t - h_0)/h_0 t$ (method of Dulong and Petit).

15. If a steel scale is to be ruled so that all millimeter intervals are correct to ±0.001 mm at the specified temperature, what is the maximum variation in the temperature that may occur during the ruling?

CHAPTER 17

CALORIMETRY

17.1. Heat Is Energy. That heat is a form of motion is an old idea, held, for example, by Francis Bacon, Boyle, and Hooke in the seventeenth century. Toward the end of the eighteenth century, however, heat was commonly believed to be an imponderable fluid, called *caloric*, which passes into a body when its temperature rises and which the body loses when its temperature drops. We have inherited from that period our terminology of heat. Although we now know that a rise in temperature means merely an increase in the kinetic and potential energy of the molecules of the material, we continue to use the definitions of heat quantities developed by Lavoisier (1743–1794), Joseph Black (1728–1799), and other eighteenth-century scientists.

In 1798 Count Rumford (Benjamin Thompson, an American Tory who fled to Europe at the time of the Revolution) concluded from qualitative experiments performed while he was in charge of cannon boring for the Elector of Bavaria that heat cannot be a substance. Observing the large amounts of heat generated in the boring operations, whether the tool was sharp or dull, Rumford decided that heat must be a form of motion. Soon after this Sir Humphry Davy (1778–1829) showed that, when two pieces of ice are rubbed together under conditions of no heat transfer from outside sources, enough heat is generated by the work done against the frictional force to melt the ice. Although the "calorists" still maintained for some time that their theory of heat was correct, the idea that heat is a form of energy gradually prevailed and became firmly established by the long series of experiments of James P. Joule from 1840 to 1878, which showed that there is an exact equivalence of heat and work. These experiments of Joule were careful, quantitative researches in which he transformed work into heat in a large variety of ways. For example, descending weights caused various surfaces to rub against each other, the heat generated being absorbed by a given quantity of water. A given amount of work done against friction always produced the same quantity of heat.

Since there is no evident change in the kinetic or potential energy of a body as a whole when it is heated, the energy given to the body must be taken up by its molecules. These are in incessant motion at any temperature, but as the temperature of the body is raised not only do its molecules move about with increased translational speeds but also the atoms in the molecules have greater vibrational and rotational energies.

The expansion of solids and liquids with heating indicates that there is also a gain in the potential energy of the molecules, for work must be done against the intermolecular attractive forces when the average distance between molecules is increased.

The conduction of heat through a body when its boundaries are warmed is easily explained by this kinetic theory of heat. For the molecules are continually colliding with each other, and the slower moving molecules gain energy in this way in their impacts with faster moving molecules. The molecules of a gas in striking the heated walls of their container rebound with greater velocity, and they pass some of this increased energy on to other gas molecules by collisions.

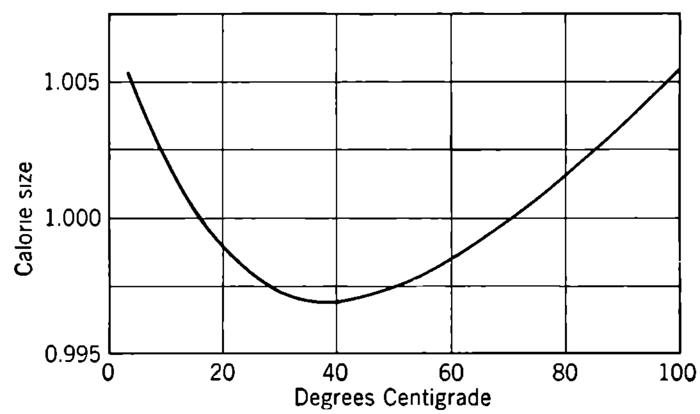


FIG. 17.1. Relative values of the calorie at different temperatures.

It is necessary to distinguish clearly between heat and temperature. With a given gas flame it will take much longer (*i.e.*, take much more heat) to warm up a pail of water than is needed to warm a cup of water to the same temperature. The temperature of a body is merely a number on some arbitrary scale expressing how hot the body is as compared with some standards. Heat, on the other hand, is a form of energy and may be measured in energy units.

17.2. The Unit Quantity of Heat. We use today the same heat unit employed by the calorists, the quantity of heat that will raise the temperature of unit mass of water one degree. Since the amount of heat required to raise the temperature of unit mass of water one degree is not the same at all temperatures (cf. Fig. 17.1), it is necessary to specify a temperature in the definition. The *calorie* is defined as *the quantity of heat required to raise one gram of water from* 14.5° *to* 15.5°*C*. The *mean calorie*, defined as one one-hundredth of the quantity of heat required to raise the temperature of 1 gm of water from 0 to 100°C, is practically the same as the 15° calorie. This calorie is often referred to as the "gram-calorie" or the "small calorie." In dietetics and most other calorimetric studies the *kilocalorie*, larger than the gram calorie by a factor of 1,000, is employed. The *British thermal unit* (Btu) *is the quantity of heat required to raise the temperature of one pound of water* through 1°F. It is easily shown that 1 Btu = 252 gm cal. We shall now drop the prefix *gram* in referring to the small calorie, while retaining the term *kilocalorie* for the larger unit.

Using these definitions of heat quantity, we compute, for example, that to raise 10 kg of water from 15 to 65°C would take 10 × 50 or 500 kcal, or 500,000 cal. Similarly, to raise 10 lb of water from 60 to 200°F would require 10 × 140 or 1,400 Btu.

17.3. Mechanical Equivalent of Heat. In Joule's experiments it was observed that if a given amount of mechanical energy, no matter how generated, was all transformed into heat in a constant amount of water the temperature rise of the water was always the same. The quantity of mechanical energy that, if entirely converted into heat, could raise the temperature of unit mass of water one degree is known as the *mechanical equivalent of heat*. This is usually designated by the letter J. The accepted value of this important constant is

$$J = 4.186 \text{ joules/cal}$$

Therefore 1 joule = 0.239 cal, and conversion to the English system of units gives 1 Btu = 778 ft lb.

Suppose that a 3,000-lb automobile moving with a speed of 60 miles/hr is brought to rest by application of its four-wheel brakes. The kinetic energy of the car is

$$\frac{3{,}000 \text{ lb} \times (88 \text{ ft/sec})^2}{2} = 11{,}616{,}000 \text{ ft poundals} = 363{,}000 \text{ ft lb}$$

This is transformed into 363,000/778 or 467 Btu in the brakes, or 117 Btu per brake.

17.4. Heat Capacity and Specific Heat. The quantity of heat needed to raise one gram of a substance one degree centigrade is different for every substance. One calorie will raise about 30 gm of mercury or platinum or about 10½ gm of copper through one degree centigrade. The *heat capacity of a body* is defined as the number of heat units needed to raise the temperature of the body through one degree. *Specific heat c is defined as the heat capacity per unit mass.* The unit of specific heat is thus the calorie per gram per degree centigrade or the Btu per pound per degree Fahrenheit. That is, the specific heat of iron is 0.12 cal gm^{-1} deg C^{-1} and is also 0.12 Btu lb^{-1} deg F^{-1}. For water the specific heat has the value 1 cal gm^{-1} deg C^{-1} only for the range 14.5 to 15.5°C. In Table 17.1 we list the specific heats of some common solids and liquids. It is necessary to specify the temperature or the temperature range for the specific heat values given, for in general the specific heat of any substance increases with its temperature. We have already mentioned the small variation in the specific heat of water over the range 0 to 100°C. For other liquids the specific heats vary appreciably with temperature. Ethyl alcohol, for example, has a specific heat of 0.456 at −100°C, but at +100°C this value has risen to 0.824. For most solids, however, the variations of the specific heats with temperature if above 200°K are so small that they may be neglected.

Table 17.1. *Specific Heats*
(Cal gm^{-1} deg C^{-1} or Btu lb^{-1} deg F^{-1})

Substance	Temp, °C	Specific heat
Alcohol, ethyl	25	0.581
Aluminum	20	0.214
Copper	20	0.092
Glycerin	0	0.540
Ice	−20 to 0	0.48
Iron (cast)	20 to 100	0.119
Lead	20	0.031
Mercury	20	0.033
Steam (atmospheric pressure)	100 to 200	0.48
Tin	18	0.054
Tungsten	20 to 100	0.034

In an experiment to measure the heat capacity of a specimen of material a finite range of temperature must be used. Therefore a *mean* specific heat for the temperature interval is determined. To raise a mass m of material of mean specific heat c from temperature t_1 to temperature t_2 requires a quantity of heat

$$Q = mc(t_2 - t_1) \tag{17.1}$$

This will also be the quantity of heat given out by this mass if it drops in temperature from t_2 down to t_1. The specific heat c_t at the temperature t would be defined by the relation

$$c_t = \frac{1}{m}\frac{dQ}{dt} \tag{17.2}$$

where dQ is the small quantity of heat liberated or absorbed by the mass m for an infinitesimal temperature change dt covering this temperature. In practice c_t may be taken as equal to c for a moderate-sized temperature interval about the temperature t.

If a 100-gm piece of cast iron is heated from 20 to 100°C, it will absorb $100 \times 0.119 \times 80$ or 952 cal. In cooling from 100 down to 20°C this iron will lose this same number of calories of heat to its surroundings.

***17.5. Internal Energy.** Why should the heat capacities of various substances be so different, and why should all specific heats decrease markedly as the temperature drops toward absolute zero? At −253°C the specific heat of copper has dropped to 0.0031, that of aluminum to 0.0039, while at −233°C carbon in the form of diamond has a specific heat nearly zero (0.0005). In seeking answers to these questions it must be remembered that the heat energy stored in a body is the total energy of all its molecules, which differ in mass for each substance and which can move only in limited paths in a solid or liquid.

The differences in the number of molecules that make up unit mass for various substances should affect their specific heats, as will also differences in energy internal to the molecules. This "internal energy" includes the vibrational or translational kinetic energy of the molecules, the kinetic and potential energy of the atoms and their electrons, which compose the molecules, and the potential energy accompanying the increase in the average distance between molecules when the body expands. Because the expansion of solids and liquids with increasing temperature is so small, the work done in the expansion against atmospheric pressure and the last-mentioned contribution to the internal energy may be neglected.

A seemingly strange observation was made in 1818 by Dulong and Petit, called the *law of atomic heats.* They found that the product of the specific heat and atomic weight, which has been named *atomic heat capacity,* is the same for all the solid elements. This relation is only approximate, but for most of the solid elements, excluding Be, B, C, and Si, the atomic heat capacities lie in the range 5.8 to 6.8 cal gm-atom^{-1} deg^{-1} and group themselves about a mean value 6.3. This near constancy of the atomic heats of all the solid elements indicates that the energy associated with the vibrations of the atoms about their equilibrium positions in the crystal lattice constitutes the main heat content of a solid. Such vibrational motion may be resolved into components along three mutually perpendicular directions, and a vibration along any one of these directions has no components in the other two directions. The atoms in a solid thus are said to have three *degrees of freedom.* By the reasoning of the kinetic theory (Chap. 20) each vibrational degree of freedom has 2 cal associated with it, so that the atomic heat of a solid should be 6 cal gm-atom^{-1} deg^{-1}, or close to the mean experimental value.

Figure 17.2 shows the manner in which atomic heats of solids fall off with decreasing temperatures. The curves for all other solid elements lie between those for lead and carbon. As the temperature approaches absolute zero, the heat capacities of all solids drop toward the value zero. The explanation of this behavior is given by the *quantum theory,* according to which the energy associated with a frequency of oscillation f is hf, where h is *Planck's universal constant of action.* The characteristic frequency f of an atom will not be excited at all unless the energy it receives from an impact equals the *quantum* hf. Furthermore, this energy hf might be sufficient to start one type of vibration but not others. At very low temperatures the impacts may not be energetic enough to excite any vibrations. The low values of the atomic heats of the light elements Be, B, C, and Si at room temperatures are caused by the fact that, since the light atoms have large vibrational frequencies, a correspondingly higher temperature is necessary to excite by impacts all the modes of vibration.

17.6. Heat of Fusion and Heat of Vaporization. Although we shall postpone to the following chapter a detailed discussion of the change of a substance from one state or phase to another, we must define at this point, because of their use in calorimetric measurements, the terms *latent heats* of fusion and vaporization. *Black* (1728–1799) was the first to

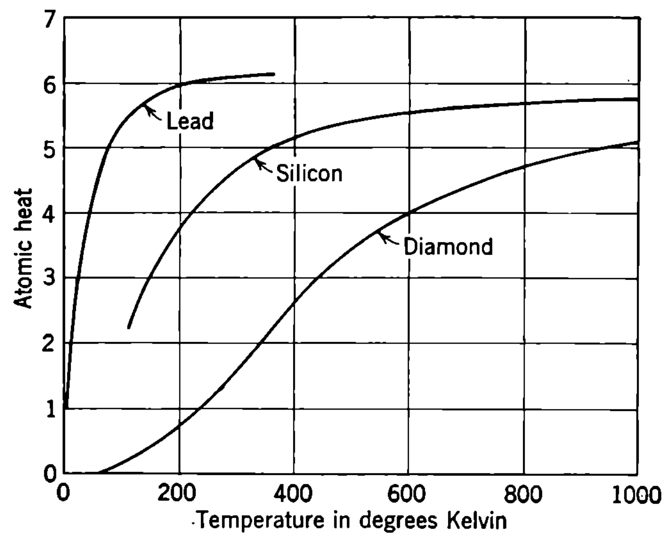


Fig. 17.2. The atomic heats of solids drop toward zero as the temperature approaches absolute zero.

show that while ice is melting or water is freezing the mixture remains at constant temperature, as does also water boiling under constant pressure or steam condensing into water. However, large quantities of heat are absorbed in the melting and vaporization processes and are released in the freezing and vapor-condensation processes; Black called these *latent heats.* The energy of motion gained by the molecules of the solid or liquid from the heat source is changed in these transformations immediately into potential energy; *i.e.*, it is consumed in working against the intermolecular forces that bind the molecules close together in the solid or liquid states.

The quantity of heat that is absorbed by unit mass of a solid to melt it without change of temperature is called the *heat of fusion.* Then the quantity of heat Q that will melt a mass m at constant temperature is

$$Q = mL \tag{17.3}$$

where L is the heat of fusion of the substance. For water under normal atmospheric pressure the melting point is 0°C or 32°F. One gram of ice in melting under these conditions absorbs 80 cal from its surroundings, while a gram of water in freezing gives out 80 cal to its surroundings. That is, the heat of fusion of water is 80 cal/gm, the corresponding value in the English system of units being 144 Btu/lb.

Similarly the quantity of heat that must be supplied to unit mass of a liquid to convert it to vapor without change of temperature is called the *heat of vaporization*. Equation (17.3) may be used to calculate the quantity of heat Q needed for this vaporization process, if L now denotes the heat of vaporization of the substance. At normal atmospheric pressure of 76 cm of mercury, water boils at 100°C, and it takes 539 cal to vaporize a gram of water under these conditions. The heat of vaporization of water is thus 539 cal/gm, the corresponding value in the English system being 970 Btu/lb.

17.7. Calorimetry. The Method of Mixtures. One of the earliest methods of measuring heat quantities is the method of mixtures. The principle of this method is simple: if heat interchange between two bodies at different temperatures takes place in a thermally insulated enclosure, the heat lost by the warmer body equals that gained by the colder body. Such an experiment would be very simple, merely weighing the two bodies and reading the initial and final temperatures, were it not for the necessity of taking considerable care to minimize or measure the heat losses from the vessel, known as a *calorimeter*, in which the heat interchange takes place. The following example illustrates the application of this principle in the determination of a specific heat:

Worked Example. A copper calorimeter cup weighing 200 gm contains 250 gm of water at 17°C. Small pieces of a metal of unknown specific heat c totaling 300 gm are heated to 99°C and then dropped into the water. The final temperature of the mixture is 19.8°C. Equating the heat lost by the metal to that gained by the water and calorimeter,[1]

$$\begin{aligned} 300 \times c \times (99 - 19.8) &= 250 \times 1 \times (19.8 - 17) + 200 \times 0.092 \\ &\qquad \times (19.8 - 17) \\ &= (250 + 18.4) \times 2.8 \end{aligned}$$

giving $c = 0.031$ cal gm^{-1} deg C^{-1}. The product of the mass and the specific heat of the calorimeter, 18.4 cal/deg in this example, is known as

[1] In this example and in others to follow, the units, which are cumbersome but simple, will not be consistently carried through as was done in the chapters on mechanics. The student is asked, however, to insert units everywhere and to verify their consistency.

its *water equivalent.* That is, a given quantity of heat will raise this calorimeter or 18.4 gm of water through the same temperature difference.

An apparatus for the measurement of the specific heat of a metal is shown in Fig. 17.3. The sample S is heated in a steam jacket to a temperature but little less than that of the live steam. It is then dropped quickly but carefully into the calorimeter C of known water equivalent, filled with a known mass of water. Any additional piece of equipment that comes into contact with the mixture, such as the stirrer, must also be represented in the heat equation.

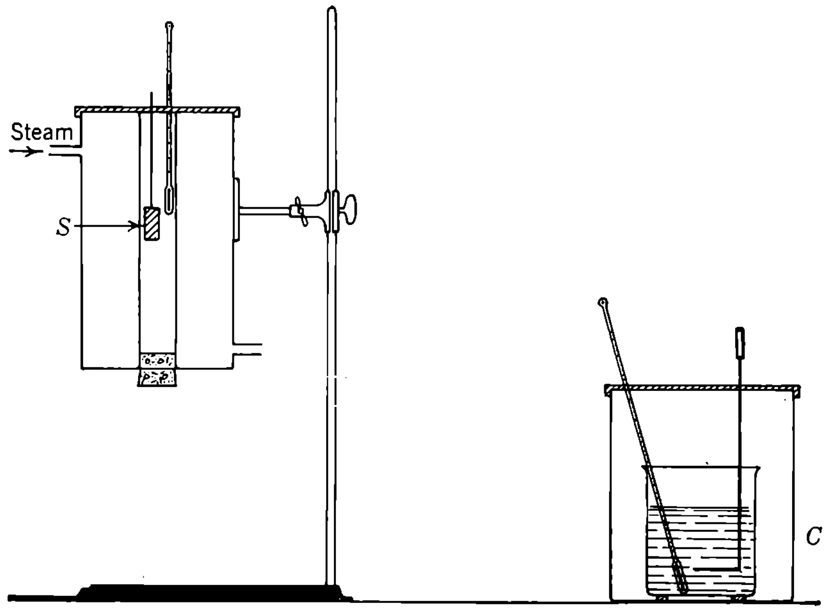


FIG. 17.3. Heater and calorimeter for measurement of the specific heat of metals by the method of mixtures.

The precision of such an experiment is ordinarily not high because of heat losses. To minimize these losses the calorimeter may have a bright, nickel-plated surface to reflect back the heat radiation, and it is always placed within another vessel to limit heat conduction to surroundings. Another good precaution is to choose the initial and final temperatures of the water so that, during the time for the mixture to come to the equilibrium temperature, heat is at first absorbed from the surroundings and then toward the end of the run given out to the surroundings. Since in this experiment the rise through the last degree of temperature takes much longer than that through the first degree, the final temperature of the mixture should be about that of the room or but slightly above it.

17.8. Latent Heats by Methods of Mixture. The latent heat of fusion of ice may be measured by dropping a piece of ice, dried with a towel to remove the water film from its surface, into slightly warmed water in a calorimeter. From the minimum temperature reached by the mixture the heat lost by the water in melting the ice may be calculated. If the initial temperature of the water is about as much above room temperature as the final temperature is below it, the heat losses are minimized.

Worked Example. Dried ice at 0°C was dropped into a calorimeter of water equivalent 10 cal deg^{-1} and containing 190 gm of water at 30°C. The minimum temperature reached by the mixture was 9.8°C, and the water content was then found to be increased by 45 gm. Therefore, $45L + 45 \times 9.8 = 200 \times 20.2$, giving $L = 80$ cal/gm.

The heat of vaporization of water may be determined by injecting saturated steam into a quantity of water, below room temperature, in a calorimeter. The initial and final temperatures of the water are noted, and the amount of steam condensed is computed from the gain in weight of the contents. To the latent heat given out by this mass of steam in condensing at the steam temperature must be added the heat gained by the cold water from this condensate cooling to the final temperature of the mixture. The sum of these two terms is equated to the mass of the cold water times its gain in temperature.

***17.9. Specific Heat of a Liquid by the Method of Cooling.** If two identical containers are filled, one with water, the other with a liquid of unknown specific heat, and both are heated to the same temperature, the lengths of time for the two to cool down to the same lower temperature are proportional to their respective heat capacities. For the rate at which heat is given off from identical containers of the two liquids at the same temperature will be the same. Hence the rate at which the temperature drops will be more rapid for the container of lower heat capacity. Natural convection currents in the liquids serve to keep the temperatures of the cooling liquids uniform enough throughout their volumes.

If the liquid of unknown specific heat c has a mass m, its heat capacity is mc. Suppose the mass of water in the other container to be w and its specific heat to be unity. Let the mass of each container be m' and its specific heat be c'. Then if the unknown liquid requires t_1 sec to cool and the water t_2 sec to cool between the same two temperatures,[1]

$$\frac{mc + m'c'}{w + m'c'} = \frac{t_1}{t_2} \tag{17.4}$$

or

$$c = \frac{wt_1 + m'c'(t_1 - t_2)}{mt_2} \tag{17.5}$$

There is a useful approximate relation, first proposed by Newton in 1701 and known ever since as *Newton's law of cooling*, governing the rate at which a hot body loses heat to its surroundings. According to this law of cooling the rate of loss of heat from a body is proportional to the difference in temperature between the body and its surroundings. If the heat capacity of the body is constant, then

[1] Note that in this section t denotes time, not temperature.

the time rate of cooling is also proportional to the temperature difference. Therefore, if T is the temperature of the body and T_0 that of its surroundings,

$$-\frac{dT}{dt} = k(T - T_0) \tag{17.6}$$

where k is a constant involving the heat capacity of the body. Integrating Eq. (17.6) we obtain

$$\ln (T - T_0) = -kt + C \tag{17.7}$$

in which C is the constant of integration. If at $t = 0$ the temperature of the body is T_1, $C = \ln (T_1 - T_0)$, and consequently

$$\ln \frac{T - T_0}{T_1 - T_0} = -kt \tag{17.8}$$

A plot of $\ln (T - T_0)$ against the time t should be a straight line if Newton's law of cooling is correct. Actually this law is a fair approximation only when the temperature difference $T - T_0$ is but a few degrees. However, the nature of the law of cooling in no way affects the principle on which this method of determining the specific heat of liquids depends: Eq. (17.4) is still true.

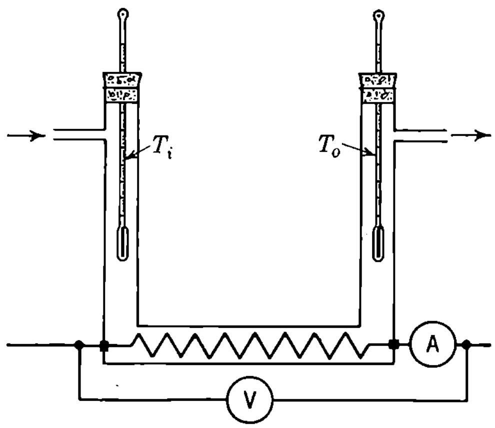


FIG. 17.4. Arrangement for determination of the specific heat of a fluid by the method of continuous flow using electrical heating.

***17.10. Continuous-flow Method.** In modern calorimetry electrical heating is much used, since electric power may be accurately measured and the relationship between the unit of electric energy and the calorie is well established (Joule's equivalent). With electrical heating, the continuous-flow method as sketched in Fig. 17.4 is an excellent way to measure the specific heat of a fluid. In the steady state, with constant electric power and uniform flow of fluid, a constant temperature difference $(T_i - T_o)$ will exist between the input and output fluid. If ρ is the density of the fluid, c its specific heat, and dV/dt the volume of fluid passing per second, then

$$\rho c \frac{dV}{dt} (T_i - T_o) = \frac{\text{volts} \times \text{amp}}{J} - \text{losses} \tag{17.9}$$

where the product volts $\times$ amp is the electric power in watts (Chap. 30). The effect of losses may be calculated by using different rates of flow but with the rise in temperature always the same.

The continuous-flow method may also be used to measure the *heat of combustion* of gaseous fuels. The heat of combustion of any substance is defined as the amount of heat evolved per unit mass (or per unit volume of a gas) upon complete burning or oxidation. A steady stream of water is passed through the calorimeter, which is so arranged (cf. Fig. 17.5) that all the heat from the burning fuel is absorbed. The heat of combustion may be calculated from the steady-state rise in temperature of the water, its rate of flow, and the rate of consumption of the fuel.

Measurement of the heat of combustion of a solid substance may be made with a *bomb calorimeter*. The specimen is placed together with oxygen under several

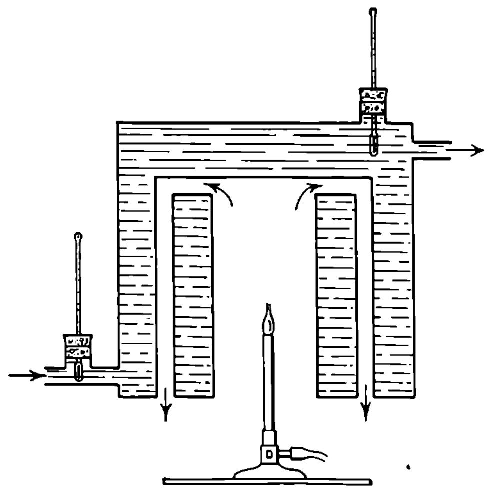


FIG. 17.5. Continuous-flow calorimeter for measuring heat of combustion of gaseous fuels.

atmospheres pressure in a stout-walled container or bomb. The bomb is then placed in water in a calorimeter, the fuel is ignited by a wire made incandescent by an electric current, and the rise in temperature of the water noted. In Table 17.2 are given the heats of combustion of a few substances.

Table 17.2. *Heats of Combustion*

Substance	Cal/gm	Btu/lb.	Btu/ft³
Acetylene	11,800	21,200	1,456
Coal, bituminous	5,000–7,000	10,000–14,000	
Coke	6,900	12,600	
Gasoline	11,500	20,750	
Furnace oil	10,500	19,000	
Methane	13,000	23,550	995
Coal gas			634
Natural gas	....	....	1,050
Fats, animal	9,500	17,000	

***17.11. Heat of Formation and Solution.** In all chemical reactions heat is either evolved or absorbed. The number of calories of heat absorbed or liberated per gram-molecular weight (mole) of compound formed from the elements under the condition of constant volume is called the *heat of formation*. Oxidation is, of course, a chemical reaction, and therefore the heat of combustion of a compound may be calculated from the heat of formation of the compound and the heats of formation of the products of combustion. The heats of formation and combustion in kilocalories per mole for a few compounds are collected in Table 17.3.

Table 17.3. *Heats of Formation and Combustion*
(kcal/mole at 20°C)

Compound	Formula	Heat of formation	Heat of combustion
Acetylene	C_2H_2	−54.3	310
Carbon (to CO_2)	C	......	94
Carbon dioxide, gas	CO_2	94.4	
Ethyl alcohol, liquid	C_2H_5OH	65.9	341
Carbon disulfide, gas	CS_2	−28.7	265
Methane	CH_4	20.3	210.9
Water:			
Liquid	H_2O	68.4	
Gas	....	57.8	

As an example of the relation between the heat of combustion of a compound and the heats of formation involved in the reaction, consider the combustion of methane according to the equation,

$$CH_4 + 2O_2 \rightarrow CO_2 \text{ (gas)} + 2H_2O \text{ (liquid)}$$

Adding the heats of formation for CO_2 and $2H_2O$ and then subtracting the heat of formation of CH_4, we get for the heat liberated in this reaction

$$94.4 + 2(68.4) - 20.3 = 210.9 \text{ kilocal/mole.}$$

If the heat of formation of a compound is negative (heat is absorbed when the compound is formed), as in the case of acetylene, the heat of combustion is usually large, for, upon burning, the heat given out includes this heat of formation as well as the heats of combustion of the elements carbon and hydrogen.

When a compound dissolves in a solvent, heat is either evolved or absorbed. This is called the *heat of solution* and is usually expressed in kilocalories per mole of compound at an indicated temperature. The number of water molecules to one of the substance is also usually specified; and if it is not, the dilution is understood to be such that any additional dilution produces no appreciable heating effects. If concentrated sulfuric acid is poured into water, the solution gets warm and the heat of solution is therefore positive (17.75 kilocal at 18°C for 200 water molecules to 1 of acid). If, on the other hand, ammonium chloride is dissolved in water, the solution becomes cooler, and the heat of solution is negative (−3.895 kilocal, also at 18°C and for 200 water molecules to 1 of the NH_4Cl).

PROBLEMS

Conversion factors: 1 kcal = 1,000 cal.; 1 Btu = 252 cal = 778 ft lb; 4.186 joules = 1 cal; 1 joule = 0.239 cal; heat of fusion of water = 80 cal/gm = 144 Btu/lb; heat of vaporization of water = 539 cal/gm = 970 Btu/lb.

1. How many calories are needed to raise the temperature of 1.2 kg of aluminum from 10 to 100°C? How many Btu are required to raise the temperature of 3 ft³ of water from 40 to 200°F?

2. A mass of 150 gm of aluminum is heated to 90°C and placed in 100 gm of a liquid at 12°C, contained in a copper calorimeter whose mass is 30 gm. The temperature of the liquid rises to 40°C. Find the specific heat of the liquid.

3. A quantity of lead shot weighing 460 gm is heated in a steam jacket to 99°C and then dropped into 200 gm of water at 16°C, contained in a copper calorimeter cup of mass 80 gm. What is the final temperature?

4. A quantity of ethyl alcohol at 30°C was mixed with some water at 12°C. The resulting temperature was 20°C. What was the ratio of the mass of alcohol to that of the water?

5. What is the result of mixing in equal parts by weight ice at 0°C and water at 60°C?

6. What is the conversion factor to change a heat of combustion in calories per gram to Btu per pound?

7. How many Btu would be required to change 1 lb of ice at −20°C to steam at 150°C and atmospheric pressure?

8. The gas company in a certain city supplies a mixture of coal gas and natural gas having a heat of combustion of 850 Btu/ft³. If in a hot-water system 0.3 of the fuel heat is wasted, how much of this gas will be required to heat the 6 ft³ of water needed for a bath from 50° to 105°F?

★9. Using a flow calorimeter with electric-power input of 200 watts, it was observed that, with a flow rate of 1,200 cm³/min of a liquid of density 0.92, a steady-state temperature difference of 5.5°C existed between the input and output liquid. Calculate the specific heat of the liquid.

★10. In a calorimeter using the method of cooling, the copper container weighs 100 gm. When it is filled with 300 gm of water, 800 sec is required for cooling between two observed temperatures. When it is filled with 275 gm of another liquid, 350 sec is required for cooling between the same two temperatures. What is the specific heat of the liquid?

★11. In a bomb calorimeter made of steel and weighing 2 kg there are 2 kg of water. Five grams of coal is burned, causing a rise in temperature of 18°C. Taking the specific heat of steel as 0.107, calculate the heat of combustion of the coal.

★12. A body is initially at a temperature of 60°C, while its surroundings are at 50°C. After 5 min its temperature is 56°C. When will its temperature be 53°C?

CHAPTER 18

CHANGE OF STATE

18.1. Solid, Liquid and Gaseous States of Matter. From X-ray studies it is known that in crystalline solids the atoms are all located at definite points in a lattice arrangement. The atoms vibrate about these lattice points, the amplitude of the vibration increasing with the rise in temperature. At the melting point, which occurs at a fixed temperature different for each crystalline substance, the amplitudes of the vibrations have become so large as to disturb the orderly arrangement of the atoms. Heat energy absorbed by the solid during the melting process is used in overcoming the attractive forces between the molecules and so producing increased molecular potential energy. Amorphous substances such as wax, tar, and glass, as well as most alloys, do not have sharp melting points but rather pass from the one state or phase to the other over a range in temperature of a few degrees. The change from the solid to the liquid state is known as *melting*, or *fusion;* the reverse change, as *freezing.* During freezing the molecules in settling down into their lattice positions have their potential energy changed back into kinetic energy, which is given as heat energy to the immediate surroundings.

When a liquid changes to the gaseous state, a process known as *vaporization*, there is a further, much larger increase in the average distance between molecules. Heat energy is absorbed by the liquid to supply this increase in potential energy, or internal energy, of the substance. In addition, a portion of the absorbed heat energy is consumed in doing *external work* in the expansion against the pressure confining the substance. In the reverse change from the gaseous to the liquid phase, known as *condensation*, energy must be given up as heat energy to the surroundings.

The transition directly from the solid state to the gaseous state, called *sublimation*, also occurs. It takes place, for instance, when "dry ice" at atmospheric pressure changes to carbon dioxide gas without going through the liquid phase. To produce sublimation there is, of course, an absorption of heat energy in order both to increase the potential energy of the molecules and to do the external work in the large expansion against the pressure of the atmosphere.

18.2. Melting and Freezing. Every pure crystalline solid has under the usually existing condition of atmospheric pressure a definite fixed melting point. The mixture of solid and liquid remains at this constant temperature, continually absorbing heat from the surroundings, until all the solid is melted. When the pure liquid cools to this same temperature, freezing begins and the mixture stays at this temperature, giving out heat to the surroundings until the mass is completely solid. By placing tubs of water in their fruit cellars on cold nights, farmers sometimes make use of the fact that sufficient heat is given off by the water in freezing (80,000 cal/kg!) to prevent damage to the fruit. As already mentioned, the amount of heat absorbed in melting or evolved in freezing per gram of substance is called the heat of fusion. This quantity may be measured by the calorimetric methods discussed in the preceding chapter. In Table 18.1 are given the melting points and heats of fusion for a number of crystalline substances. Note that the ratio of cal/gm to the corresponding Btu/lb is 5/9.

Table 18.1. *Melting Points and Heats of Fusion*

	Melting point		Heat of fusion	
	°C	°F	cal/gm	Btu/lb
Alcohol, ethyl	−114.4	−174	24.9	44.8
Aluminum	658	1,216	76.8	138.2
Copper	1,083	1,980	42.	75.6
Lead	327	621	5.86	10.55
Mercury	− 39	− 38	2.82	5.08
Nitrogen	−210	−346	6.09	10.96
Oxygen	−219	−362	3.30	5.94
Platinum	1,755	3,190	27.2	49.0
Silver	961	1,762	21.07	37.9
Tin	232	450	14.0	25.2
Water	0	32	79.71	144

Most substances contract on freezing, so that the solid sinks in the liquid. Water, however, expands when it freezes, resulting in such effects as the bursting of water pipes, rupture of plant cells, and splitting of rocks. In addition to water, iron, bismuth, antimony, and a few alloys such as type metal (58% Pb, 26% Sn, 15% Sb, 1% Cu) expand on solidifying. Because of this property excellent castings result when these metals in the molten state are poured into molds.

Even though a fixed melting point exists, there is no sharp transition from the crystalline state to the liquid state in which all the molecules are in random translational motion. X-ray analysis shows that near the melting point the solid is not entirely crystalline, while just after melting there is still much crystalline structure in the liquid. It has also been revealed by X rays that many amorphous solids are really made up of microscopic crystals. Other amorphous solids, such as glass, are *undercooled liquids* of very high viscosity.

When a crystalline substance is heated until it is liquid at a temperature well above the melting point and it is then allowed to cool, a plot of its temperature against time is known as a *cooling curve*. Such a curve for pure iron is shown in Fig. 18.1. At 1539°C the liquid freezes, as is indicated by the horizontal portion of the curve at this temperature. But this is not the only transition point, for in cooling to room temperature the solid goes successively through four crystalline forms, or phases. As the iron in cooling passes from one of these phases into the next, there is a rearrangement of the atoms, with accompanying evolution of heat, similar to the heat of fusion. This is easily observed in a length of iron wire heated by an electric current to a bright orange-yellow color. On cooling, as it approaches 900°C it is red in color, but at about this temperature it suddenly brightens, indicating transition from the γ to the β crystalline structure. Such changes in the arrangement of the atoms in their lattice structure have been observed for many crystalline solids from these transition points in the cooling curves.

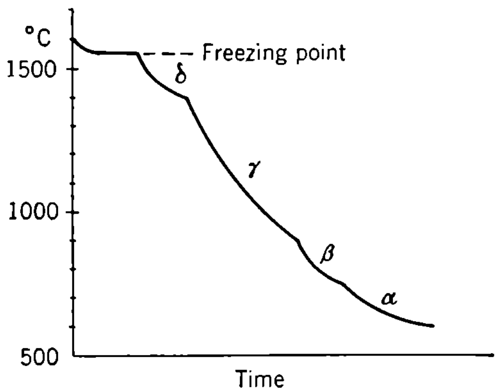


FIG. 18.1. Cooling curve for iron.

Any cooling liquid if quiescent and free from impurities may be *undercooled* several degrees below its normal freezing point without solidification. This condition is unstable, however; and if a particle of the solid is dropped into the undercooled liquid or if the liquid is agitated, crystallization immediately occurs. There will then be such a rapid evolution of the latent heat of fusion that the mixture will usually rise to a temperature several degrees above the freezing point before settling back to that temperature as the solidification becomes complete.

18.3. Effect of Pressure upon the Melting Point. On a substance, such as water, which contracts on melting, external work is done during the melting process. Since greater pressure increases this external work, the melting will thereby be aided, so that it will occur at a lower temperature. If, however, as with most other solids, expansion accompanies the melting, increase in pressure will raise the melting point. The effect of increase in pressure on the melting point of water is represented in Fig. 18.2. This lowering of the melting point is not large, for to drop the melting point 1°C the pressure must be raised to about 135 atm.

This curve represents the only conditions of pressure and temperature under which the solid and liquid can exist together in equilibrium. The region of this diagram to the left of the curve represents the solid state, and for equilibrium for any combination of pressure and temperature lying to the right of the curve the substance must be entirely in the liquid state. Following the lowering of the melting point with great increase in pressure, it has been found that a minimum value of −22°C occurs at about 2,500 atm pressure. For even greater pressures P. W. Bridgman has shown that water is transformed back into the solid phase again, and at extremely high pressures he has demonstrated that ice exists in at least five different forms. For one of these forms the melting point at about 20,000 atm is 80°C.

If a wire with weights at each end is looped over a block of ice, it slowly melts its way through the ice, which freezes again above the wire. This process is known as *regelation*. The ice immediately *below* the wire, being under increased pressure, but at 0°C, is above its melting point. A little of it therefore melts, with absorption of heat, below the wire. This water at slightly below 0°C flows around the wire, above which it freezes again because in that region the pressure is normal. The heat used up in melting below the wire is provided by the heat of fusion of the water freezing above it. The formation and flow of glaciers are an example of regelation.

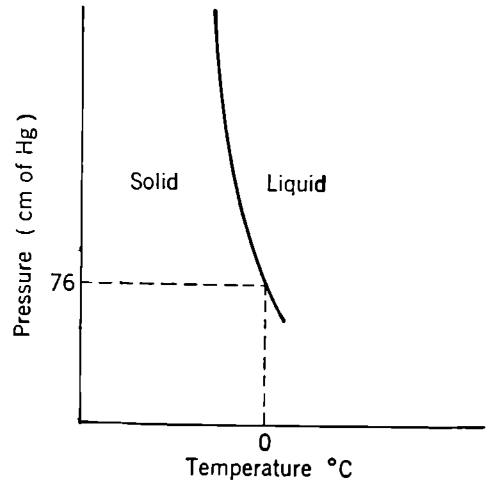


FIG. 18.2. Effect of pressure on melting point of water.

***18.4. Impurities and the Melting Point.** The presence of even a slight amount of any impurity lowers the melting point of a substance. Thus sea water has a lower freezing point than pure water. For dilute solutions the lowering of the freezing point is approximately proportional to the amount of the dissolved material. If the temperature of a solution is lowered beyond its freezing point, the pure solvent only will at first freeze out of the solution. An antifreeze mixture of alcohol and water begins to freeze as a mushy mixture of ice and liquid. Upon lowering the temperature further the alcohol content of the remaining solution rises until a certain concentration is reached, and then the mixture crystallizes. Such a mixture is called a cryohydrate.

This effect of impurities in lowering the melting point is in accord with our discussion of the nature of the melting process. If foreign particles are embedded in the crystal lattice, it is fairly evident that their presence will result in the breaking up of the regular arrangement of the atoms at a lower temperature.

In an alloy composed of two metals the presence of small amounts of either metal lowers the melting point of the other, the drop in the melting point being proportional to the amount of the minor component, up to a certain limit. That combination of two metals which gives a minimum melting point is called a *eutectic*[1] *mixture*. Figure 18.3 is a plot of the melting point of solder, which is a lead-tin alloy, for various percentages by weight of the two components. The eutectic comes at 181°C for the composition tin 63 per cent, lead 37 per cent. If one started with a molten mixture of 20 per cent tin and 80 per cent lead at a temperature of 400°C and allowed it to cool, the process would proceed along the dotted curve of the diagram. Upon reaching the point A, lead would begin to freeze out. Latent heat of fusion would be released to retard the cooling. Finally, when the eutectic composition is reached, the residue would freeze, the temperature remaining constant at 181°C until solidification is complete. Starting with an 80 per cent tin, 20 per cent lead mixture, the tin would begin freezing out first at B, continued cooling leading again to the same eutectic.

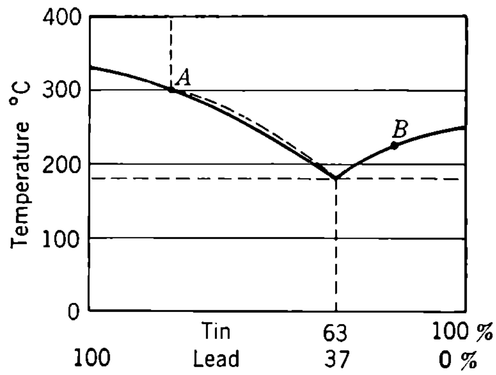


FIG. 18.3. Melting-point curve for lead-tin alloys.

By introducing a third metal, bismuth, into this mixture a eutectic alloy with a melting point of 96°C may be made. With more constituents an even greater lowering of the melting point may be effected. An example is Wood's metal (50% bismuth, 25% lead, 12.5% tin, and 12.5% cadmium), which melts at 65.5°C.

18.5. Vaporization. The molecules in a liquid are moving incessantly in a random manner with a rather wide distribution of velocities about a mean value that increases with the temperature. If a molecule at the surface of the liquid has a velocity with an upward component normal to the surface, it will shoot out of the liquid. Possibly the attractive force from the molecules in the surface will pull it back into the liquid; but if its velocity exceeds a certain critical value, it will escape into the region above the liquid and form part of the vapor. If the faster moving molecules are thus constantly leaving the liquid, the average kinetic energy of the remaining molecules in the liquid must be lowered. Now, as will be seen in Chap. 20, the temperature of the liquid is proportional to the average kinetic energy of its molecules, and hence the process of evaporation must *cool* the liquid. This cooling effect is a matter of common experience, illustrated, for example, in the cooling of the skin by the evaporation of perspiration.

[1] Greek *eu*, well + *tekein*, to melt.

Table 18.2. *Saturated Vapor Pressure of Water*

Temp	Pressure, cm Hg
0°C	0.458
20	1.751
40	5.513
60	14.92
80	35.51
100	76.0
140	271.0
180	751.4

If the space above the liquid surface is confined, some of the molecules of the vapor in their random motion strike back into the liquid. As the density of the vapor increases, the number of molecules returning to the liquid per second increases, until finally a state of equilibrium is reached in which the number returning to the liquid balances the number leaving. The vapor is then said to be *saturated*, and this maximum vapor pressure is called the *saturated vapor pressure*. Values of the saturated vapor pressure for water at several temperatures are given in Table 18.2 and are plotted in Fig. 18.4. The points on this curve represent the conditions under which liquid and vapor exist together in equilibrium. If at any of these points the pressure is increased without increasing the temperature, all the vapor will condense, while if the temperature is increased without increasing the pressure all the liquid will vaporize. The region to the left of the curve in this diagram represents the all-liquid condition, that to the right of the curve the all-vapor condition.

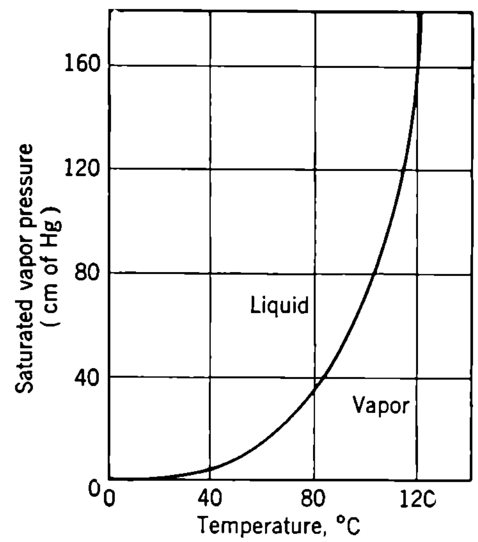


Fig. 18.4. Pressure of saturated water vapor.

The vapor in the confined space above the liquid will increase in density up to the saturated condition even though other gases may be occupying the space. Thus, if air at the normal 76 cm of mercury pressure is present and the temperature is 20°C in the enclosed space above a liquid water surface, water vapor contributing an additional partial pressure of 1.75 cm will form, so that the total pressure becomes 77.75 cm of mercury. The presence of the other gas retards the rate of evaporation, however, because of collisions between molecules of the gas and of the vapor. If the space above the liquid surface contains nothing but the vapor and the volume of this space is suddenly increased, as by raising the gastight piston in a cylinder containing nothing but the liquid and its vapor in equilibrium (Fig. 18.5), the increase in the evaporation of liquid to restore the vapor to its saturated pressure takes place practically instantaneously. If the piston is then suddenly lowered, the rate of return of the vapor molecules to the liquid immediately increases until the equilibrium saturated vapor density (and pressure) for the existing temperature is again established.

18.6. Boiling. When bubbles of saturated vapor form in the interior of a mass of liquid and have sufficient pressure to be stable, growing by accretion of more molecules as they rise to the surface, the liquid is said to be boiling. These vapor bubbles in the interior of the liquid will not be stable unless the pressure of their vapor is at least equal to the external pressure on them. Therefore, *the boiling point of a liquid is that temperature at which its vapor pressure equals the atmospheric pressure on the liquid.*

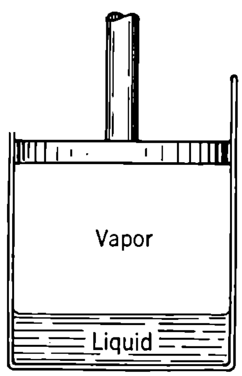


FIG. 18.5. Saturated vapor pressure is independent of volume above the liquid.

At normal barometric pressure of 76 cm of mercury, the boiling point of water is 100°C; in fact, this is arbitrarily taken as one of the fixed points of the thermometer scale. But it will boil at a lower temperature if the pressure above the liquid surface is reduced. At an altitude where the normal barometer reading is 70 cm of mercury, water boils at 97.7°C. From the data in Table 18.2 it is evident that, if the pressure above the water in a closed vessel is reduced to 35.51 cm of mercury, boiling will begin at 80°C. Thus a saturated vapor pressure curve such as Fig. 18.4 may be said to be a *boiling-point* curve. The essential equipment to obtain this in the laboratory is sketched in Fig. 18.6. The liquid is placed in the boiler B, which is connected through the condenser C to a manometer M and to a vacuum pump. The thermometer T for reading the boiling temperature at any pressure should have its bulb just above but not in the liquid, for the temperature of the liquid may be slightly higher owing to the presence of impurities.

Impurities in general raise the boiling point of a liquid, the elevation of the boiling temperature being, for small amounts of impurities, proportional to the amount present. In cooking, salt or sugar added to the water raises the boiling temperature somewhat. The rise in the boiling point is about 0.5°C for each mole of any nonvolatile substance dissolved in a liter of water.

18.7. Heat of Vaporization. Work Done in Expansion. The amount of heat required to change unit mass of a liquid to vapor without change of temperature is called the *heat of vaporization.* This quantity may be determined by one of the calorimetric methods outlined in the previous chapter. As seen in Sec. 17.6, most of the heat energy absorbed by the substance to vaporize it goes into increasing the potential energy of the molecules, *i.e.*, into work done against the intermolecular attractive forces in producing the large increase in the average distance between molecules. A sizable fraction of this heat energy, however, goes into external work done in the expansion of the substance against the confining pressure. When the vapor changes back to the liquid state, the heat of

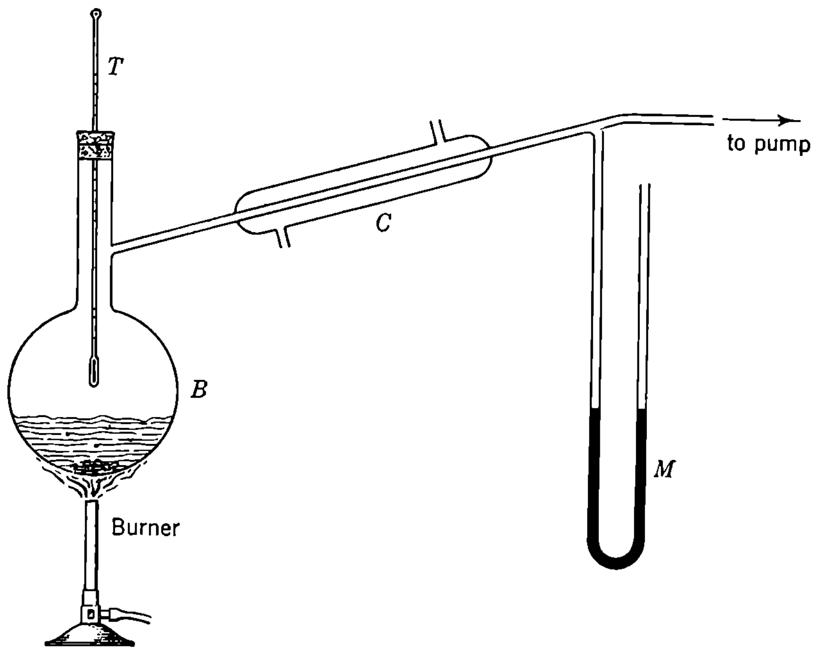


FIG. 18.6. Apparatus for determination of boiling-point curve.

vaporization is given up to the surroundings, with conversion of the molecular potential energy into kinetic energy. For example, in steam heating systems it is chiefly the heat of vaporization of the steam in condensing that warms the radiators.

The value of the heat of vaporization of any substance depends on the temperature at which the change from liquid to vapor occurs. For water it is 539 cal/gm at the normal boiling point of 100°C, but it is 596 cal/gm for water boiling under a pressure of 0.46 cm at 0°C, while at 180°C the heat of vaporization of water has fallen to 478 cal/gm. At the critical temperature (Chap. 19), at which the distinction between liquid and vapor vanishes and which occurs at 374°C for water, the heat of vaporization becomes zero. In Table 18.3 are given the normal boiling points and corresponding heats of vaporization for a number of common liquids.

Table 18.3. *Normal Boiling Points and Heats of Vaporization*

	Boiling pt		Heat of vaporization	
	°C	°F	cal/gm	Btu/lb
Alcohol, ethyl	78.3	109	204	368
Helium	−268.6	−451.5	6	10.8
Mercury	357	675	65	117
Nitrogen	−195.6	−320	47.6	85.5
Oxygen	−182.9	−297	50.9	91.6
Sulfur dioxide	138	280	50	90
Water	100	212	539	970

One gram of water at 100°C occupies a volume of approximately 1 cm³, whereas 1 gm of saturated steam at this temperature has a volume of 1,676 cm³. Before computing the exact amount of work done in this expansion, let us consider the general problem of the work done in an expansion against a constant pressure. Suppose that the substance fills the volume between the end of the cylinder and the gastight, frictionless piston in position a (Fig. 18.7). Atmospheric pressure p is exerted against the other side of the piston, which has an area A. A certain amount of heat is now supplied, enough just to vaporize all the liquid in this case, the resulting expansion pushing the piston through a distance x in the cylinder. Since the total force against the piston has been constant and of value pA throughout, the work that has been done is

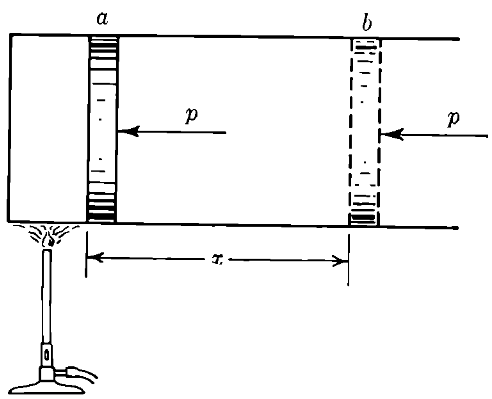


FIG. 18.7. Work is done by the expansion in the vaporization process.

$$W = Fx = pAx = p \times (\text{vol change}) \tag{18.1}$$

We may now compute the external work done when 1 gm of water is converted to steam at atmospheric pressure:

$$\begin{aligned} W &= (76 \times 13.6 \times 980)\text{dyne cm}^{-2}\,(1{,}676 - 1)\text{cm}^3 \\ &= 1.013 \times 10^6 \text{ dyne cm}^{-2}\ 1{,}675 \text{ cm}^3 = 169.7 \text{ joules} \end{aligned}$$

and since 4.186 joules = 1 cal,

$$W = \frac{169.7}{4.186} = 41 \text{ cal}$$

Therefore, of the 539 cal absorbed by the gram of water, 498 cal is consumed in producing the increase in molecular potential energy and 41 cal in doing external work.

Equation (18.1) holds when a substance, say a gas, expands against *constant* pressure. If the pressure varies during the expansion, the increment dW of work done for a small volume change dV is

$$dW = p\,dV \tag{18.2}$$

and hence the total work done for the finite expansion from the original volume V_0 to a volume V_1 is

$$W = \int_{V_0}^{V_1} p\,dV \tag{18.3}$$

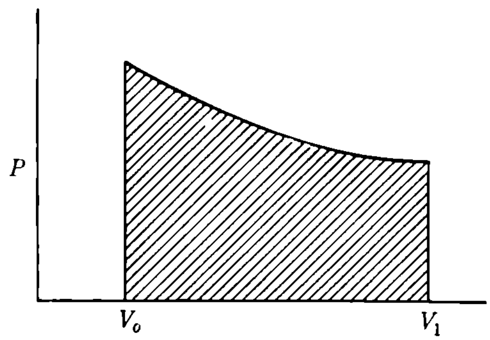


FIG. 18.8. The shaded area represents the work done by the substance in expanding against the pressure P.

To evaluate this integral the functional relationship between p and V must be known. Graphically, if the values of p are plotted against V as in Fig. 18.8, the area under the curve represents the value of the integral between the volume limits given. If the pressures are known from a gauge calibrated in lb/in^2 and the volumes are measured in cubic feet, the gauge readings must be converted to lb/ft^2 by multiplying by 144 in order to give W in foot pounds. Finally, considering this work done in expansion as positive, if a vapor condenses or a gas contracts in volume, then the external pressure does the work and we must regard W as a negative quantity.

18.8. Sublimation. The change directly from the solid to the vapor state can occur under the proper conditions of temperature and pressure. Common examples of this process, known as *sublimation*, are the vaporization of iodine crystals and of "dry ice." The latter *sublimes* at normal atmospheric pressure, for carbon dioxide cannot exist in the liquid state at a pressure below about 5 atm. If ice is heated when it is under a pressure less than 4.6 mm of mercury, it will pass directly to the vapor state. Just as for the solid-liquid and liquid-vapor transitions, a curve may be drawn connecting all the points of corresponding pressure and temperature at which the solid and vapor phases can exist together in equilibrium. Such a line in the *pt* plane is known as the "sublimation curve."

In Fig. 18.9 are plotted in one diagram the three equilibrium curves for water. Each of the three areas of this diagram represents all the conditions under which just the one state can exist. The curve PA represents all the pressure-temperature combinations for which liquid and gas may exist in equilibrium with each other. Similarly, the curve PB gives all the pt combinations for which solid and liquid may exist together in equilibrium, while PC is the plot of the solid-vapor equilibrium conditions. These three curves meet in a common point, where all three states—solid, liquid, and vapor—must exist in equilibrium with each other, and this is the sole pressure-temperature combination at which freezing and boiling can occur simultaneously. The point P is called the *triple point*, and for water it corresponds to a pressure of 4.6 mm of mercury and a

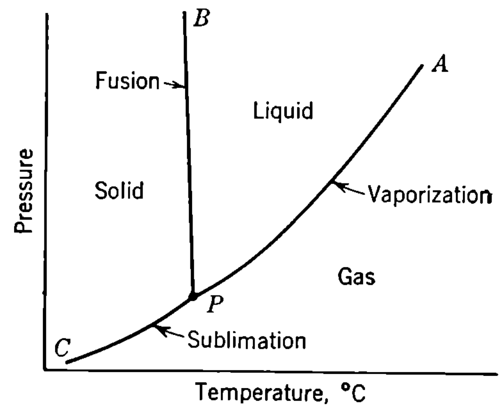


FIG. 18.9. The three equilibrium curves meet in a common point, the triple point.

temperature of +0.0072°C. The temperature for the triple point of any substance is very near its normal freezing point, since the change of freezing point with pressure is very slight.

PROBLEMS

1. How much heat is liberated when 10 ft³ of water freezes at 32°F?

2. A piece of copper weighing 250 gm is heated to 400°C and placed on a large block of ice. Assuming that all the heat given up by the copper goes into melting ice, how much ice is melted?

3. What must be the velocity of a lead bullet at 50°C so that when it strikes a target, if all the heat is developed within the bullet, it will just be melted?

4. How much steam at 140°C must be added to 1 kg of ice at −20°C to give nothing but water at 0°C?

5. A pressure cooker is filled with steam at a gauge pressure of 15 lb/in² (gauge pressure is the pressure in excess of atmospheric pressure). What is the approximate temperature within the cooker?

6. Heat is supplied to 500 gm of ice at −10°C, converting it finally to steam at 100°C. If the heat is supplied at a constant rate of 1,000 cal/min, how long will this take? Plot a graph with time as abscissa and temperature as ordinate.

7. How much heat will be liberated when 1 lb of steam at 212°F is changed to ice at 32°F?

8. One pound of water is boiled at 212°F and atmospheric pressure, becoming 26.8 ft³ of steam. Calculate the external work of the expansion, in foot pounds. What is the increase in internal energy, in Btu?

9. Take a mass of water substance across the triple-point diagram at some constant pressure above the triple point, explaining just what happens as the lines are crossed. Do this also for a constant temperature of $-1°C$.

10. What is the total force on the end of a boiler 2 ft in diameter if the temperature of the water inside is 140°C? (Neglect the weight of the water.)

11. The skating edge of one of a pair of ice skates has an area of 300 mm^2. The skater weighs 175 lb. Assuming that the drop in the melting point of ice varies linearly with pressure at the rate of 1°C drop for each 135 atm increase in p, calculate the lowering of the melting point under the runner of this skate. Do you believe that this is the main reason for the small amount of friction between the steel skate and the ice?

12. One thousand gallons of water is heated in a boiler from 70 to 302°F, at which temperature water boils under a pressure of 69.04 lb/in.^{2} If the heat of vaporization at 302°F is 907 Btu/lb and the average specific heat of water in this temperature range is 1.01 Btu lb^{-1} deg F^{-1}, how much heat is required to raise the temperature and evaporate this water? How many cubic feet of natural gas would produce this heat?

CHAPTER 19

PROPERTIES OF GASES AND VAPORS

19.1. Boyle's Law. A gas is a fluid that will expand so as to occupy completely any volume into which it is placed. For gases, as for liquids, the shear modulus of elasticity is zero, as described in Chap. 13. In the present chapter we shall discuss those properties of gases that are temperature dependent. Since all gases may be liquefied, there is no real difference between a gas and a vapor. Below its critical temperature (Sec. 19.9), however, a gaseous substance may properly be referred to as a vapor. But an unsaturated vapor does obey all the gas laws.

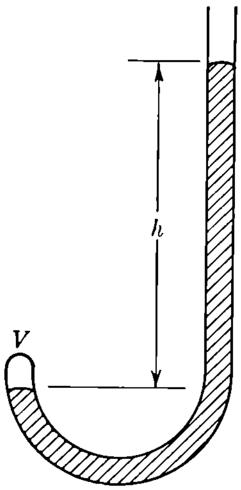


FIG. 19.1. Boyle's apparatus for pressures over atmospheric.

Robert Boyle in 1662 discovered a simple relationship between the pressure and the volume of a confined gas. *For a given mass of gas at constant temperature the product of the pressure and the volume of the gas is constant.* For pressures above atmospheric Boyle used a tube of the form sketched in Fig. 19.1. Air was trapped in the closed end of the tube at V by mercury, and the pressure on this air could be varied by regulating the height of the mercury in the open arm. The pressure for a difference in level h of the mercury in the two arms is $B + h$, where B is the atmospheric pressure in centimeters of mercury on the open end. If p is the pressure and V the volume of a mass of gas, then

$$pV = C \qquad \text{(for const temp)} \qquad (19.1)$$

where the value of the constant C is, of course, proportional to the mass of gas used. If p_1 and V_1 are the pressure and volume of the gas under one condition, while p_2 and V_2 are another pair of values for the same gas at the same temperature,

$$p_1V_1 = p_2V_2 \qquad \text{(temp const)} \qquad (19.2)$$

Careful measurements show that this law is not perfectly accurate, the inaccuracy becoming noticeable at high pressures (Sec. 19.8). For all gases at low pressures and at temperatures well above their boiling points, however, Boyle's law is accurate enough for all practical purposes. It is found convenient to speak of an *ideal gas*, an imaginary gas that obeys Boyle's law exactly for all pressures. This ideal gas has simple properties and is closely approximated by real gases at very low pressures and by helium and hydrogen even at ordinary pressures. From another point of view an ideal gas is one devoid of intermolecular forces. In Chap. 20 we shall show that such a gas satisfies Boyle's law.

19.2. Isothermal Expansion of an Ideal Gas. In Fig. 19.2 are plotted the pressure-volume relations of an ideal gas for two different temperatures. These so-called *isothermal* (Greek *isos*, equal; *thermos*, heat) curves, or pV curves for constant temperature, are equilateral hyperbolas, asymptotic to the p and V axes. The temperature t_2 is higher than t_1, for the constant C increases with temperature.

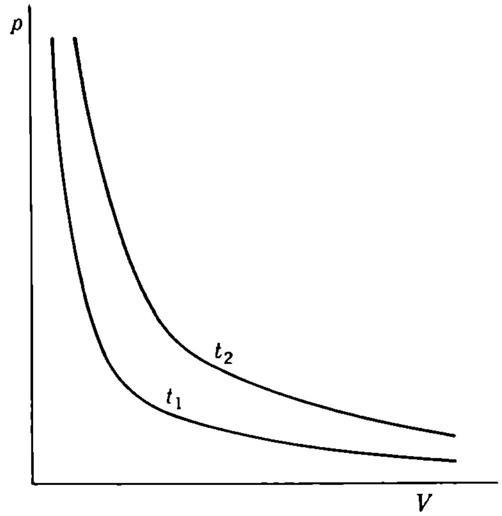


FIG. 19.2. Isothermal curves (pV = const) for an ideal gas.

When any gas is compressed rapidly by a pump, the gas is heated. A rapid expansion, conversely, cools the gas. Most pV changes are therefore not isothermal processes. To ensure the validity of isothermal relationships we shall have to assume, then, that the compressions and expansions are carried out so slowly that heat transfer to the surroundings can take place to maintain the temperature of the gas constant.

Suppose that a given mass of an ideal gas at a pressure p and volume V has its volume decreased by an amount dV by application of a slight additional pressure dp. Now the bulk modulus of elasticity (Chap. 4) is

$$M = \frac{dp}{-dV/V} = -V\frac{dp}{dV}$$

where the minus sign indicates that for increase in pressure the volume decreases. By Boyle's law, $p = C/V$, so that $dp/dV = -C/V^2 = -p/V$. Hence

$$M = p \tag{19.3}$$

That is, for small strains the bulk modulus of elasticity of an ideal gas is equal to the initial pressure, if the compression is carried out isothermally. For a sudden compression with a rise in temperature the stress-to-strain ratio is larger than this (Sec. 19.6).

If the gas expands isothermally from a volume V_1 to a volume V_2, the work done by the gas is

$$W = \int_{V_1}^{V_2} p\, dV \tag{19.4}$$

In view of Boyle's law, $p = C/V$, and therefore

$$W = \int_{V_1}^{V_2} C\,\frac{dV}{V} = C \ln \frac{V_2}{V_1} = p_1V_1 \ln \frac{V_2}{V_1} = p_2V_2 \ln \frac{V_2}{V_1} \tag{19.5}$$

And since $V_2/V_1 = p_1/p_2$, Eq. (19.5) may also be written

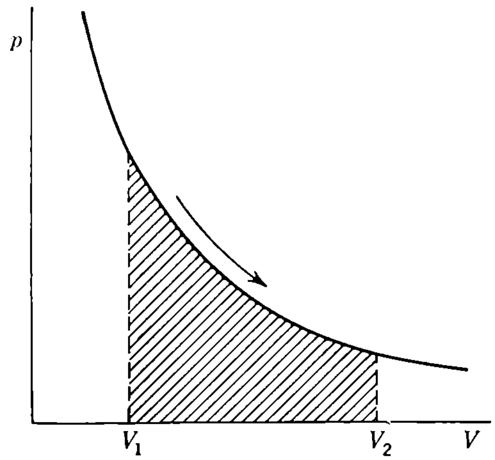


FIG. 19.3. Work done in isothermal expansion is given by shaded area.

$$W = p_1V_1 \ln \frac{p_1}{p_2} = p_2V_2 \ln \frac{p_1}{p_2} \tag{19.6}$$

On a pV diagram this work is represented by the area under the isothermal curve bounded by the V_1 and V_2 limits (Fig. 19.3). In an expansion, $V_2 > V_1$, and W is positive (work is done by the gas), while in a compression, $V_2 < V_1$, so that $\ln V_2/V_1$, and hence W, is negative. A negative value of W means that an external force does work *on* the gas.

Worked Example. Two cubic feet of air at 100 lb/in.² gauge pressure expands isothermally to atmospheric pressure (15 lb/in.²). Calculate the work done. By Eq. (19.6)

$$W = 115 \text{ lb/in.}^2 \times 144 \text{ in.}^2/\text{ft}^2 \times 2 \text{ ft}^3 \times \ln \frac{115 \text{ lb/in.}^2}{5 \text{lb/in.}^2}$$
$$= 3.31 \times 10^4 \times 2.0373 \text{ ft lb} = 6.73 \times 10^4 \text{ ft lb}$$

19.3. Thermal Expansion of Gases. Absolute Temperature Scale. There is a large variety of ways in which a gas may expand when heated, for both the pressure and volume may change simultaneously. The experimental results of greatest significance, however, are obtained by allowing the gas to expand at constant pressure. This may be accomplished by having the gas in a cylinder closed by a frictionless piston with the constant pressure of the atmosphere backing it (Fig. 19.4*a*). It is found that the increase in volume is closely proportional to the original volume and to the temperature increase. If the original temperature is 0°C, the volume being V_0, then the volume V_t at t°C is given by

$$V_t = V_0(1 + \beta t) \qquad \text{for } p = \text{const} \tag{19.7}$$

where the quantity β is known as the *volume coefficient of expansion* of the gas. This linear relation between volume and temperature, with pressure constant, is shown in Fig. 19.4*b*.

This relation is often referred to as Charles's law, in honor of Jacques Charles (1746–1823), who in 1787 discovered that all gases have the same expansion coefficients. Actually it was Gay-Lussac who in 1802 first published experimental results confirming this law. Later work showed that the law is only a close approximation, just as is Boyle's law. Still it should be noted that the volume coefficients of expansion of all gases are very nearly equal, whereas for liquids and solids this is not at all the case. The lower the pressure, the more nearly do all gases exhibit the

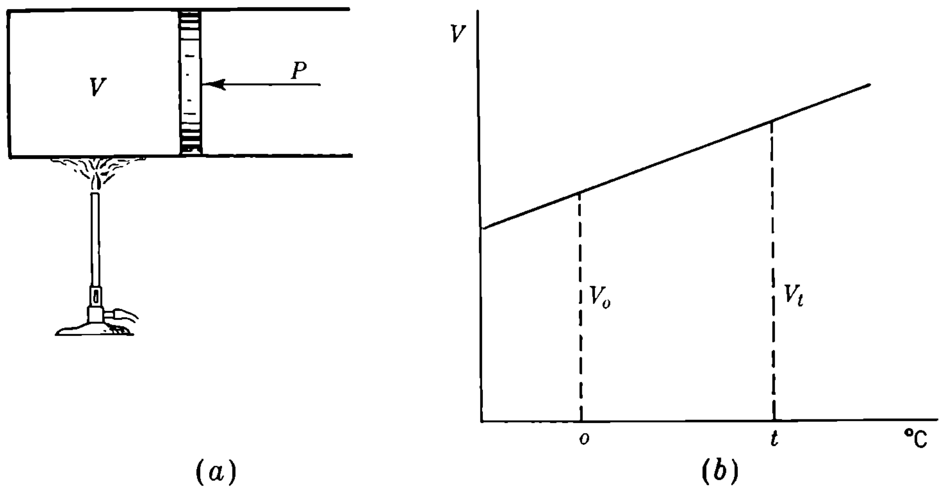


FIG. 19.4. Expansion of a gas at constant pressure.

same value of β as well as conform to Boyle's law. This common value of β is 0.003660 deg C^{-1} which may be taken as the volume coefficient for an ideal gas. Since β is expressed in reciprocal degrees, its numerical value depends on the size of the degree. The decimal 0.003660 is very nearly $1/273$, and so it may be said that the volume of a gas at constant pressure increases by $1/273$ of its value at 0°C for each centigrade degree rise in temperature.

When a mass of gas is kept at a constant volume as it is heated, its pressure also varies linearly with the temperature. If the original temperature is 0°C and the pressure is p_0, then the pressure p_t at t°C is given by

$$p_t = p_0(1 + \beta t) \qquad \text{for } V = \text{const} \qquad (19.8)$$

where we represent the *pressure coefficient* of the gas by the same symbol β used for the volume coefficient, since for our hypothetical ideal gas the two coefficients are identical, while for real gases the two are both very nearly $1/273$ deg C^{-1}. Equation (19.8) is often called Gay-Lussac's law.

Giving β its numerical value $1/273$ in Eq. (19.8) we may write

$$p_t = p_0\left(1 + \frac{t}{273}\right) = p_0\left(\frac{273 + t}{273}\right) \tag{19.9}$$

The quantity $(273 + t)$ is a temperature that we shall denote by T, referred to a zero point 273° below 0°C. This zero point is called *absolute zero*, and the temperatures calculated from this zero are called *absolute temperatures*. The pressure variation with temperature of a gas held at constant volume is then

$$p_t = p_0\left(\frac{T}{T_0}\right) \tag{19.10}$$

where T_0 denotes the ice point on the absolute scale. The temperature scale based on absolute zero is commonly known as the Kelvin scale, the designation of a temperature being T°K. An ideal gas when used in a constant-volume gas thermometer (Sec. 16.2) should give the Kelvin thermodynamic scale (Sec. 21.7) exactly. The precise value of absolute zero is −273.18°C, the lowest temperature that can possibly be attained (Fig. 19.5).

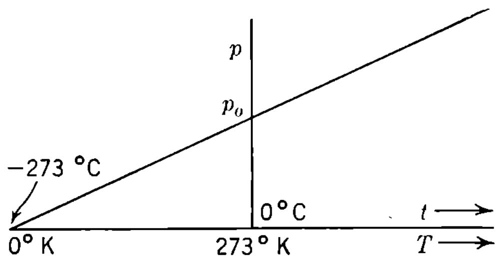


FIG. 19.5. At 0° on the absolute temperature scale an ideal gas would have zero pressure.

19.4. The Equation of State of an Ideal Gas. Equations (19.1), (19.7), and (19.8) give the relations between the pressure, volume, and temperature of an ideal gas, in each case with one of these quantities held constant. It is convenient to combine all three variables, p, V, and t, into one equation. Such a relation is called an *equation of state* of the substance. For an ideal gas we may derive such an equation by combining Boyle's law with Eq. (19.10) in the following manner: Let the gas be initially at the temperature T_0 (= 0°C), with pressure and volume p_0 and V_0, respectively. Then let the gas be heated at constant volume to the temperature T (Fig. 19.6). By Eq. (19.10) the pressure p_t is now $p_t = p_0(T/T_0)$. Next, holding the temperature constant at T°K, let the pressure-volume values change from p_t, V_0 to p, V. By Boyle's law, $p_tV_0 = pV$. By eliminating p_t between these two equations,

$$\frac{pV}{T} = \frac{p_0V_0}{T_0} = \text{const} \tag{19.11}$$

The value of this constant ratio of the product of p and V to the absolute temperature of the gas will depend on the kind of gas and on the quantity of it. Now according to a suggestion made in 1811 by Avogadro, *equal volumes of different gases at the same temperature and pressure contain equal numbers of molecules.* By this law a *mole* of any gas, which is the number of grams equal numerically to its molecular weight (O_2 = 32, N_2 = 28, He = 4, etc.), under standard conditions (p = 1 atm, t = 0°C), should occupy the same volume, 22,400 cm³. Hence, if the volume V contains exactly one mole of *any* gas, we may write

$$\frac{pV}{T} = R \tag{19.12}$$

where R is written for the combination p_0V_0/T_0 and is called the *gas constant per mole.* Its constancy can be seen from the fact that p_0 = 1

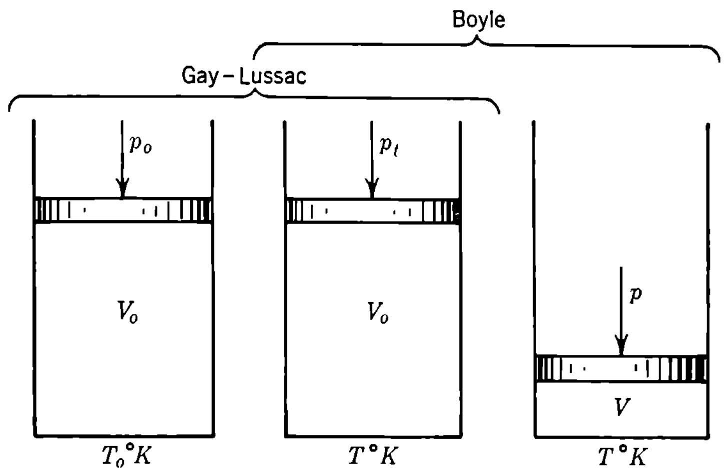


Fig. 19.6. Combination of Boyle's law and Gay-Lussac's law.

atm, T_0 = 273°K, and V_0 = 22,400 cm³. If V refers to the volume of n moles of *any* gas, then we may write as the final form of the *equation of state*

$$\boxed{pV = nRT} \tag{19.13}$$

Avogadro's law is strictly true only for an ideal gas, but it is a good approximation for real gases. To evaluate R in cgs units, take n = 1 mole and assume standard conditions,

$$R = \frac{1.013 \times 10^6 \text{ dyne cm}^{-2} \times 22{,}400 \text{ cm}^3}{273.18 \text{ deg}} = 8.314 \times 10^7 \text{ ergs deg}^{-1} \text{ mole}^{-1}$$

Worked Example. A tank of oxygen gas has a volume of 3 ft³ and shows a gauge pressure of 150 lb/in.² at a temperature of 70°F. How many pounds of oxygen does the tank contain? On taking 15 lb/in.² = 1 atm the gas pressure in the tank is 11 atm. One cubic foot is 28.3 liters, and 70°F is $273 + (70 - 32) \times 5/9 = 294.1°K$. Substituting in Eq. (19.13),

$$n = \frac{11 \times 1.013 \times 10^6 \times 3 \times 28.3 \times 10^3}{8.314 \times 10^7 \times 294.1} \text{ or } 38.8 \text{ moles}$$

Since one mole of oxygen weighs 32 gm, the contents of this tank weigh $(38.8 \times 32)/454$, or 2.73 lb.

19.5. Specific Heats of a Gas. When any body is heated, part of the heat may be used in doing external work ΔW in its volume expansion against any external forces acting on the body. But a greater portion of the heat energy Q absorbed by the body always goes into increasing its *internal energy* U. The internal energy is composed of (1) the total kinetic energy of all the molecules, to which the temperature of the body is proportional; (2) the energy of the atoms within the molecules; and (3) the potential energy of the molecules with respect to neighboring molecules to which they are bound by (attractive) forces. Because of the conservation of energy the heat-energy input should equal the sum of the increments in these several forms of internal energy, ΔU, plus the external work done if the body expands. Stating this mathematically,

$$\Delta Q = \Delta U + \Delta W \tag{19.14}$$

where the three terms must, of course, all be in the same energy units. As will be seen later, this equation is the *first law of thermodynamics*. Any of the quantities ΔQ, ΔU, or ΔW may, of course, be negative.

In Chap. 17 we defined the specific heat of a substance as the quantity of heat absorbed or given out when unit mass of the substance changes its temperature by one degree. Similarly the *molar specific heat of a gas*, C, is defined as the quantity of heat required to raise one mole of the gas 1°C. Since specific heats vary with the temperature, if ΔQ represents the quantity of heat passing into or out of a mass of n moles while it changes in temperature by Δt, the molar specific heat at temperature t in the interval Δt is

$$C = \frac{\Delta Q}{n\,\Delta t}$$

or, in differential notation,

$$C = \frac{1}{n}\frac{dQ}{dt} \tag{19.15}$$

The quantity of heat involved in raising the temperature of a gas one degree depends upon whether the volume or the pressure is held constant. If the volume is constant, the external work $\Delta W = 0$. We therefore distinguish between the molar *specific heat at constant volume*, C_v, and the molar *specific heat at constant pressure*, C_p. (Actually one can define any number of specific heats for a gas if *both* p and V are allowed to vary in the heating process, but these two are of greatest interest. For a solid $C_p = C_v$, of course, because of the negligible expansion with temperature.) Suppose that we have n moles of a gas in a volume V at a pressure p and an absolute temperature T. Writing Eq. (19.14) in differential notation[1] we have for C_v, since $dW = 0$,

$$C_v = \frac{1}{n}\left(\frac{dQ}{dT}\right) = \frac{1}{n}\frac{dU}{dT} \tag{19.16}$$

But since, for the constant-pressure case, $dW \neq 0$, we have, for C_p,

$$C_p = \frac{1}{n}\left(\frac{dU}{dT} + \frac{dW}{dT}\right) = C_v + \frac{1}{n}\frac{dW}{dT} \tag{19.17}$$

Now $dW = p\,dV$, and hence, from the equation of state for an ideal gas Eq. (19.13), we have

$$\frac{1}{n}\frac{dW}{dT} = \frac{1}{n}p\frac{dV}{dT} = \frac{1}{n}nR = R$$

Therefore,

$$\boxed{C_p - C_v = R} \tag{19.18}$$

where the universal gas constant R must be expressed in the same units used for the molar specific heats. The difference between the two molar specific heats should, then, be the same for all gases if the specifications for an ideal gas were actually met.

If C_p and C_v are determined experimentally for any gas, Eq. (19.18) forms a means for evaluating the mechanical equivalent J. This is essentially what J. R. Mayer (1814–1878) did when he made the first calculation of J in 1842. For example, hydrogen at 15°C has $C_p = 6.783$ and $C_v = 4.812$, both in cal mole^{-1} deg C^{-1}. From the equation of state, $R = 8.314 \times 10^7$ ergs deg C^{-1} mole^{-1} of H_2. Hence

$$J = \frac{8.314 \times 10^7 \text{ ergs mole}^{-1} \text{ deg C}^{-1}}{1.971 \text{ cal mole}^{-1} \text{ deg C}^{-1}} = 4.2 \times 10^7 \text{ ergs/cal}$$

The measurement of C_p can always be made by a continuous-flow method (Sec. 17.10) without much difficulty. The direct determination of C_v for a gas is rather difficult, however, because of the fact that the gas has such a small mass compared with that of its container. Fortunately the ratio $\gamma = C_p/C_v$ is easily obtained from a measurement of the speed of

[1] In order to avoid confusing the student we purposely do not use partial derivatives.

sound in the gas (Sec. 37.7), and hence C_v may be calculated if C_p is known. Knowing C_p, Eq. (19.18) may also be used to calculate C_v. The specific heats of a few common gases are listed in Table 19.1.

Further discussion of the specific heats of gases is found in our treatment of the kinetic theory in the next chapter. We shall there consider the theoretical values of the specific heats and the partition of the internal energy of polyatomic molecules.

19.6. Adiabatic Processes. Any thermodynamic change during which no heat is allowed to enter or escape from the substance is called an *adiabatic process* (Greek *a*, not; *dia*, through; *batic*, marching). This can be realized in practice either by having the process take place in an enclosure with such well-insulated walls that heat is not transmitted through them or by having the whole process occur in such a short interval of time that no appreciable amount of heat can be conducted through the walls. We require the relation between the pressure and volume of an ideal gas in an adiabatic change, a relation that is the counterpart of Boyle's law, pV = const, for an isothermal change.

Table 19.1. *Specific Heats of Gases at Atmospheric Pressure*

Gas	Temp, °C	c_p, cal gm^{-1} deg C^{-1}	C_p, cal mole^{-1} deg C^{-1}	C_p/C_v
Alcohol, ethyl	90	0.406	18.704	1.13
Air	100	0.2404	6.924	1.401
Ammonia	15	0.5232	8.910	1.310
Argon	15	0.1253	5.005	1.668
Carbon dioxide	15	0.1989	8.752	1.304
Hydrogen	15	3.389	6.783	1.410
Nitrogen	15	0.2477	6.941	1.404
Oxygen	15	0.2178	6.970	1.401
Water	100	0.4820	8.686	1.324

In an adiabatic process $\Delta Q = 0$; hence we may write Eq. (19.14) in the form

$$dU = -p\,dV \tag{19.19}$$

That is, the change in the internal energy of the gas equals the external work done. From Eq. (19.16), $dU = C_v n\,dT$, and thus

$$C_v n\,dT = -p\,dV \tag{19.20}$$

From Eq. (19.13), we have

$$p\,dV + V\,dp = nR\,dT \tag{19.21}$$

Upon eliminating dT between these two equations there results

$$\frac{C_v(p\,dV + V\,dp)}{R} = -p\,dV \tag{19.22}$$

Now, $R = C_p - C_v$, and $C_p/C_v = \gamma$. Making these substitutions in Eq. (19.22) and rearranging terms we obtain

$$\frac{dp}{p} + \gamma\frac{dV}{V} = 0 \tag{19.23}$$

Integration of this equation gives

$$\ln p + \gamma \ln V = \text{const} \tag{19.24}$$

or

$$pV^\gamma = \text{const} \tag{19.25}$$

This is the equation of an adiabatic curve. Plotted in a pV diagram as in Fig. 19.7 it gives a steeper curve at any point such as A than the isothermal that passes through the same point.

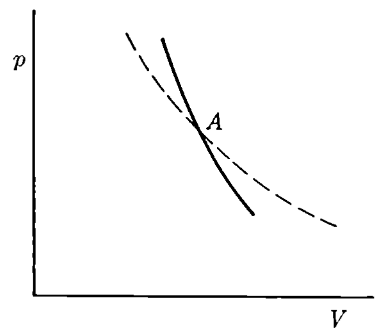


FIG. 19.7. An adiabatic curve is steeper than an isothermal curve (dashed line) at the same point.

A sudden compression of a gas is usually an adiabatic process. If the strain is small, the adiabatic bulk modulus of elasticity $M = -V\,dp/dV$ is obtained at once from Eq. (19.23). The result is

$$M = \gamma p \quad \text{(adiabatic change)} \tag{19.26}$$

where p is the initial pressure of the gas.

***19.7. Expansion of a Gas.** We have already noted that, if there are forces between the molecules of a gas, the change in the average distance between molecules accompanying an alteration in the volume of the gas involves some change in the internal energy. (It is evident that there are strong forces binding the molecules together in the solid and liquid states, but it is not at all obvious that for a gas, with the much larger distances between molecules, these intermolecular forces are appreciable.) In the famous Joule-Thomson porous-plug experiment direct evidence was found for the existence of intermolecular forces in a gas. The gas at constant high pressure was allowed to expand through a plug of tightly packed cotton (a small-bore needle valve could have been used) into a low-pressure region (Fig. 19.8). A slight cooling was found for all gases except hydrogen and helium. This cooling is evidence that there is indeed a small attractive force between the gas molecules, the gain in potential energy in the expansion causing a corresponding reduction in the kinetic energy of the molecules. The cooling is approximately proportional to the pressure difference on the two sides of the plug and is greater the lower the initial temperature of the gas. Even hydrogen and helium are cooled by expansion if their temperature is first lowered below the so-called "temperature of inversion," −80.5°C for hydrogen and about −238°C for helium.

In an ideal gas whose molecules are not supposed to exhibit any forces on each other, this temperature drop in a free expansion should not occur. The cooling effect in a "Joule-Thomson expansion" found in real gases is employed in all methods of liquefying air and the other gases that at one time were thought to be permanent gases. If the expansion is so arranged that a real gas, instead of expanding freely, pushes a piston against a back pressure, this energy, too, must be supplied by the gas at the expense of the kinetic energy of the molecules and a further cooling effect results.

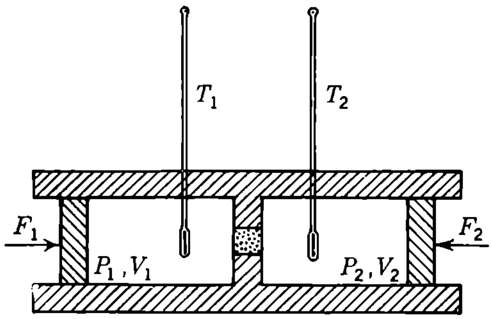


FIG. 19.8. $T_1 \neq T_2$ in this porous-plug experiment.

***19.8. Deviations from Boyle's Law.** If Boyle's law were an accurate relation, holding for all gases no matter how high the pressure, the plot of pV against p would be a horizontal straight line parallel to the p axis. The first careful investigation of the deviations from Boyle's law was carried out by Regnault in 1847, and his measurements were later extended to high pressures by Amagat. The variation of the product of p and V for a given mass of several gases is sketched in Fig. 19.9. For all gases except hydrogen and helium the curves first drop to a minimum value that depends both on the kind of gas and on its temperature, and then for still higher pressures the pV products show steady increase with p. The curves for hydrogen and helium have an upward slope at all pressures, as do those of other gases at high temperatures.

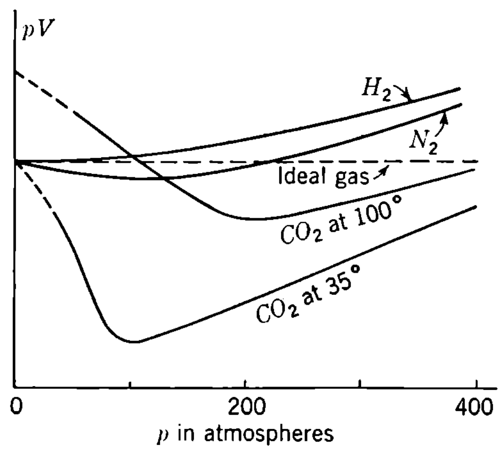


FIG. 19.9. Deviations of a given mass of several gases from Boyle's law at high pressures.

Over the range of pressures and temperatures plotted in Fig. 19.9 liquefaction of these gases does not take place. The investigation of the pressure-volume relations of a gas at and near the liquefying point is a matter of very considerable importance. A careful study of the liquefaction of carbon dioxide was carried out by Thomas Andrews (1813–1885) in 1863. His results are typical of the behavior of gases near liquefaction, for similar pVT relations, but with different numerical values, hold for any gas as it is being liquefied. The next section describes Andrews's work.

19.9. The Critical Point. In Andrews' experiment, carbon dioxide was contained in a glass tube. To measure the pressure required to compress the gas to various measured volumes, mercury was forced into the tube. The tube could be kept at a constant temperature during these operations. Isothermal curves obtained by Andrews in the temperature range from 13 to 48.1°C are plotted in Fig. 19.10. The curve for 48.1°C resembles the typical hyperbola for a gas obeying Boyle's law. As the temperature

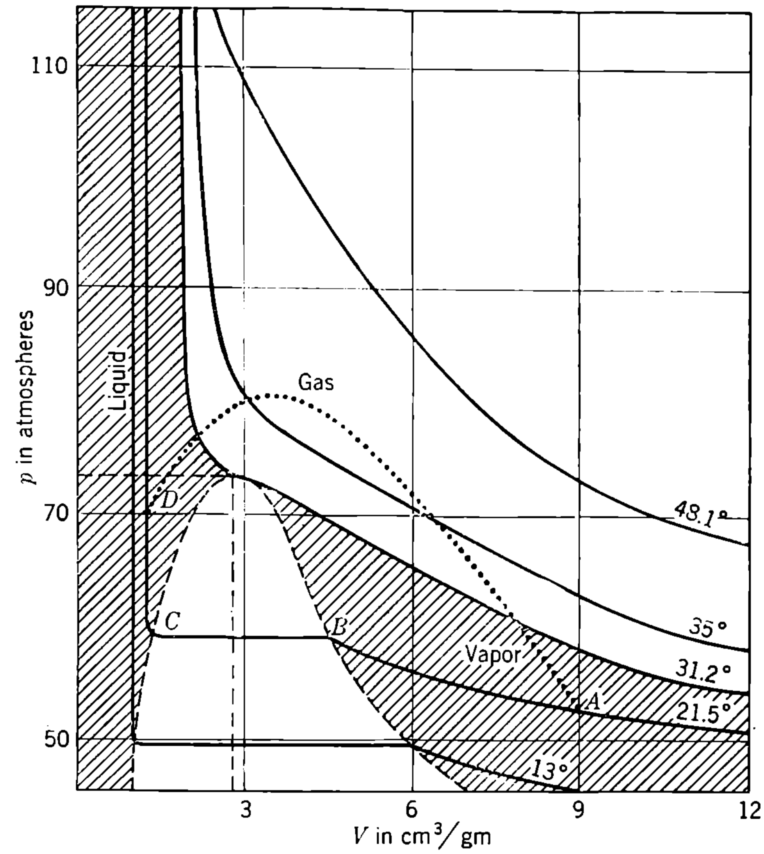


FIG. 19.10. Isothermals for carbon dioxide.

is lowered a few degrees below this, however, the isothermals in this pV plane become distorted. The curve for 31.2°C is known as the *critical isothermal*. No part of it is horizontal, but there is a marked point of inflection at $p = 73$ atm, and for still higher pressures the curve rises almost vertically, indicating that the substance is incompressible like a liquid.

Proceeding along the 21.5°C isothermal from a point such as A, we see that the gaseous CO_2 is at first rather easily compressed into a small volume by increase in the pressure. At B the gas begins to liquefy, and one observes a definite meniscus separating liquid and vapor in the tube. There is then a considerable decrease in volume but no increase in pressure until the point C is reached. Here the substance is all in the liquid state, and further increase in pressure produces but a minute decrease in the volume. The isothermal for 13°C has a similar but longer horizontal portion, indicating that there is a greater difference in density between the liquid and its vapor at the lower temperature. In fact, as the temperature is raised to the critical value 31.2°C, the horizontal portions of the isothermals get shorter, disappearing altogether at this temperature. Above 31.2°C no liquefaction of the CO_2 gas is possible, however much the pressure is raised.

This temperature above which the gas cannot be liquefied by application of pressure alone is called the *critical temperature.* The pressure that is just sufficient to liquefy the gas at this temperature is called the *critical pressure.* The dashed line through the ends of the horizontal portions of the isothermals below the critical isothermal surrounds an area representing all the physical conditions in which the liquid and its vapor can be in equilibrium with each other. The highest point of this dashed curve is called the *critical point,* and the corresponding volume is the *critical volume.* The part of the dashed curve to the right of the critical point is called the *saturation curve.* It represents all pVt conditions for saturated vapor, and its abscissas are the specific volumes (V per gram or the reciprocal of the density) of the saturated vapor. The portion of the dotted curve to the left of the critical point is the *liquid curve* whose abscissas are the specific volumes of the liquid. At the critical point the liquid and the saturated vapor have the same density.

It is possible to pass from a point such as A, where the substance is a gas, to a point D, representing a liquid state, by passing along the dotted curve without crossing the dashed curve, *i.e.*, without having the liquid distinct from the vapor at any time. There is *continuity of state* above the critical temperature; there is, in fact, never any distinction between the

Table 19.2. *Critical Constants for Gases*

Gas	Crit temp, °C	Crit pressure, atm	Density, gm/cm³
Air	−140.8	37.2	0.35
Alcohol, ethyl	243.1	63.1	0.276
Ammonia	132.4	111.5	0.235
Carbon dioxide	31.2	73.0	0.460
Helium	−267.9	2.26	0.069
Hydrogen	−239.9	12.8	0.031
Nitrogen	−147.1	33.5	0.430
Oxygen	−118.8	49.7	0.430
Water	374.0	217.7	0.4

liquid and vapor states. All gases exhibit properties like this, but there is a wide variation in critical constants. Critical data for a few common gases are assembled in Table 19.2.

★19.10. Liquefaction of Gases. In order to liquefy a gas it must be cooled below its critical temperature, and then by application of sufficient pressure it may be liquefied. As Table 19.2 indicates, air and some other gases have very low critical temperatures. The commercially successful method of liquefying such gases was developed by Linde in 1895, using the regenerative expansion method (Fig. 19 11). A compressor raises the pressure of the air to about 4,000

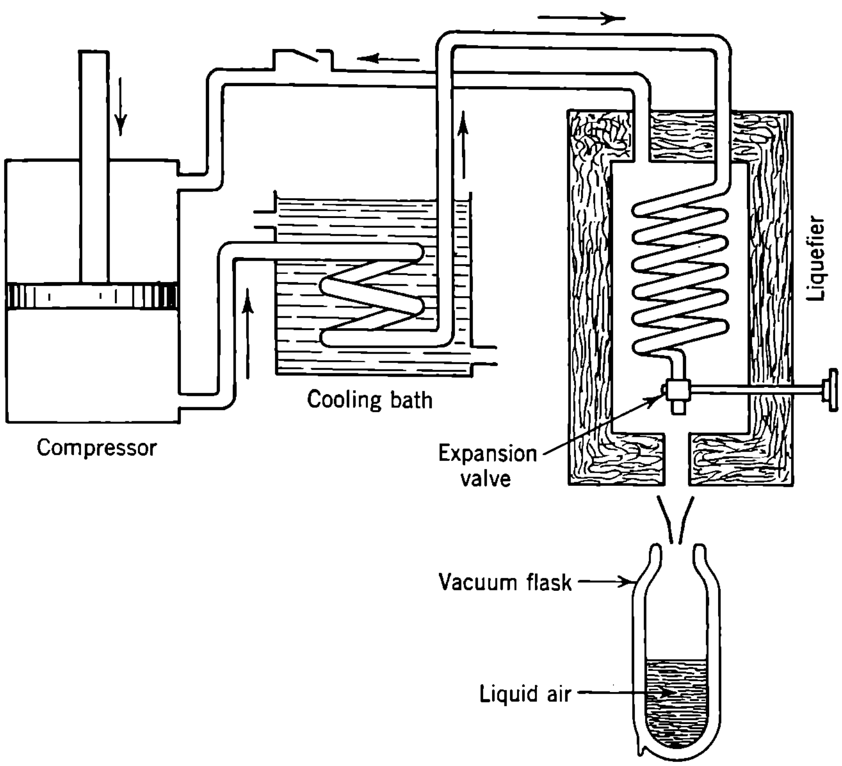


FIG. 19.11. Schematic arrangement of apparatus for liquefying air.

lb/in.^{2}, the heat generated in the gas by the compression being removed in water-cooling tanks. The high-pressure air then passes through the many turns of tubing in the liquefier, issuing at the bottom through an adjustable needle valve into a space in which p is atmospheric. This expansion produces a considerable cooling (0.25°C per atmosphere decrease in pressure, and this cooling increases with decreasing temperature), and the air thus cooled flows up around the coil in the liquefier, thus cooling the oncoming compressed air. This regenerative process goes on, gradually lowering the temperature at which expansion takes place, until finally some of the air liquefies on expansion and flows into the flask below.

The yield of liquefied gas may be increased considerably by precooling the compressed gas in some manner before it enters the liquefier. For example, if the coil in the precooler is surrounded by a mixture of dry ice and some solvent such as carbon tetrachloride, the high-pressure air is cooled to about $-80°C$. This principle is used in liquefying hydrogen, the precooling being accomplished with liquid nitrogen. Further cooling needed to reduce the hydrogen gas below its critical temperature is produced by the expansion. Liquid hydrogen may in turn be used to precool helium, which is the most difficult of all gases to liquefy. P. Kapitza has used a regenerative balanced-expansion method for liquefying helium without the aid of liquid hydrogen. In Kapitza's apparatus the compressed gas is made to do external work, which cools it more. This procedure in addition to the precooling with liquid nitrogen and cooling from the Joule-Thomson effect results in yields of as much as several liters of liquid helium per hour. This method of producing liquid helium is enjoying a great vogue at the present time, for it makes possible many researches on the behavior of matter at temperatures near absolute zero.

***19.11. Equations of State. van der Waals' Equation.** It is obvious from our discussion of the isothermals of a real gas that the equation of state of an ideal gas [Eq. (19.13)] is increasingly inaccurate the closer the approach to the liquid state. Two simplifying assumptions were made in the derivation of the ideal-gas equation, and these must be improved upon if we wish to obtain a general equation that connects the three variables p, V, and T and that will hold even for a condensed state of a real gas. One of these improvements is to recognize the existence of forces of attraction between the molecules; the other is to allow for the fact that the volume available for the gas to expand into must be reduced by an amount of the order of the actual volume occupied by the molecules of the gas. Of the several improved equations of state which have been produced, that suggested by J. D. van der Waals (1837–1923) is the most generally used. The van der Waals equation, applicable both to the gaseous and to the liquid state, is

$$\left(p + \frac{a}{V^2}\right)(V - b) = nRT \tag{19.27}$$

where a and b are constants characteristic of each gas.

To explain simply the form of the van der Waals correction to the pressure, it is to be noted that the forces of attraction between the molecules constitute an unbalanced force directed into the gas and hence form a slight addition to the pressure p. Empirically it was found that this correction has the form a/V^2. A detailed treatment, beyond the scope of this text, shows that the b constant is a volume about three times that which would be occupied by the gas molecules packed most closely together as in the solid state. A plot of an isothermal as given by Eq. (19.27) and corresponding to, say, the 21.5°C isothermal of CO_2 in Fig. 19.10 is given in Fig. 19.12. Between the points B and C the curve is continuous, cutting the experimentally observed horizontal portion BC at three points. The reason for this is that the equation is cubic in V, with all three roots real in this region. For the pressure p_c and the temperature T_c at the critical point, the three roots of the equation are equal. Above the critical point the equation has only one real root, and for large values of V and T it correctly represents a curve that is nearly hyperbolic.

It is, of course, impossible to reproduce experimentally the portion of the curve in Fig. 19.12 lying between maximum and minimum, for this represents an increase in volume with an increase in pressure. It is, however, possible to obtain experimental points on the left side of the curve below C and points on the right

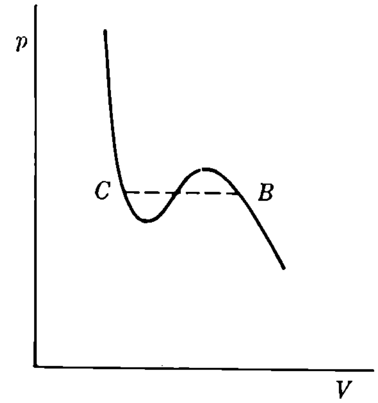


FIG. 19.12. Isothermal according to van der Waals' equation.

side above B. The points below C are obtained by *superheating* a liquid, while those above B represent an *undercooled* vapor.

PROBLEMS

1. An air bubble at the bottom of a lake 10 m deep has a volume of 1 cm³. Calculate the volume of the bubble just as it reaches the surface of the water, if the barometer stands at 750 mm.

2. Because of air in the space above the mercury in a barometer tube that is 80 cm long the mercury stands at 70 cm when the correct barometer reading is 76 cm. Where will the mercury stand when the correct barometer reads 74 cm?

3. A cylinder of oxygen gas has a volume of 3 ft³, and its pressure gauge reads 2,000 lb/in.² What volume would this gas occupy at atmospheric pressure if its temperature did not change?

4. A tire is inflated to a gauge pressure of 35 lb/in.² when the temperature is 20°C. While the car later is traveling at high speed in the hot sun, the temperature of the tire becomes 50°C. Assuming no change in the volume of the tire and no leakage, what is then the pressure in the tire?

5. If a gas has a volume of 10 ft³ at a gauge pressure of 20 lb/in.² and a temperature of 60°F, what volume will it occupy if its gauge pressure is increased to 30 lb/in.² and the temperature is raised to 80°F?

6. Some dry air occupying a volume of 140 cm³ at 17°C is heated at constant pressure to 50°C. What is the new volume?

7. An oxygen cylinder has a volume of 2 ft³. What will be the gauge pressure in this cylinder when 5 lb of oxygen has been forced into it at a temperature of 17°C?

8. What volume will be occupied by 10 gm of methane gas (molecular weight 16) at a pressure of 2 atm and at 25°C?

9. How much external work is done when 10 gm of air (molecular weight 28.8) is heated from 15 to 65°C at a constant pressure of 76 cm of mercury?

10. What pressure is exerted by 2 gm of nitrogen gas contained in a 3-liter flask at 20°C?

11. Calculate the work done in compressing 2 ft^3 of gas isothermally into 1 ft^3, the initial pressure being 75 cm of mercury.

12. An ideal gas undergoes an adiabatic expansion from condition p_1, V_1 to condition p_2, V_2. Show that the external work done is $[1/(\gamma - 1)]\ (p_1V_1 - p_2V_2)$.

13. Some air at 27°C and 76 cm of mercury pressure is suddenly compressed to half its volume. If the compression is adiabatic, what will be the new temperature and pressure? [Combine Eqs. (19.25) and (19.13) to eliminate either p or V, as the case may be.]

***14.** The constants in the van der Waals equation for methane are $a = 0.0045$ and $b = 0.00191$ for p in atmospheres and V in ft^3. If 1 ft^3 of methane at 0°C and $p = 1$ atm is compressed to 0.05 ft^3 isothermally, what is the new pressure and by how much does it differ from that calculated with Boyle's law?

15. Find the density of water vapor at 25°C and a pressure of 20 mm of mercury.

16. How much heavier is a volume of 1 liter when filled with carbon dioxide at 100°C and $p = 76$ cm of mercury than when filled with hydrogen at 0°C and the same pressure?

17. Because of air in the space above the mercury surface in a barometer tube the mercury column is but 40 cm long and the space above the mercury is 40 cm long when a good barometer reads 75 cm. Some liquid ether is then introduced into this space. To what level will the mercury column fall? Take the saturated ether vapor pressure as 40 cm and that of mercury as 1×10^{-4} cm of mercury, respectively.

CHAPTER 20
KINETIC THEORY

20.1. The Kinetic Theory of Gases. We have made numerous references to the molecular structure of matter and to the random thermal motion of the molecules of a gas. The existence of molecules as the smallest entities characteristic of a given chemical compound was firmly established experimentally during the first half of the nineteenth century by Dalton, Gay-Lussac, Avogadro, and many others. What we now know as Avogadro's law, *viz.*, that equal volumes of different gases at the same temperature and pressure contain equal numbers of molecules, was actually at the time (1811) only a hypothesis made by Avogadro. A proof of this law came later when it was shown that the ratios of the molecular weights of various gases are the same as the ratios of their densities, which can be true only if the number of molecules per unit volume is the same for all gases. There are departures from exact equality of these ratios for all gases, however, the deviations being the greatest for those gases which fail to obey Boyle's law. Avogadro's law is also, then, only a good approximation.

Just after the middle of the last century the mechanical nature of heat was firmly established, principally by the many accurate experiments of Joule. It then became evident that the remarkably simple laws governing the thermal behavior of ideal gases could be explained by a theory based upon a simple model. The *kinetic theory of gases* was developed during the second half of the century by J. C. Maxwell (1831–1879), L. Boltzmann (1844–1906), and others into one of the important branches of theoretical physics. The theory in its simplest form is based upon the following assumptions regarding the gas molecules and their behavior:

1. In an enclosed gas at a fixed temperature the molecules are moving in a purely random manner, so that on the average the number moving in any direction is the same as that in any other direction.
2. The volume occupied by the molecules themselves is negligible in comparison with the total volume of the enclosure.
3. The molecules exert no forces on each other except at the instant of collision.
4. The impacts of the molecules with each other and with the walls are perfectly elastic. If this were not the case, energy would be lost by the gas because of collisions of the molecules with the walls, with resulting drop in temperature.

We have already seen that the second and third of these assumptions

It is necessary to assume Eq. (20.5) in order to reconcile the kinetic-theory result with the empirically determined equation of state. In other words, Eq. (20.5) constitutes the kinetic-theory definition of temperature.

If $\nu = N/V$ is the number of molecules per unit volume, then Eq. (20.3) may be written

$$p = \frac{1}{3}\frac{Nm}{V}\overline{v^2} = \frac{1}{3}\nu m\overline{v^2} = \frac{2}{3}\nu E_1 \tag{20.7}$$

where $E_1 = E/N$ represents the average kinetic energy per single molecule. Placing this value of p in Eq. (20.6), we have

$$\tfrac{2}{3}\nu V E_1 = nRT \tag{20.8}$$

Now $\nu V/n$ equals the number of molecules N_0 in a mole of any gas, or *Avogadro's number*. The value of N_0 is 6.02×10^{23} mole^{-1}. In terms of N_0, Eq. (20.8) may be written

$$\tfrac{2}{3}N_0 E_1 = RT \tag{20.9}$$

or

$$E_1 = \frac{3}{2}\frac{R}{N_0}T = \frac{3}{2}kT \tag{20.10}$$

where $k = R/N_0$ is an important quantity called the *gas constant per molecule*, or the *Boltzmann constant*. We see that the average kinetic energy of a molecule of *any* gas is the same for a given temperature. The value of k is 8.314×10^7 ergs deg^{-1} mole$^{-1}/6.02 \times 10^{23}$ mole^{-1} $= 1.38 \times 10^{-16}$ ergs deg C^{-1}. Since this is the same for all molecules, it may be called one of the *universal constants* of the physical world.

Summary of Symbols Used in This Section:

N = total number of molecules present in gas
n = number of moles of gas
ν = number of molecules per cm^3
N_0 = number of molecules per mole (Avogadro's number)
m = mass of a single molecule
$M = Nm$ = total mass of gas
$\rho = M/V$ = density of gas
E = total energy of gas
$E_1 = E/N$ = energy per molecule

20.3. Molecular Velocities. The mean-square velocity of the molecules of any gas may be computed from Eq. (20.3). By taking the square root of this mean-square velocity we obtain what is known as the *root-*

mean-square (rms) *velocity*. Thus,

$$\sqrt{\overline{v^2}} = \sqrt{\frac{3p}{\rho}} \tag{20.11}$$

Because of the relatively small ρ the molecules of hydrogen gas should have the largest rms velocity. For hydrogen at $p = 1$ atm and 0°C, $\rho = 9 \times 10^{-5}$ gm/cm³, and hence

$$\sqrt{\overline{v^2}} = \sqrt{\frac{3 \times 76 \text{ cm} \times 980 \text{ cm/sec}^2 \times 13.6 \text{ gm/cm}^3}{9 \times 10^{-5} \text{ gm/cm}^3}} = 1.84 \times 10^5 \text{ cm/sec}$$

The most convenient method of computing the rms velocity of the molecules of any gas at any temperature is to employ Eq. (20.9). We leave it as an exercise for the student to show that from this equation

$$\sqrt{\overline{v^2}} = \sqrt{\frac{3RT}{\mu}} \tag{20.12}$$

where μ is the molecular weight of the gas. It is to be noted that, according to this kinetic-theory relation, the molecular velocity would be zero at the absolute zero of temperature.

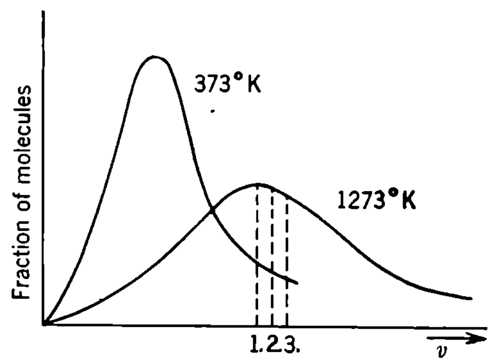


FIG. 20.2. Maxwellian distribution of velocities for molecules of a gas. 1, 2, and 3 are the most probable, the arithmetic mean, and the root-mean-square velocities respectively.

So far we have been concerned with the *mean* velocity of the molecules. There is a large variation in the actual velocities, and the velocity of a given molecule changes continually as it collides with other molecules. The instantaneous value of the velocity might be nearly zero, or it might be much larger than the rms value. Since the number of molecules is huge and since there is complete randomness about their motions, the well-known methods of statistics are applicable for deriving the manner in which the molecules are distributed over the whole range of velocities. It turns out that the form of the distribution is similar to that for the variation of the number of bullet holes with distance from the bulls-eye of a target, provided that a large number of shots are fired by persons of equal skill. The distribution for the molecular velocities was first calculated by Maxwell and by Boltzmann. In Fig. 20.2 we plot such a Maxwellian distribution for a given number of molecules for two different temperatures.

Since the two curves in this figure each represent the same total number of molecules, the areas under the two curves must be equal. Each curve starts from zero, passes through a maximum, and then comes asymptotically down to the v axis. As the temperature increases, the maximum, which represents the most probable velocity, moves toward larger velocities. It is to be noted that there are many molecules at 373°K with velocities greater even than the most probable velocity at 1273°K. The *arithmetic mean* of all the velocities $(1/N \sum_{i=1}^{N} v_i)$ comes at a value somewhat larger than the most probable value, and the rms velocity $\sqrt{\sum_{i=1}^{N} v_i^2/N}$ is still larger. About 0.6 of the molecules have velocities greater than the most probable velocity. The fact that a small fraction of the molecules may have quite large velocities even for moderate temperatures means that hydrogen molecules high in the earth's atmosphere may frequently have velocities greater than the velocity of escape (Chap. 11).

Experimental evidence for the existence of the random motion of the molecules in a gas or liquid is supplied by a discovery made by Robert Brown, a botanist, in 1827. He noticed that, with a good microscope, colloidal particles suspended in water or smoke particles in a gas are observed to dance about in a rapid, irregular manner. This *Brownian movement* is caused by the unequal bombardment of the particles on different sides by the very much smaller molecules of the liquid or gas. Careful study of the motions of these colloidal particles reveals that they have a mean kinetic energy equal to that of a gas at the existing temperature.

Further evidence for molecular motion comes from measurements of the rate of *diffusion* of the molecules of one fluid through those of another fluid. The diffusion rate depends among other factors on the relative thermal velocities of the two types of molecule.

***20.4. Specific Heats of Ideal Gases. Equipartition of Energy.** The molecules of a monatomic gas may be considered as little spherical particles. Now a particle constrained to move along a line, either straight or curved, is said to have one *degree of freedom.* If the particle is restricted to move only on a surface, it is said to have two degrees of freedom. If, however, the particle can move freely in three-dimensional space, it has three degrees of freedom. In general, a molecule has as many degrees of freedom as coordinates necessary to describe its position. As compared with a monatomic molecule a diatomic or polyatomic molecule has additional possibilities for its motion, hence more degrees of freedom than the three associated with the translation of its center of mass in space.

The diatomic molecule may be thought of as a dumbbell-shaped body with an elastic connection between the two mass particles (Fig. 20.3). The rotation of this molecule about its center of mass C may be described in terms of just two coordinates, since there would be a negligible moment of inertia and hence negligible energy associated with rotation about the line connecting the two atoms. Such a molecule has then two degrees of freedom of rotation, which, together with the three degrees of freedom for the translation of the center of mass in space, makes five degrees of freedom in all. In a polyatomic, nonlinear molecule there is an additional rotational degree of freedom (three coordinates used in describing the rotation), or a total of six. Also, in both diatomic and polyatomic molecules there is the possibility that the atoms or groups of atoms may vibrate with respect to each other. These vibrations should be more prominent the higher the temperature. They form another means of storing energy and constitute other degrees of freedom.

The total internal energy of a monatomic gas is just the kinetic energy of translation of its molecules. For this three-dimensional motion there are three degrees of freedom; and since the molecular motion is perfectly random, exactly one-third of the total energy is said to be associated with each degree of freedom. This is an example of the *principle of equipartition of energy*, which says that *each* degree of freedom has associated with it an energy equal to $\frac{1}{2}kT$ per molecule.

Let us see how well these ideas agree with the observed values of the specific heats of various gases. From Eq. (20.9) we may write for the total kinetic energy of the molecules in a mole of an ideal gas

$$\boxed{N_0E_1 = \tfrac{3}{2}RT} \qquad (20.13)$$

For a monatomic gas this is the total internal energy per mole, and by the equipartition principle we have for the energy per degree of freedom

$$\tfrac{1}{3}N_0E_1 = \tfrac{1}{2}RT \qquad (20.14)$$

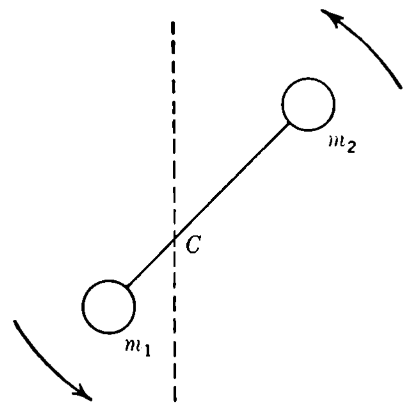


FIG. 20.3. A diatomic molecule.

From Eq. (19.20) the molar specific heat of a gas at constant volume is

$$C_v = \frac{1}{n}\left(\frac{dQ}{dT}\right) = \frac{1}{n}\frac{dU}{dT} \qquad (20.15)$$

Therefore for a mole of a monatomic gas

$$C_v = \frac{d(\frac{3}{2}RT)}{dT} = \frac{3}{2}R = 2.98 \text{ cal mole}^{-1} \text{ deg}^{-1} \qquad (20.16)$$

or close to 1 cal mole^{-1} deg^{-1} for each degree of freedom. If one assumes equipartition, the existence of five degrees of freedom for a diatomic molecule would then indicate that a diatomic gas should have a molar specific heat of 5 cal mole^{-1} deg^{-1}. For a polyatomic gas with six degrees of freedom the theoretical value of C_v is 6 cal mole^{-1} deg^{-1}.

In practice the specific heat at constant pressure C_p is the quantity usually measured, using the continuous-flow method, and then C_v is calculated either from the relation $C_p - C_v = R$ or from the ratio $\gamma = C_p/C_v$ as determined from measurement of the speed of sound in the gas.

For a monatomic gas such as argon, helium, or mercury vapor, γ should be $\frac{5}{2}R/\frac{3}{2}R = 1.67$. A diatomic gas like hydrogen, oxygen, nitrogen, or air should have γ equal to 1.4, while γ for a triatomic gas such as carbon dioxide or water should be 1.33. Comparison of these theoretical predictions for C_v and for γ with the experimental values *at room temperature* is made in Table 20.1. Inspection of this table shows that for the simpler molecules the agreement between the theoretical and experimental values is excellent. For triatomic or more complex molecules some energy may go into the vibrations of the atoms, with the result that both C_p and C_v may be expected to be larger and their ratio smaller than the theoretical values based upon the six degrees of freedom.

Table 20.1. *Theoretical and Experimental Values of* C_v *and* $\gamma = C_p/C_v$

Gas	Atoms per molecule	Degrees of freedom	C_v, cal mole^{-1} deg C^{-1}		γ	
			Theory	Experiment	Theory	Experiment
Argon	1	3	2.98	2.98	1.67	1.67
Helium				2.98		1.67
Mercury				—		1.67
Air	2	5	4.97	4.95	1.40	1.401
Hydrogen				4.80		1.410
Oxygen				5.04		1.401
Nitrogen				4.93		1.404
Carbon dioxide	3	6 or more	>5.96	6.75	<1.33	1.304
Water				6.67		1.324
Ether (35°C)	15	>6	>5.96	30.9	<1.33	1.08

***20.5. Variation of Specific Heats with Temperature. Quantum Theory.** The good agreement between the theoretical and experimental values of C_v at ordinary temperatures for the diatomic molecules listed in Table 20.1 disappears at high temperatures. For example, at 2000°C, hydrogen has a C_v equal to 5.8 cal mole^{-1} deg C^{-1}. This increase in the amount of energy needed to raise the temperature of the gas by one degree must be absorbed into the vibrational degree of freedom, but the gradual increase of the specific heat with temperature indicates that only a fraction of the molecules receive enough energy in their collisions to set their atoms into vibration. Also, the values of C_v for the diatomic molecules in Table 20.1 indicate that near room temperature these molecules have no vibrational degrees of freedom. The specific heat of hydrogen has been investigated down to 20°K. At low temperatures the number of degrees of freedom seems to decrease, so that below 60°K C_v for hydrogen has the value 3 cal mole^{-1} deg C^{-1} found for monatomic gases. At this low temperature the hydrogen molecules do not even have their rotational degrees of freedom activated.

The fact that a molecule must receive a certain definite amount of energy in order to have one of its rotational or vibrational degrees of freedom activated is contrary to the classical laws of mechanics. For, classically, a body capable of vibrating would be set vibrating with small amplitude no matter how small the blow. Many types of experiment have now convinced physicists that, if a molecule has a natural fundamental frequency of vibration f, it will vibrate only when its energy is hf or some multiple of this amount. The h is a *universal constant of action* discovered by Max Planck in 1900. The quantity hf is called a *quantum of energy* and the theory the *quantum theory*. There is now much evidence that all the vibrational, rotational, and electron motions in molecules are *quantized; i.e.*, they can be excited only when the energy received by an impact exceeds a critical amount characteristic of the particular atoms involved. For a light diatomic molecule like H_2, hf is greater than $E_1 = \frac{3}{2}kT$ for $T = 300°K$, and therefore this vibration is not produced by impacts with other H_2 molecules having only this amount of energy. For a heavy diatomic molecule like iodine, on the other hand, the natural vibration frequency f is so much smaller that, for 300°K, E_1 is larger than hf and nearly all the iodine molecules will have at least one quantum of vibrational energy.

***20.6. The Mean Free Path of Gas Molecules.** In its random thermal motion a molecule in a gas at atmospheric pressure makes many collisions with the other gas molecules. For this reason the diffusion of one gas through another is a very slow process if mass motion of the gas through convection currents or turbulence is not present. If a strong odor released in one corner of the room is soon detected in the far corner, the molecules of the odorous compound must have been carried this distance by convection. Upon collision any molecule is just as likely as not to be set into motion in the reverse direction. The average distance l that a molecule travels between collisions with other molecule, called the *mean free path*, is an important quantity in the kinetic theory. An expression for l in terms of the diameters of the molecules and their number ν per cubic centimeter can be derived as follows:

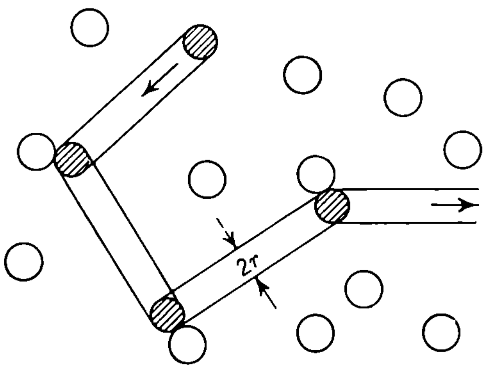


FIG. 20.4. Illustrating the broken cylindrical volume swept out by one of the gas molecules in its random motion.

Let us assume that the molecules are spheres of effective radius r. One of these molecules, traveling with an arithmetical mean velocity $\bar{v}$, sweeps out in 1 sec a cylindrical volume $\pi r^2 \bar{v}$ cm³, but of course this volume has a kink or bend in it for each collision (Fig. 20.4). The average volume occupied per molecule is $1/\nu$ cm³. Now any molecules having their centers within a radial distance $2r$ from the axis of the cylinder project into it and will be struck by this molecule. Per second, $4\pi r^2 \bar{v}\nu$ of these collisions will occur. The mean free path l is then the distance the molecule travels in a second, $\bar{v}$ cm, divided by the number of collisions in that same time; thus

$$l = \frac{\bar{v}}{4\pi r^2 \bar{v}\nu} = \frac{1}{4\pi r^2 \nu} \text{ cm} \qquad (20.17)$$

Since ν varies directly with the pressure, the mean free path is inversely proportional to the pressure of the gas.

This calculation is oversimplified, for we have assumed that all the gas molecules except the one are at rest. Maxwell showed that by taking into account the fact that all the molecules are in motion with various speeds the mean free path should be given by

$$l = \frac{1}{4\pi r^2 \nu \sqrt{2}} \text{ cm} \tag{20.18}$$

It is easy to see qualitatively that, since all the molecules are in motion, the chance of collisions will thereby be increased and thus the value of l as given by Eq. (20.17) should indeed be somewhat lowered.

If any two of the three quantities l, r, and ν are known, Eq. (20.18) may be used to solve for the third. The number ν of molecules per cubic centimeter for an ideal gas under standard conditions is Avogadro's number, 6.02×10^{23} mole^{-1}, divided by the volume of a mole, 2.241×10^4 cm^3, as determined from the ideal-gas equation. Therefore

$$\nu = \frac{6.02 \times 10^{23}}{2.241 \times 10^4} = 2.69 \times 10^{19} \text{ cm}^{-3}$$

This is often called the Loschmidt number. The mean free path may be measured directly, using the modern method of molecular beams, and may be calculated indirectly from measurements of the viscosity, diffusion, or heat conduction of the gas. For the ordinary gases under standard conditions l has values between 1 and 2×10^{-5} cm, and the fairly good agreement between the values for a given gas as calculated by the different methods gives one confidence in the general correctness of the kinetic theory.

The value of r may be determined from viscosity measurements (Sec. 15.7), X-ray diffraction patterns, and molecular spectra and from the constant b of van der Waals' equation. All these methods yield values of r that are fairly consistent with each other, with r lying in the range 1 to 3×10^{-8} cm for the ordinary gases. In this connection it should be mentioned that, since the molecules are not spheres actually, it is difficult to say exactly what their diameters are. The outer portions of all atoms and molecules consist of a rather diffuse distribution of negative electrons, and so the studies of the several phenomena mentioned may be expected to yield slightly different values for the size of any molecule. It is really very satisfying that the information from the data of a number of rather unrelated types of experiment (plus their theoretical interpretation) about these molecules, which are so small that they cannot be seen and measured directly, is all quite consistent.

For hydrogen at normal atmospheric pressure and 0°C the following approximate values are found; $l = 1.42 \times 10^{-5}$ cm, $2r = 2.4 \times 10^{-8}$ cm, and $\bar{v} = 169{,}500$ cm/sec.

20.7. The Mass of an Atom. The outstanding property that all atoms possess is that of mass. Mass is, of course, an important parameter in the study of the motion of atoms by themselves, as well as in the study of the motion of collections of atoms, *i.e.*, of solid bodies and fluids. Since Avogadro's number is so large, we know that ordinary quantities of materials contain very large numbers of atoms. From Avogadro's number the mass of an atom can at once be found. If A is the atomic weight of a substance, then the mass m of one atom is

$$m = \frac{A}{N_0} \tag{20.19}$$

Thus for hydrogen, the lightest atom, $A = 1.008$ gm and $m = 1.6 \times 10^{-24}$ gm. The hydrogen molecule, which contains two atoms, has twice this mass. Similarly, for iron, $A = 55.85$ gm, and $m = 9.16 \times 10^{-23}$ gm; for one of the heaviest atoms, uranium, $A = 238.07$ gm, and $m = 3.93 \times 10^{-22}$ gm. The range of masses is just the range of atomic weights and varies by a factor of more than 200 from the lightest to the heaviest atom.

***20.8. The Size and Shape of Atoms.** By a procedure exactly analogous to that of the preceding section the size of an atom can be determined. The volume of one mole of atoms is A/ρ, where ρ is the density of the substance. The average volume V occupied by one atom is therefore

$$V = \frac{A}{N_0\rho} \tag{20.20}$$

or, from Eq. (20.19),

$$V = \frac{m}{\rho}$$

The average volume occupied by one atom is, of course, a volume much larger than the atomic volume itself unless, in the substance under consideration, the atoms are in contact with one another. Thus the average volume that an atom or molecule occupies in an ideal gas is, since one mole occupies 22.4 liters at 0°C,

$$V = \frac{2.24 \times 10^4}{6.02 \times 10^{23}} = 3.7 \times 10^{-20} \text{ cm}^3$$

If the temperature is lowered, the density of the gas increases and V decreases, until the gas becomes a liquid. The atoms still have some heat motion, but a further reduction in temperature causes the substance to solidify. In a solid, V is still smaller, and the atoms may be said to be truly in contact with one another. A reduction in temperature produces little change in volume, and the compressibility of a solid is extremely high. To find the volume of a hydrogen atom we should use the density of solid hydrogen, therefore, in Eq. (20.20). This density has been determined to be $\rho = 0.081$ gm/cm^3, and hence, for hydrogen, $V = 1.008/6.0 \times 10^{23} \times 0.081 = 2.1 \times 10^{-25}$ cm^3. For iron, $\rho = 7.8$ gm/cm^3, $A = 55.8$ gm, and $V = 1.18 \times 10^{-23}$ cm^3; and the corresponding quantities for lead are $\rho = 11.4$ gm/cm^3, $A = 207.2$ gm, and $V = 3.4 \times 10^{-23}$ cm^3. We shall examine in more detail later (Fig. 49.1) how V varies from the lightest to the heaviest atoms.

To find the linear dimensions of an atom from the volume, it is necessary to know the shape of the atom as well. If the shape of a hydrogen atom is assumed to be spherical, then the volume just calculated corresponds to that of a sphere of diameter 3.4×10^{-8} cm.

It is interesting to compare this diameter with the diameter of the orbit of the circulating electron in a hydrogen atom. In Sec. 11.9 the orbital diameter was found to be 1.05×10^{-8} cm and hence smaller than the distance between atoms, as might be expected.

Some clue to the shape of the atoms of an element might be obtained from the form and symmetry of crystals of the element. Iron forms crystals with cubic symmetry, and the arrangement of the atoms within the crystal is indicated in Fig. 20.5. The structure is called a body-centered cubic structure. Sulfur, on the other hand, forms crystals in long needles. We cannot conclude from this, however, that iron atoms have a cubic shape, whereas sulfur atoms are needlelike, since, if iron and sulfur are combined to form iron sulfide crystals, a new and complicated structure results. Some elements, moreover, form two kinds of stable crystal. Carbon, for example, exists either as a diamond or as graphite. Thus it is difficult to find evidence for a definite shape that is characteristic of an atom of an element.

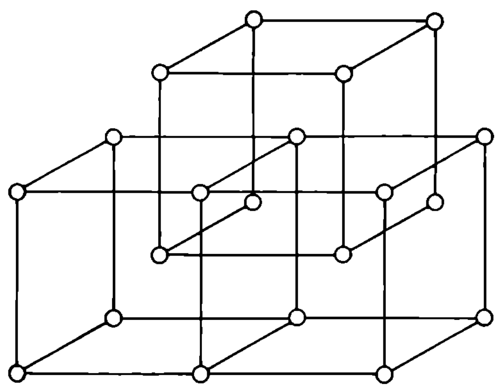


FIG. 20.5. The body-centered cubic structure of iron and many other metals.

Some additional evidence on the shape and size of atoms can be obtained by a detailed study of the collisions in a monatomic gas. The value of the radius of an atom of a gas obtained from the coefficient of viscosity agrees, although not exactly, with that from the simpler method indicated above. It would be expected, however, that the shape and size of an atom in a solid body where it is close to many other atoms might be different from those of an atom in a gas where the atoms spend most of the time far away from their neighbors. The wave-mechanical theory of atomic structure tells us that atoms do not have sharp boundaries, and consequently the concept of shape loses much of its significance.

***20.9. The Production of a High Vacuum.** Most of the discoveries of the past 50 years in atomic physics have been made with high-vacuum apparatus, and today the manufacture of many articles such as incandescent lamps, radio tubes, neon signs, X-ray tubes, vacuum bottles, etc., requires the use of high-vacuum techniques on a large scale. During the evacuation process all glass parts are usually heated to drive off occluded vapors and gases, and metal fittings within the vacuum may be outgassed by high-frequency induction heating. In radio tubes the last traces of gas are removed by a chemical "getter," which is usually some alkali or alkaline-earth metal distilled into the tube to combine with residual oxygen, nitrogen, or water vapor.

The modern *rotary pump* usually consists of an eccentric rotor C revolving in a cylindrical casing (Fig. 20.6). Once each revolution the gap space between rotor and cylinder wall is connected at A with the vessel being evacuated and fills with air, which is then compressed and forced out through a valve in the outlet tube B. A plunger D is pressed against C by a spring and prevents air from leaking back from B to A. The pump works in oil, which acts as a vacuum seal. Such pumps can reduce the gas pressure to about 10^{-3} mm of mercury. They are used as *fore pumps* to produce the preliminary vacuum for mercury- or oil-diffusion pumps, with which a vacuum of 10^{-7} mm of mercury may be obtained.

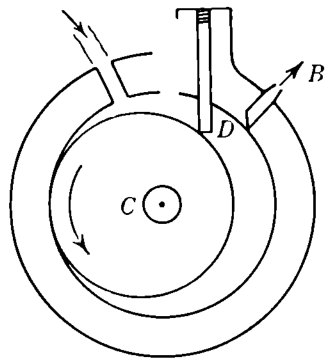


FIG. 20.6. A rotary vacuum pump.

A single-stage mercury-diffusion pump is sketched in Fig. 20.7. Mercury is heated in the boiler B, producing vapor, which passes up the tube, issues at the nozzle O, is condensed by a water jacket, and falls back into the boiler to be vaporized again. At O the fast-moving mercury atoms collide with air molecules that have diffused in from A, which connects with the apparatus being exhausted. The air molecules are thus given a downward velocity component and are removed through the outlet C by the fore pump. Often these pumps have two or three nozzles or stages in series for more rapid pumping and a better

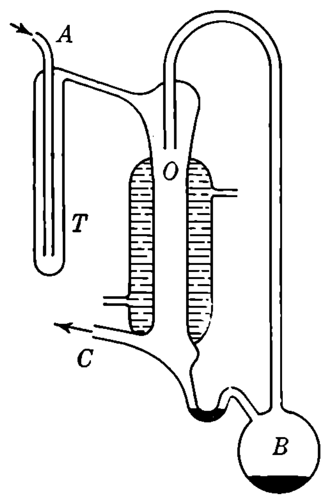


FIG. 20.7. Single-stage mercury diffusion pump.

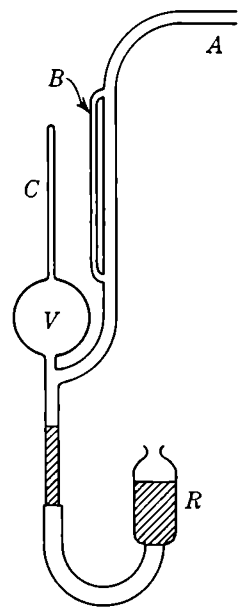


FIG. 20.8. McLeod gauge for measuring low gas pressures.

ultimate vacuum. A trap T cooled by immersion in liquid air or in a dry-ice and organic solvent mixture must be used to freeze out mercury vapor that diffuses back toward the evacuated apparatus. Recently fractionating diffusion pumps operating with a low-vapor-pressure oil or a liquid such as n-butyl phthalate have come into use because for many purposes no liquid-air trap is needed to prevent these vapors from diffusing into the evacuated apparatus.

To measure these low pressures it is customary to use a *McLeod gauge*, a *Pirani gauge*, or an *ionization gauge*. A simple form of McLeod gauge is shown in Fig. 20.8. The gauge is connected at A to the apparatus being evacuated. When the mercury in the lower tube is raised, it traps gas at the low pressure p into the accurately known volume V from the apparatus. Further raising of the mercury level compresses this gas into a known volume of the capillary C. The pressure of this compressed gas is given by the difference in level of the mercury in C and that in the parallel capillary tube B, which has the same bore as that of the C tube to minimize error due to capillarity. The original pressure of the gas trapped in V may then be computed with the aid of Boyle's law.

One weakness of the McLeod gauge is that it does not indicate correctly the pressure of condensable vapors. For fairly low but not the lowest pressures the Pirani gauge is satisfactory. It consists merely of a wire carrying an electric current in the low-pressure region. Since the heat conducted away from the wire is proportional to the pressure of the gas, one need only measure the electric resistance of the wire, for this resistance in turn varies linearly with the temperature of the wire. For measurement of pressures lower than 10^{-4} mm of mercury the *ionization gauge* is now commonly used. This is actually just a triode thermionic vacuum tube, which functions very well as a low-pressure gauge since the ionization current in the residual gas for fixed electron current is proportional to the pressure.

PROBLEMS

1. If 10^{22} molecules per second of hydrogen moving with a speed of 10^5 cm/sec strike 2 cm^2 of wall at an angle of 45°, how much pressure do they exert on the wall? The mass of the H_2 molecule is 3.32×10^{-24} gm.

2. Calculate the rms velocity at $T = 300°K$ of (a) helium atoms and (b) uranium hexafluoride molecules (mol wt = 352).

3. What is the total average kinetic energy of the molecules of 1 gm of oxygen gas at 300°K?

4. How much slower does a molecule move at 4°K than at room temperature (20°C)?

5. One gram of hydrogen and 2 gm of helium are introduced together into a completely evacuated bulb of radius 20 cm. Find the total pressure exerted by this gas mixture on the walls of the bulb if the temperature is 20°C. (From Dalton's law of partial pressures, total pressure equals sum of partial pressures.)

6. The best vacuum yet attained in experimental apparatus is 10^{-9} mm of mercury. How many molecules of residual gas still remain in 1 cm^3 at 0°C in this so-called vacuum?

★7. Compare the rms velocity of hydrogen molecules at liquid-air temperature (−190°C) and at the surface of the sun (6000°C).

8. If of 100 molecules 10 have a speed of 1 cm/sec, 10 have a speed of 2 cm/sec, etc., with the last 10 having a speed of 10 cm/sec, calculate (*a*) their rms velocity and (*b*) their arithmetic-mean velocity. Note that (*a*) is larger than (*b*).

★9. Heat is supplied to a quantity of ether vapor ($\gamma = 1.08$). What fraction of this heat is used in increasing the translational energy of the ether molecules?

★10. The diameter of the argon atom (argon is monatomic) is 2.2×10^{-8} cm. Calculate the mean free path of argon atoms under standard conditions.

11. At what centigrade temperature is the rms velocity of a molecule one-half its value at 20°C?

★12. Generalizing Eq. (20.16) to the case of a molecule with f degrees of freedom, show that then $\gamma = 1 + 2/f$. Calculate from the experimental value of γ for ether (Table 20.1) the number of degrees of freedom for this molecule.

CHAPTER 21
THE LAWS OF THERMODYNAMICS

21.1. The First Law of Thermodynamics. Thermodynamics is the science that describes the changes occurring in bodies as they are heated, or do external work, or change their state. Any of these effects may take place singly or in combination. Bodies of particular interest in this respect are gases, such as air; vapors, such as steam; or mixtures, such as gasoline vapor and air.

The subject is dominated by two laws, the first of which has already been stated in Sec. 19.5. Suppose that a body is heated, receiving ΔQ units of heat. During this process it does ΔW units of work. Writing ΔU for the increase in *internal energy* of the body, we may express the first law of thermodynamics in the form

$$\Delta U = \Delta Q - \Delta W \tag{21.1}$$

The quantity ΔQ in this equation, as everywhere else in this chapter, is assumed to be expressed in the same energy units as ΔU and ΔW, for instance, ergs (calories times J). In words, the first law says that the increase in internal energy equals the heat received by the substance, less the work done by the substance. The internal energy is not a directly measurable quantity, for it includes the kinetic energy of the molecules, the energies of vibration and rotation of the constituents of the molecules, and the potential energy due to the intermolecular forces. In the form of Eq. (21.1) the law does no more than assert the validity of the principle of conservation of energy. Its usefulness arises from the fact that a thermodynamic system, when taken through a transformation that finally restores the original state, incurs no change in internal energy ($\Delta U = 0$) despite the fact that ΔQ and ΔW individually are not zero.

Let us apply the first law to a very simple process, the flow of heat from a body at a temperature T_2 to another, at temperature T_1, with which it is in contact. In this process the hotter body loses an amount of internal energy ΔU_2, the colder one gains ΔU_1, and if the two bodies are insulated from their surroundings these two quantities are numerically equal. The total change in internal energy of the system (consisting of the two bodies in contact) will thus be zero. Similarly the heat lost by one body will equal that gained by the other, so that $\Delta Q = \Delta Q_1 + \Delta Q_2$ will also be zero. No external work was done during the process; hence Eq. (21.1) is obviously satisfied, its left- and right-hand sides being zero.

But imagine now that heat, instead of flowing from the hotter to the colder body, had passed from the colder to the hotter. All quantities, ΔU_2, ΔU_1, ΔQ_2, ΔQ_1, would then have their signs reversed, and Eq. (21.1) would still be satisfied. Such a process, however, does not occur in nature, for heat of itself always flows from high to low temperatures. Hence the first law of thermodynamics, while always true, is not sufficiently restrictive in selecting those processes which actually take place. It must be supplemented by another law which somehow specifies the *direction* in which heat changes occur. This is the function of the second law of thermodynamics.

21.2. The Second Law of Thermodynamics. The oldest and probably the simplest statement of the second law is due to Rudolf Clausius (1822–1888), who announced it in 1850 in the form: *Heat cannot flow from lower to higher temperatures in a self-acting manner.* The qualifying phrase "in a self-acting manner" is extremely important; for it is possible to transfer heat from cold to hot bodies, as in a household refrigerator. The point is that this cannot be done without the expenditure of *external work.*

The history of physics knows innumerable attempts to disprove the laws of thermodynamics. Even today amateurs endeavor to construct perpetual-motion machines that run forever while doing work, in violation of the first law of thermodynamics. None have ever worked. Technically these fictitious devices are known as "perpetual-motion machines of the first kind." Similarly the scientific literature of the last century abounds with descriptions of devices, like steamships that cross the ocean by drawing heat energy from the surrounding water, that contradict the second law. These, known as "perpetual-motion machines of the second kind," are equally futile inventions. For in order to extract heat from the water an engine would have to cool the water, and this can be achieved only by doing an amount of external work greater than the energy released by the water. Experience shows that the second law is never contravened and is quite worthy of forming one of the pillars of physical science.

Strangely enough the second law has never been derived from the laws of dynamics by purely logical reasoning. It can be shown to represent the *most probable* course of events, but not the only possible course. Therefore it should be regarded as a principle which operates above the laws of dynamics, on a plane more general than that of Newton's laws, for example.

To be more fruitful in application the second law must be stated in a precise mathematical way. This requires the introduction of a new physical quantity, called *entropy*, which will now be defined.

***21.3. The Meaning of Entropy.** We consider again an example of heat transfer taking place between two bodies in contact (Fig. 21.1). They constitute what we shall call the "thermodynamic system." If $T_2 > T_1$, a small quantity of heat dQ will be transferred from body 2 to body 1 in some small interval of time. The ratio dQ/T will be seen to have a peculiar significance in thermodynamics. For the present we note that, during the transfer under consideration, body 2 loses the ratio dQ/T_2 while body 1 gains dQ/T_1. The system therefore gains

$$\frac{dQ}{T_1} - \frac{dQ}{T_2}$$

This is a positive quantity. If heat flowed from body 1 to 2, it would be negative. Natural processes occur in such a way as to make this difference positive.

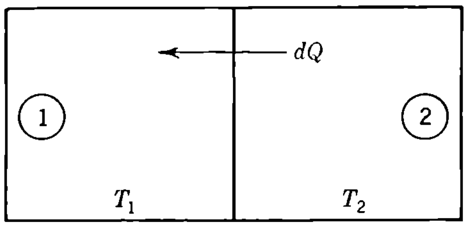


FIG. 21.1. $\dfrac{dQ}{T_1} - \dfrac{dQ}{T_2} > 0.$

For this reason it is indicated that the ratio be given a name and a symbol. We define

$$\boxed{\frac{dQ}{T} = dS} \tag{21.2}$$

T being the absolute temperature, and call S the *entropy*, and dS the change in entropy caused by the transfer of dQ at the temperature T. In our example the increase in entropy of body 1 is $dS_1 = dQ/T_1$, that of body 2, $dS_2 = -dQ/T_2$; and we have seen that the entropy of the system,

$$dS = dS_1 + dS_2 > 0$$

What has here been illustrated by reference to a simple instance is actually a very general situation: the increase in entropy, always defined by Eq. (21.2) for an infinitesimal process, is found to be positive in all processes that occur of themselves in groups of bodies that are insulated from their surroundings. When a finite transfer of heat over a finite range of temperatures takes place, Eq. (21.2) may be generalized to read

$$\Delta S = \int_{T_1}^{T_2} \frac{dQ}{T} \tag{21.3}$$

but only with one important proviso: the transfer of heat considered in the evaluation of the integral must take place so slowly that the system is always in internal equilibrium. Such processes are of great interest in thermodynamics; they are called *quasi-static*.

To illustrate the use of Eq. (21.3) in a different way we calculate the entropy change that one mole of an ideal gas undergoes in an expansion from volume V_1

to volume V_2. In every part of the process the first law is obeyed; hence

$$dU = dQ - dW$$

But for an ideal gas, unless the temperature changes, there can be no dependence of U on volume—there are supposedly no intermolecular forces. Therefore $dU = 0$, and $dQ = dW = p\,dV$ as was shown in Sec. 18.3. We thus see that

$$\Delta S = \int \frac{p\,dV}{T} \tag{21.4}$$

Now, for an ideal gas, $pV = nRT$

and Eq. (21.4) becomes

$$\Delta S = \int nR\frac{dV}{V} = nR \ln \frac{V_2}{V_1} \tag{21.5}$$

Again we see that ΔS is positive, for $V_2 > V_1$. Spontaneous condensations, for which ΔS would be negative, do not take place.

We are now in a position to formulate the second law of thermodynamics in the following simple way: In *natural processes* taking place in *isolated systems* the entropy never decreases.

There are processes, to be sure, in which the change in entropy (of an isolated system) is zero, as, for example, all purely dynamical motions in which $dQ = 0$, and many thermodynamic changes. Such processes are called *reversible*. To give a numerical example, we compute the entropy increase due to the melting of 1 gm of ice. In this process, T = const = 273°, and 80 cal (latent heat of fusion) is taken in. Hence

$$\Delta S = \int \frac{dQ}{T} = \frac{1}{T}\int dQ = \frac{Q}{T} = \frac{80}{273}\text{ cal/deg} = 1.23\text{ joule/deg}$$

On the other hand, when a substance of mass m having a constant specific heat c is heated from T_1 to T_2 degrees absolute the entropy change is

$$\Delta S = \int_{T_1}^{T_2} \frac{mc\,dT}{T} = mc \ln \frac{T_2}{T_1}$$

***21.4. Entropy and Probability.** From its definition [Eq. (21.3)] the reader would hardly suspect the entropy to bear any relation to the probability of a system's state. Yet it is possible to show, by considerations given in books on statistical mechanics, that *entropy is a measure of the probability for the arrangement of molecules to which the state corresponds.* To make this clearer we may compare the molecules with the cards in a deck. If they are dealt at random to four players, it is very unlikely that each player will receive all the cards of one suit, very likely that they are pretty evenly distributed in the different suits. Now it is possible to calculate numerically the probabilities of the various possible distributions of cards among the players; distributions with high probability have a large entropy.

Similar reasoning can be applied to the distribution of molecules in space. When a volume V is available to the molecules of a gas, the state in which they occupy this volume uniformly is more probable than the one in which they are all crowded into part of it. For this reason the entropy of the ideal gas after expansion is greater than before, as indeed we found to be true in the preceding section.

The quantitative relation between the entropy S of a system containing N molecules and the probability w for the arrangement of molecules is

$$S = k \ln w \tag{21.6}$$

This may be shown to be equivalent to definition (21.3), although it cannot be done in this book. But we shall prove that Eq. (21.6) agrees with Eq. (21.5), which was derived for the expansion of an ideal gas.

If nothing is known about the position of a given molecule, then the probability that it may be found in a volume V is proportional to V; the greater V, the greater the chance of finding it in V. Hence the probability of finding a *single* molecule in V is

$$w_1 = cV$$

where c is some constant. Now the probability that N molecules shall be present simultaneously in V is the N-fold product of w_1; in other words

$$w = (cV)^N$$

According to Eq. (21.6),

$$S = kN(\ln c + \ln V)$$

The difference in entropy between a state of volume V_2 and a state of volume V_1 is therefore

$$S = kN(\ln c + \ln V_2) - kN(\ln c + \ln V_1)$$
$$= kN \ln \frac{V_2}{V_1} = nR \ln \frac{V_2}{V_1}$$

and this conforms exactly with Eq. (21.5) because $R = N_0 k$ [cf. Eq. (20.10)].

Relation (21.6) is quite general and shows the meaning of entropy in terms of probability. With this new meaning of entropy the second law of thermodynamics asserts that every natural system changes toward states of greater and greater probability—which is not particularly surprising. Heat flows from high to low temperatures because the state in which all temperatures are equal, *i.e.*, in which all molecules have the same average kinetic energy, is more probable than any other. The second law holds for the same reason that a new deck of cards, when shuffled, takes on a random distribution.

From this point of view a violation of the second law is not an impossibility! If the deck is shuffled long enough, order among the cards may return! If we wait long enough, the molecules of a gas might temporarily all be in one corner of its container or the water in a glass might suddenly freeze on a hot summer day. But if the probabilities for such occurrences are computed, they turn out to be unbelievably small, so small in fact that a period of time far longer than the age of the universe must elapse before such freak phenomena should once take place.

21.5. Thermodynamic Cycles. A set of quantities or variables, such as pressure p, volume V, and temperature T, are said to define the thermodynamic state of a system. In the present section it is well to think of the system as a gas or vapor enclosed in a container of variable volume such as a cylinder with piston. A change from one state to another is called a "thermodynamic process," or "transformation." Such a change can be represented in a number of ways, of which the pV plot is the most common.

An illustration is given in Fig. 21.2, which represents a process in which a gas is taken from a state p_1, V_1 to another whose variables have the values p_2, V_2. The temperature T will in general have different values at different points of the path.

The work done by the gas during the passage from A to B is, according to Sec. 18.3,

$$W_{AB} = \int_{V_1}^{V_2} p\, dV$$

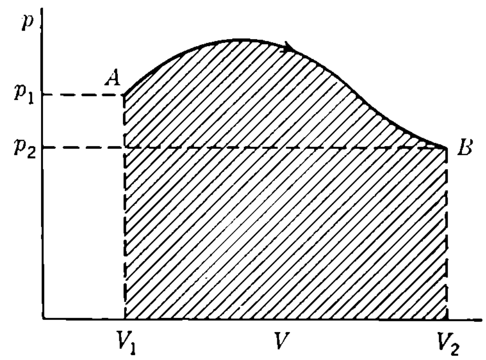


Fig. 21.2. pV diagram.

and this is equal to the shaded area in Fig. 21.2.

A transformation that returns the gas to its original state is called a *cycle*. Figure 21.3 represents a cycle, the gas being taken from A to B and back to A. During the passage from A to B the gas expands and does an amount of work equal to the area under the upper curve. In passing from B to A the gas is compressed and receives an amount of work equal to the area under the lower curve. The net work done by the gas during the cycle is the area on the pV plot that is enclosed by the cycle. In a cycle, although the substance returns to its original state, the net work done is not zero. The reason is that the gas has received heat while performing the cycle and this heat has been transformed into work.

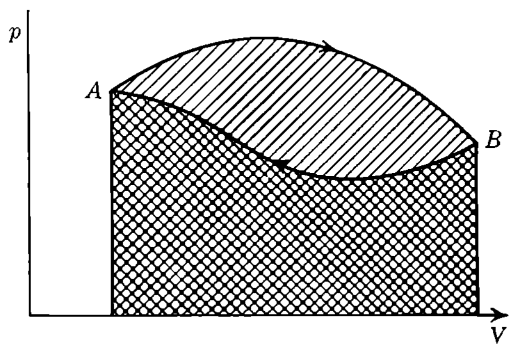


Fig. 21.3. Diagram of a cycle.

If the cycle in Fig. 21.3 had been traversed in the opposite sense, so that what was previously an expansion has now become a compression, and vice versa, the work done by the gas is the area under the lower curve, the work done on the gas is the area under the upper curve, and the work enclosed by the cycle is no longer the work done *by* the gas but that done *on* the gas during the cycle. In general, then, if a cycle on the pV

diagram is traversed in clockwise manner, its area represents work done by the system; if it is traversed in counterclockwise manner, it represents *negative* work done *by* the system, *i.e.*, work done *on* the system by external agencies.

Among the infinite variety of possible cycles, several simple ones are of great practical interest, and one of these will first be singled out for special consideration.

21.6. The Carnot Cycle. Let the "working substance," which need not be a gas but will here be assumed to be one for the sake of simplicity, be enclosed in a cylinder with a piston. The piston and the vertical

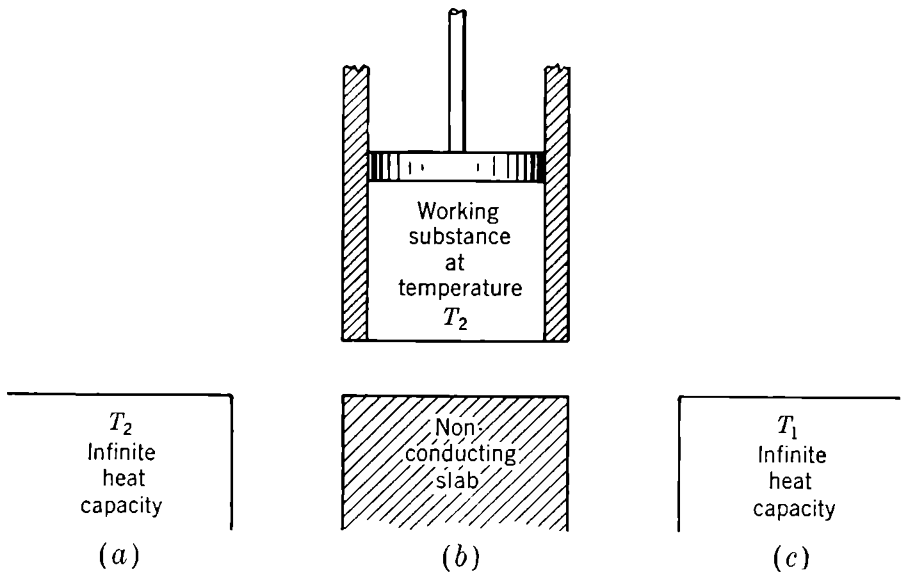


FIG. 21.4. Assembly for Carnot cycle.

walls of the cylinder are made of insulating material, but the base is conducting (Fig. 21.4). In addition to the cylinder assembly we assume the presence of three bodies, all shown in Fig. 21.4. The first is a large body at temperature T_2, a body that for practical purposes has an infinite heat capacity. The second is a nonconducting slab; the third, a body with infinite heat capacity at a temperature $T_1 < T_2$.

In its initial state the working substance is represented by the point A of Fig. 21.5. It is then subjected successively to four processes, which will form a cycle.

1. We place the cylinder upon the heat reservoir a and allow the substance to expand very slowly. During the course of this expansion the gas will be at temperature T_2, and it will take in Q_2 cal of heat from the reservoir. In the diagram (Fig. 21.5) this part of the cycle is represented by the *isothermal* AB. If the gas were an ideal one—which we

are not assuming—the curve AB would be represented by $pV = RT_2$; in general it will be different.

2. Place the cylinder on the slab b, which seals it against heat conduction, and let the gas expand farther to the point C in the diagram. BC is an adiabatic curve for the gas; for an ideal gas it would have the form $pV^\gamma = \text{const}$, as was shown in Sec. 19.6. During passage from B to C the temperature of the gas drops to T_1, for the gas does work but receives no energy from the outside.

3. Put the cylinder on reservoir c, and compress the gas slowly, isothermally. In this process, heat in the amount of Q_1 cal is given off because work is being done on the gas while its temperature remains unaltered. This corresponds to CD in Fig. 21.5. The point D is so chosen that it can be connected with A by an adiabatic curve.

4. Finally we return the cylinder to the slab b, and compress the gas adiabatically until the initial point A is reached. The cycle is now complete.

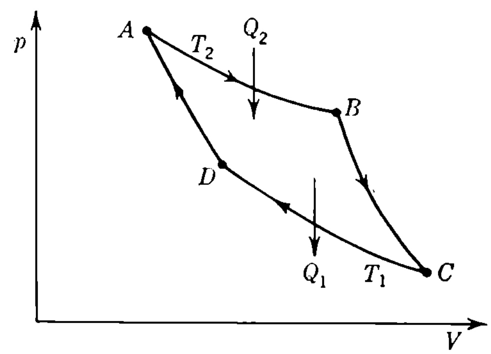


FIG. 21.5. Carnot cycle.

21.7. Efficiency of Engines; Absolute Temperature. The Carnot cycle typifies the action of many engines, although the actual cycle performed by a given heat engine differs markedly from it. The essential facts are these: Q_2 cal of heat is taken from a reservoir at a high temperature T_2; Q_1 cal is given off to another reservoir (the condenser of a steam engine) at a lower temperature T_1. In return for the heat received the working substance performs work equal to the area enclosed by the cycle.

We may apply the first law [Eq. (21.1)] to this cycle, noting that $\Delta U = 0$. Further, writing simply W for ΔW, the work done by the gas, and putting $\Delta Q = Q_2 - Q_1$, we have

$$Q_2 - Q_1 = W \tag{21.7}$$

an equation that directly expresses conservation of energy.

By the "efficiency" of an engine is always meant the work output divided by the work input. In the case of a heat engine the former is W, and the latter might seem to be $Q_2 - Q_1$. As a matter of fact, however, the heat rejected, Q_1, is usually not retrievable and should therefore also be counted as lost. Hence the efficiency of a heat engine is

$$E = \frac{W}{Q_2}$$

and this becomes, by virtue of Eq. (21.7),

$$E = \frac{Q_2 - Q_1}{Q_2} \tag{21.8}$$

Now there are two remarkable things about a Carnot engine, the first proved by Sadi Carnot (1796–1832) himself, the second by Lord Kelvin (1824–1907). One is that the efficiency calculated from a Carnot cycle is the greatest possible efficiency which any engine can attain. It is an *upper limit* to the efficiencies of real engines. The other is that the heats Q_1 and Q_2 are in the same ratio as the *absolute temperatures* T_1 and T_2. This rather unexpected fact is of very great importance and has been used to formulate and determine absolute temperatures in a manner wholly independent of the properties of thermometric substances, the efficiency of a Carnot cycle being the same for every working substance.

Utilizing the second fact, expressible in the form

$$\frac{Q_2}{Q_1} = \frac{T_2}{T_1}$$

we may rewrite Eq. (21.8).

$$E = \frac{T_2 - T_1}{T_2} \tag{21.9}$$

No heat engine can have a greater efficiency than this. Only if T_1, the temperature at which heat is rejected, is zero can the efficiency be 100 per cent.

A steam engine that takes in heat at 200°C, or 473° abs, and discharges its exhaust at 20°C, or 293° abs, cannot have an efficiency greater than

$$\frac{473 - 293}{473} = 0.38 = 38\%$$

even if it is mechanically perfect.

21.8. Refrigerators. The basic principle of refrigeration is easily understood in connection with Fig. 21.5, which, as already stated, represents a cycle for an idealized sort of heat engine. When the cycle is traveled in reverse, it becomes a refrigerator cycle.

In the passage from A to D the gas expands adiabatically, lowering its temperature while doing work. From D to C it expands isothermally, withdrawing Q_1 cal from reservoir c in Fig. 21.4. The compression from C to B is adiabatic, that from B to A isothermal. In this latter part of the cycle, Q_2 cal is given up to the *warmer* reservoir a of Fig. 21.4. As shown in Sec. 21.6, the work done *on* the gas is the area enclosed by the

Carnot diagram. The effects of the cyclic process are therefore: (1) a removal of Q_1 cal from a colder body; (2) a transfer of Q_2 cal to a hotter body; (3) an expenditure of mechanical work on the working substance. In short the system has fulfilled the function of a refrigerator.

From an engineering point of view, of course, this refrigerator, consisting of the assembly drawn in Fig. 21.4, is most unsatisfactory. A more practical arrangement, used in the commercial manufacture of ice (working substance NH_3) and in electric household refrigerators (working substance SO_2 or C_2H_5Cl), is shown in Fig. 21.6. On the downstroke of the compressor piston the substance, existing as a vapor in the cylinder, is forced into the condenser coils on the right, where it becomes a liquid. The latent heat of condensation is carried off by the cooling water, which continually flows around the condenser coils. In passing through the throttle valve V the liquid expands, aided by the upstroke of the piston, so that in the evaporator the substance becomes a vapor again. The heat needed for expansion, and also the latent heat of vaporization, is withdrawn from the water or brine in a tank surrounding the evaporator coils.

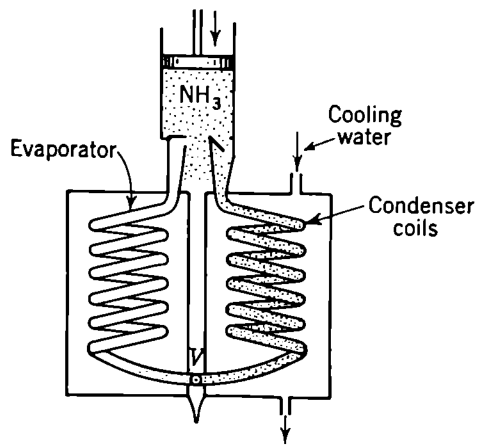


FIG. 21.6. Refrigerator.

This process, it will be observed, contains all the essential features of the Carnot refrigerating cycle; differences are that the substance changes its state and that the contours of the diagram are not exactly isothermals and adiabatics. Also, the real cycle is not truly reversible and has a lower efficiency.

***21.9. Other Cycles.** The working substance of a steam engine is water, whose cycle differs appreciably from that of Carnot because it changes its state from liquid to vapor in the process. The ideal thermodynamic behavior of water in a steam engine is represented by the *Rankine* cycle (William J. Rankine, 1820–1872), sketched in Fig. 21.7. From A to B, water is compressed (isothermally) to the pressure inside the boiler. Since water is nearly incompressible, almost no change in volume takes place in this process. Between B and C there occurs an isothermal vaporization (boiling) of the water; the piston is pushed out in the cylinder, and a certain amount of heat Q_2 is taken from the furnace. From C to D there occurs an adiabatic expansion of the steam, and from D to A an isothermal condensation of steam into water; this requires removal of an amount of heat Q_1 from the water. The work done is again the area of the cycle.

In practice the Rankine cycle is, of course, not realized. But the actual cycle can be traced out on an *indicator diagram,* in which is plotted automatically the pressure in the cylinder vs. the position of the piston, the latter being proportional to V. The result approximates the Rankine cycle, as shown qualitatively by the dotted curve in Fig. 21.7.

An internal-combustion engine, used in automobiles, airplanes, and so forth, has a number of cylinders, one of which is sketched in Fig. 21.8. A complete cycle consists of four

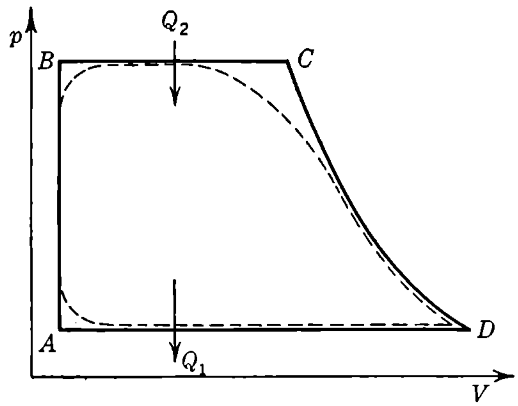


FIG. 21.7. Rankine (steam-engine) cycle.

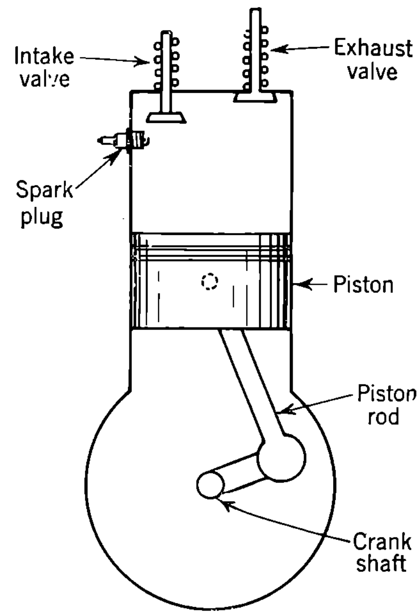


FIG. 21.8. Internal combustion engine.

strokes; its action can be illustrated by reference to the *Otto* cycle drawn in Fig. 21.9. On the first downward stroke, fuel enters through the open valve. This corresponds to the intake part of the Otto cycle, AB, where the volume of the fuel increases without increase in pressure. On the upstroke the piston compresses the fuel, closing the valves. This process corresponds to the adiabatic curve from B to C. Near the end of this upward stroke the explosion occurs, temperature and pressure of the cylinder contents rise greatly and almost without change in volume (part CD of the Otto cycle), and an amount of heat Q_1 is taken in. The piston is now ready to go down again; the hot gases expand (power stroke; portion DE of cycle) nearly adiabatically. At the end of this stroke the exhaust valve opens, and the pressure in the cylinder drops nearly to atmospheric as indicated by EB. Here heat Q_2 is given up by the exhaust gases. Finally on the fourth stroke (upward) the exhaust gases are pushed out of the cylinder, a process that corresponds to BA. The cycle is now completed and about to repeat itself.

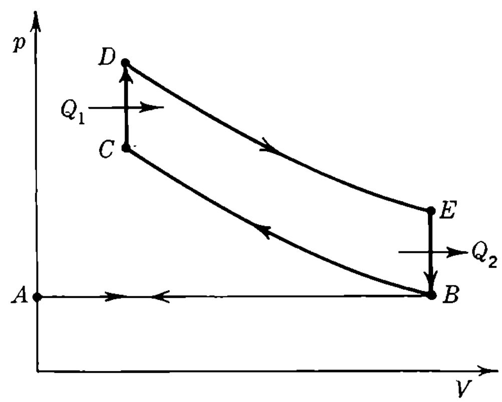


FIG. 21.9. Otto (gasoline engine) cycle.

The *heat pump*, a modern device for cooling a building during the summer and heating it in winter, is a gas engine operating between the temperature of the building and that of the ground. One set of coils runs through the former; one set is embedded in the ground. During the cold season, gas is compressed into the house coils, where the heat of compression is released. The same gas is then allowed to expand into the ground coils, where it momentarily attains a temperature lower than that of the ground and thus absorbs heat from it. Then the cycle starts over again, the gas being compressed into the house coils, and so forth.

In the warm season the cycle is reversed: gas is compressed into the ground coils and expanded into the house coils, so that the ground is heated and the house is cooled.

***21.10. Heat Death of the Universe.** Before closing this chapter we shall give one final thought to the second law of thermodynamics. To some philosophers it has had frightening aspects, for it forebodes a future state in which the entropy of the universe is a maximum, in which all things are of one common temperature, in which there is no relative motion of parts. Indeed this is the condition toward which the second law propels the universe, a condition in which life is impossible and which is therefore known as the "heat death" of the universe. To remove this unpleasant prospect, scientists have claimed that the second law is not of universal validity; that it holds only for inorganic and not for living matter; that the living cell produces what has been called "ektropy" (negative entropy); that mental processes "use up" entropy and therefore violate the second law.

Actually there is evidence for none of these assertions. So far as we know, the second law holds for the processes taking place in living matter as accurately as in the inorganic realm. The heat death of the universe cannot be forestalled by such specific proposals. But there are a few general observations that should properly be made concerning it.

First the heat death is an extremely long time away. No doubt the planetary system, and man, will have disappeared long before its occurrence. Second there may be distant regions of our vast universe, not known to us, where the second law actually fails. Cosmic rays continually *decrease* the entropy of the earth upon which they impinge, and their origin is not clear at present. Last, and perhaps most important, the universe may not be a closed, or isolated, system, in which case the second law is inapplicable to it as a whole. The heat death, with its unhappy prospects for our beloved universe, is therefore by no means a certainty.

PROBLEMS

1. Using the first law of thermodynamics, show that when a substance undergoes an isochoric (constant-volume) change the increase of internal energy equals the heat absorbed.

2. A gas is heated and receives 30 cal from a burner. At the same time it expands against a constant external pressure of 1 atm, its volume increasing from 1 to 1.5 liters. What has been the change in internal energy?

3. A gas is heated and receives 1.2 Btu from a burner. At the same time it expands against a constant external pressure of 2 atm, its volume increasing from 0.6 to 0.9 ft^3. What has been the change in internal energy in foot pounds?

★**4.** Why is the entropy change in a (reversible) adiabatic process zero?

5. A Carnot engine works between 500 and 20°C. What is its efficiency?

6. A Carnot engine has its low-temperature reservoir at 50°F, and its efficiency is 30 per cent. What is the temperature of the other reservoir? What should it be if the efficiency were 50 per cent?

★**7.** Consider the entropy of water at 32°F to be zero. Calculate (*a*) the entropy of 1 lb of ice at 32°F; (*b*) the entropy of 1 lb of ice at 0°F (specific heat of ice is 0.5 Btu lb^{-1} deg F^{-1}); (*c*) the entropy of 1 lb of water at the normal boiling point; (*d*) the entropy of 1 lb of steam at 1 atm and 212°F. Express your answers in Btu/deg F abs.

★**8.** One pound of air (which for the present purposes may be considered as an ideal gas) expands slowly until its volume is doubled. Compute its change in entropy.

9. Plot an exact Carnot cycle on a pV diagram for one mole of an ideal gas between temperatures 300 and 100° abs. Let the point A correspond to $p = 1$ atm, the point B to 0.5 atm. Compute, graphically or otherwise, the work done in this cycle. Take $\gamma = 1.5$.

CHAPTER 22

TRANSFER OF HEAT

22.1. Types of Heat Transfer. Wherever a difference in temperature exists, a flow of heat energy occurs from the region of higher to that of lower temperature. Sometimes the engineer wishes to better the heat transfer, as, for example, from the firebox to the hot water in a boiler or from the motor block to the circulating coolant in a hot motor. In other instances the problem is to minimize the heat flow, say through the walls or roof of a dwelling or through the insulated sides of a refrigerator. In all cases of heat transfer there are involved one or all of the following processes: *convection*, *conduction*, and *radiation*.

Convection is heat transfer by the mass motion of fluid matter, either liquid or gas. It is brought about by the changes in density of the fluid caused by heating. Familiar examples of convection are the draft in a stove, the warming of all the air in a room by a hot stove in one corner, and the circulation of the water in a hot-water heating system.

Conduction is the transfer of heat by matter but without any motion of the matter itself. When one end of an iron poker is placed in a fire, in a short time the handle becomes warm. The kinetic theory makes this understandable; for the molecules of the iron at the hot end are set into violent motion, and because of the strong forces binding the molecules together this increased motion is passed along successively to the other molecules down the length of the rod. The good conductors of heat are the metals, and of these the metals which are the best electric conductors are also the best heat conductors. Since electric conduction is explained as a drift of electrons, more or less free, through the metal, it must be that the motion of these electrons is in good measure responsible for the conduction of heat.

Radiation is the process of the conversion of the heat energy in matter into electromagnetic radiation, similar in nature to light, and the subsequent reconversion of this radiant energy into heat by absorption by any substance upon which the radiation falls. Heat radiation and light travel through empty space at the huge speed of 186,000 miles/sec. We receive energy from the sun by this process, and a beam of radiant energy is sent forth by reflection from the parabolic mirror of an electric heater. Most of the heat transfer from steam radiators to the air in a room, however, is by conduction and convection. Still it is true that, if you hold your hand to one side of a hot steam radiator, the warmth you receive comes by the radiation process. In fact the walls of an enclosure and all the objects within are continuously radiating energy to each other whether they are at the same or different temperatures.

22.2. Convection. Of the three methods of heat transfer, convection is the easiest to understand, for it involves only the usual laws of fluid mechanics. The household water heater provides a good illustration of convection in a liquid. The water in the copper coil C (Fig. 22.1) becomes heated by conduction through the metal pipe. This hot water, being less dense, rises to the top of the tank, while cold water from the bottom of the tank enters the coil and is in turn heated. Thus by convective circulation all the water in the storage tank becomes heated.

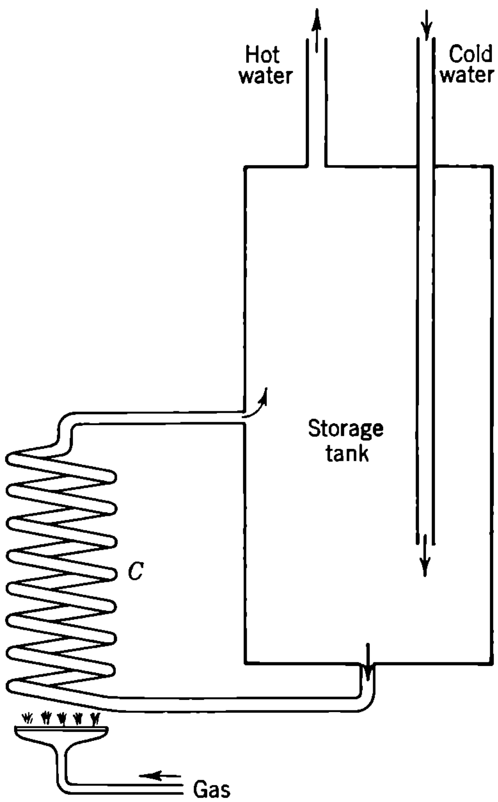


FIG. 22.1. Convective circulation of the water in a hot-water tank.

All heating systems for dwellings make large use of convective circulation for the distribution of the heat. The draft in a stove or chimney is produced by convection. For the heated air in a chimney, being lighter than the corresponding column of cold air, will experience a buoyant force, by Archimedes' principle. The draft is better the taller the chimney; for the greater the pressure difference between the top and bottom of the chimney, the greater this buoyant force.

In cold weather the surface water in a lake is the first to cool, thus becoming heavier than the warmer layers below. This surface water consequently sinks and is replaced by the lighter water from below. This new surface water is then cooled, sinks, and is replaced by more of the warmer water from below. Thus by convection the whole body of water is cooled to 4°C, the temperature at which the water has its maximum density. Further cooling of the surface water makes it lighter, with the result that it does not sink and all convective circulation ceases. It is the surface then which freezes first, so that marine animal life is protected and rivers in cold climates do not become choked with ice building up from the bottom.

Ocean currents and winds are often large-scale examples of convection. The Gulf Stream and the Japan Current are at least to some extent caused by convection, the warm surface ocean currents originating near the equator being balanced by a return cold current near the ocean floor. Land and sea breezes are local convection currents in the atmosphere. During the morning the land becomes warmer than the near-by water, both because it absorbs the solar radiation more rapidly and because of its lower specific heat. The warmer air above the land then expands, rises, and is replaced by the cooler sea air. At night the land loses its heat by radiation more rapidly than does the sea, and hence the return current of cooler air is from the land to the sea. As a final example we should mention the trade winds, which blow steadily in regions within a few degrees of latitude of the equator. The considerable heating of the atmosphere near the equator causes this air to expand and rise, and the cooler air comes in from north and south of the equator to replace it. Because of the rotation of the earth these return currents are somewhat deflected from the north-south line.

The analytical treatment of convection currents forms a rather difficult branch of hydrodynamics and will not be considered here. Dispensing with approximate analyses we shall limit our discussion of convection to the few examples cited above, which should serve to emphasize that the convective circulation of fluids is a very common and important process in the physical world.

22.3. Conduction. The amount of heat that is transferred by conduction between two surfaces, one of which is at a higher temperature than the other, depends greatly on the character of the substance filling the intervening space. As already mentioned, the metals are the good conductors of heat energy. Poor conductors, or good heat insulators, are nonmetallic substances such as dry wood, cloth, leather, and glass. Porous materials like cork, sawdust, asbestos, and rock wool are, because of their fine subdivision and their air holes, the best heat insulators. The heat conducted through any cross section of the substance depends also on the *temperature gradient*, dT/dx, at the point and on the area of the section. Consider the heat flow after steady-state conditions have been reached in a bar of length L and cross-sectional area A, the temperatures of the ends of the bar being T_2 and T_1, with $T_2 > T_1$ (Fig. 22.2). The average temperature gradient along the entire bar is the ratio $(T_2 - T_1)/L$, and at any particular section x the temperature gradient is the ratio of the temperature difference dT between two parallel sections flanking that at x to the distance dx between them as this interval becomes infinitesimally small. Finally it is obvious that the amount of heat conducted through the bar should be proportional to its cross-sectional area A.

Denoting quantity of heat by Q, we may then write, for the rate of flow of heat through any section of area A,

$$\frac{dQ}{dt} = -KA\frac{dT}{dx} \tag{22.1}$$

where the factor of proportionality K is a constant called the *thermal conductivity* of the substance. The minus sign indicates that the heat flow dQ/dt is positive, *i.e.*, from left to right in Fig. 22.2, when the temperature gradient dT/dx is negative in this direction. If the rate of flow of heat is in calories per second, the area in square centimeters, and the temperature gradient in centigrade degrees per centimeter, the units for the conductivity K are calories divided by (second·centimeter·degree centigrade). Engineers in English-speaking countries often express the temperature gradient in Fahrenheit degrees per inch, the area in square feet, and the rate of heat flow in Btu per hour. With these units K is the number of Btu that is conducted in one hour per square foot of material one inch in thickness, with the two surfaces at a temperature difference of one degree Fahrenheit. In Table 22.1 are listed average values of thermal conductivities for several materials.

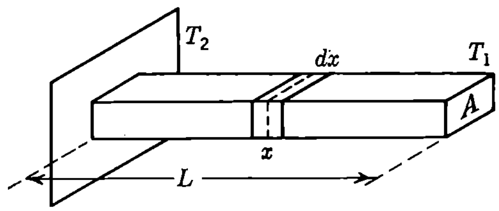


Fig. 22.2. Heat flow along a bar with end temperatures T_2 and $T_1 (T_2 > T_1)$.

Table 22.1. *Thermal Conductivities at Ordinary Temperatures*

Substance	cal cm^{-1} sec^{-1} deg C^{-1}	Substance	cal cm^{-1} sec^{-1} deg C^{-1}
Aluminum	0.49	Cement	0.0007
Brass	0.26	Cork	0.0001
Copper	0.92	Glass	0.0025
Iron	0.15	Ice	0.005
Lead	0.08	Silk	0.0001
Silver	0.98	Water	0.0013
Steel	0.11	Air	0.000057
Substance	**Btu in. ft^{-2} hr^{-1} deg F^{-1}**	**Substance**	**Btu in. ft^{-2} hr^{-1} deg F^{-1}**
Hair felt	0.26	Sawdust	0.41
Rock wool	0.26	Celotex	0.34
Glass wool	0.29		

To measure the thermal conductivity of a metallic substance the specimen should be in the form of a rod or bar. The surface of the bar must be well insulated to prevent heat losses to the surroundings. Heat is supplied electrically at one end, and in the steady state the heat arriving at the other end may be determined by having a continuous flow of water in contact with the end of the bar. If the rate of flow of the water and its temperature rise are known, the quantity of heat arriving at this end may be computed. The temperature gradient down the bar is best measured by thermocouples buried in small holes in the bar at known intervals down its length. The input electric power in watts divided by the mechanical equivalent of heat, 4.186 joules/calorie, gives the heat input in calories/sec. An arrangement for such measurements is shown in Fig. 22.3.

Essentially the same procedure may be used for the measurement of the thermal conductivity of a heat-insulating material, but because of the low

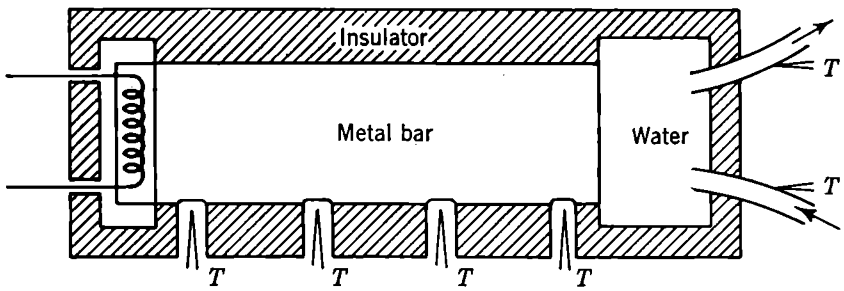


FIG. 22.3. Determination of the thermal conductivity of a bar. T indicates a thermocouple for temperature measurement.

conductivity it is necessary to have the specimen in the form of a sheet or slab of fairly large surface. Conductivities of liquids or gases are measured only with difficulty because of ever-present convection. To minimize convection in the fluid sample, heat must be supplied at the top and removed at the bottom. Inspection of Table 22.1 shows that air has relatively a very small thermal conductivity. This accounts for the insulating qualities of materials such as the five at the end of Table 22.1, for all such substances have much air trapped in pores or between the small particles of finely divided matter.

***22.4. Heat-flow Calculations.** We may apply Eq. (22.1) to get the temperature at any point along a bar (Fig. 22.2) when the heat flow has reached a steady state and assuming perfect insulation along the sides of the bar. Since the heat flows at a constant rate, we set $dQ/dt = H$, a constant. Then,

$$\frac{dT}{dx} = -\frac{H}{KA} \tag{22.2}$$

and since H/KA is independent of x, we obtain, by integration,

$$T = -\frac{H}{KA}x + C \tag{22.3}$$

where C is a constant of integration. To evaluate this constant we take $x = 0$ at the hot end where $T = T_2$, while, at the other end, $x = L$, $T = T_1$. Upon placing these values in Eq. (22.3), $C = T_2$, and

$$T_1 = -\frac{H}{KA}L + T_2 \tag{22.4}$$

Therefore Eq. (22.3) may be written

$$T = -\frac{H}{KA}x + T_2 \tag{22.5}$$

On solving Eq. (22.4) for H and substituting in Eq. (22.5), there results

$$T = \frac{(T_1 - T_2)x}{L} + T_2 \tag{22.6}$$

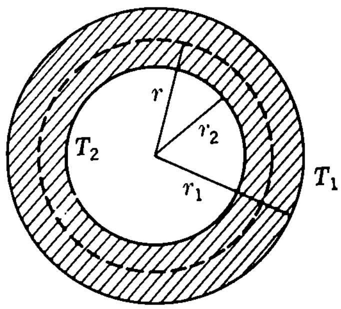


FIG. 22.4. Conduction of heat through a pipe wall.

showing that the temperature varies linearly with distance along the bar.

A more complicated case is the important one of heat transfer through the cylindrical wall of a pipe. Suppose the pipe of length L to carry steam at a temperature T_2, with the outer surface at temperature T_1, the inner and outer radii being r_2 and r_1, respectively (Fig. 22.4). Take the radius of any cylindrical surface, such as the dotted one in the sectional drawing of Fig. 22.4, to be r. Then, in Eq. (22.2), $A = 2\pi rL$, and hence we may write the steady-state equation of heat flow

$$\frac{dT}{dr} = -\frac{H}{2\pi rLK} \tag{22.7}$$

or

$$dT = -\frac{H}{2\pi LK}\frac{dr}{r} \tag{22.8}$$

Integrating between the limits of T and r for the inner and outer surfaces of the pipe,

$$\int_{T_2}^{T_1} dT = -\frac{H}{2\pi LK}\int_{r_2}^{r_1}\frac{dr}{r} \tag{22.9}$$

we have

$$T_1 - T_2 = -\frac{H}{2\pi LK}\ln\frac{r_1}{r_2} \tag{22.10}$$

Therefore the rate of flow of heat radially out through the pipe wall is

$$H = \frac{2\pi LK(T_2 - T_1)}{\ln(r_1/r_2)} \tag{22.11}$$

This radial heat flow through cylindrical walls may conveniently be used to measure the conductivity K of the walls. The inner surface may be heated electrically, or the cylinder itself may carry the electric current. When steam or other vapors and liquids are in contact with the pipe walls, thin films of these liquids probably adhere to the walls. Such films have rather low thermal conductivities and hence definitely affect the heat flow to the surrounding medium.

22.5. Radiation. The radiant energy emitted by a body is not heat, but upon absorption by another surface it is converted back into heat energy. Heat radiation is, in fact, electromagnetic radiation like light, radio waves, and X rays, and it exhibits such optical properties as reflection, refraction, dispersion, interference, and polarization. Electromagnetic waves consist of an electric and a magnetic force varying periodically and directed transverse to the direction of propagation, and the transmission of such disturbances is a property of empty space. The experimentally produced radiation spectrum extends in wavelength from roughly 10^{-9} mm (hard gamma rays) to effectively infinity (radio waves from a slowly oscillating charge). The eye is sensitive merely to the short range from about 4×10^{-4} mm (violet) to 7.5×10^{-4} mm (red). Heat radiation comprises the range of wavelengths longer than 7.5×10^{-4} mm, hence infrared (Latin *infra*, below), merging at about 1 mm into the short-wavelength limit of the so-called "microwaves," which may be produced by electric circuits.

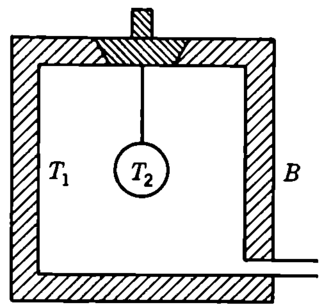


FIG. 22.5. A body in an evacuated box comes to thermal equilibrium with the walls.

Suppose that the box B (Fig. 22.5) has its walls maintained at a constant temperature T_1 and that we suspend inside it, successively, various small objects, some with well-polished, reflecting surfaces, others with dull, blackened surfaces, and all at different temperatures T_2. Each time we promptly evacuate the box. Regardless of their initial temperatures T_2 and the nature of their surfaces the bodies always come to the same temperature T_1 as that of the walls. The body on the whole radiates to the walls. If, however, T_2 is initially lower than T_1, the net exchange of radiation is from the walls to the body. Therefore, when the two come to thermal equilibrium, there must be equal amounts of radiation streaming both ways. Now, of the radiant energy from the walls incident on the body, part is reflected, and part is absorbed. After thermal equilibrium at the temperature T_1 is established, though, each of the bodies must emit radiation at the same rate as it absorbs it. That is, if one of the bodies is a good absorber, it must also be a good emitter, and vice versa. And since the amount of radiation from the walls incident on each of these bodies per second is the same, the body that is a poor absorber must be a good reflector. Therefore a good reflector is a poor emitter.

An instructive experiment corroborating this was first performed in 1804 by Sir John Leslie. A hollow metal cube with one side highly polished, another painted with dull black paint, another with a mat-white paint, etc., is filled with hot water, thus bringing the whole to the same temperature. When each face in turn is presented to a thermopile (Chap. 31), which connected to a sensitive galvanometer forms one of the best radiation-measuring devices, the surfaces exhibit marked differences in the amount of radiation being emitted. The largest current in the thermopile-galvanometer circuit is produced by the dull black surface, showing that while it is the best absorber it is also the best emitter. The polished surface, on the other hand, reflects most of the radiation incident upon it back into the cube and emits but little to the thermopile.

There are many practical applications of these facts. In the double-walled vacuum, or Thermos bottle, the space between the two walls is evacuated to eliminate heat flow by conduction and convection. These interior walls are also silvered to reduce the emission of radiant energy. In the walls of dwellings and refrigerators a layer of smooth, nontarnishing aluminum foil will aid materially in reducing the heat transmission through the wall because of the high reflecting power and hence low emitting power of the shiny metal surface.

A body having a black surface that absorbs perfectly all the radiation striking it is called an ideal *black body*. Such a body would also be an ideal radiator and would therefore emit more radiation than any other surface at a given temperature. No actual surface is perfectly black, for even lampblack reflects about 1 per cent of radiation incident on it. Basing his ideas on experiments by John Tyndall (1820–1893) concerning the total radiant energy emitted by a hot platinum wire, Joseph Stefan (1835–1893), in 1879, conjectured that the total radiation from a heated body is proportional to the fourth power of its absolute temperature. A few years later, in 1883, L. Boltzmann (1844–1906) showed that this fourth-power law is to be expected theoretically for a perfect radiator. *Stefan's law* is usually written

$$W = \sigma T^4 \tag{22.12}$$

where W is the rate of emission of radiant energy per unit area. If this is in ergs $\text{sec}^{-1}\ \text{cm}^{-2}$, the constant σ has the value 5.7×10^{-5}, and if W is given in watts/m^2 (mks system of units), σ has the numerical value 5.7×10^{-8}. For any real emitting surface we should write this law in the form

$$W = e\sigma T^4 \tag{22.13}$$

where e is the *emissivity* of the surface. For copper, e has a value of about 0.3. According to Eq. (22.12) a black body at 300°K radiates about 462,000 ergs sec^{-1} cm^{-2}. At 1000°K, when the body just begins to glow with a dark-red color, the rate of emission is about 125 times greater than at room temperature (300°K).

As we have already seen, a body is continuously exchanging radiation with its surroundings. If a body at temperature T_2 is surrounded by walls at a temperature T_1, the net energy radiated is

$$W = e\sigma(T_2^4 - T_1^4) \tag{22.14}$$

Further discussion of the emission and absorption of radiation will be found in Chap. 48.

PROBLEMS

1. An icebox has an inner volume 2 by 4 by 1½ ft, and the walls may be assumed to be made entirely of cork 2 in. thick. How much ice will melt in 1 day if the interior is to be maintained at 42°F while the outside temperature is 72°F? Take the conductivity of cork as 0.3 Btu in. ft^{-2} hr^{-1} deg F^{-1}.

2. A room 12 by 18 by 12 ft is completely insulated on all surfaces by 3 in. of rock wool. If the temperature of this room is to be maintained at 70°F while the temperature outside is 30°F, how much heat must be supplied to the room per hour to compensate for the loss by conduction through the walls?

3. Window glass has a thermal conductivity of approximately 7.24 Btu in. ft^{-2} hr^{-1} deg F^{-1}. Calculate the amount of heat conducted in 1 hr through a window whose glass measures 3 by 4 ft and is ⅛ in. thick.

4. One end of a copper bar 4 cm^2 in cross section and 40 cm in length is in a steam bath, while the other end is in a mixture of ice and water. Assuming that the sides of the bar are perfectly insulated, find the rate of heat flow in the bar and the temperature at a point 10 cm from the hot end.

***5.** A steam pipe 100 m in length and 5 cm in diameter is covered by a tight-fitting cylindrical jacket of insulating material 2 cm thick and of conductivity 0.0003 cal cm^{-1} sec^{-1} deg C^{-1}. The temperature of the steam pipe is 100°C, and that of the outer surface of the jacket is 20°C. Calculate the heat lost through the jacket per hour.

***6.** Calculate an expression for the rate of flow of heat through a spherical shell of inner radius R_1, outer radius R_2, and conductivity K.

7. A compound wall consists of three layers each 2 cm thick and of thermal conductivities 0.01, 0.02, and 0.03 cal cm^{-1} sec^{-1} deg C^{-1}, respectively. Calculate the conductivity of a single layer of material 3 cm thick that will transmit the same heat flow for the same temperature difference.

8. Calculate the energy in watts radiated per square centimeter by a block of copper at 400°C and at 1000°C.

9. How many calories per second will be emitted by a spherical black body 10 cm in diameter at a temperature of 600°C?

CHAPTER 23

METEOROLOGY

23.1. The Atmosphere. Meteorology, the study of the weather, is one of the most recent branches of physical science. Its development has been retarded, not for lack of interest—this being proverbially present among all peoples and at all ages—but by the difficulties that beset this science. Meteorology involves a thorough investigation of the earth's atmosphere, an extremely difficult subject.

If a uniform gas is enclosed in a vessel, its behavior is easy to understand in terms of the laws of fluid mechanics (Chap. 13). If it is uniformly heated, the laws of thermodynamics (Chap. 21) enable us to predict the changes in temperature, pressure, or volume that will occur. The situation is less simple when the heating is not uniform; for then convection currents will be created. A further complication arises when the walls of the vessel are not smooth but irregular, so that the currents will suffer resistance near the walls of the vessels. All these difficulties are minor in comparison with those which arise when the gas itself is not homogeneous but contains condensable vapors, like water. Finally, imagine that the whole vessel with its contents is in a state of rotation and you have a fair appreciation of the troubles with which the meteorologist has to contend. For the atmosphere is indeed a system of gases and vapors enveloping the rough surface of the earth like a relatively thin film, unequally heated, and rotating with the earth.

The ultimate source of atmospheric heat is the sun. However, as the sun's rays pass through the atmosphere on their way to the earth, only about 10 per cent of their energy is absorbed, the remainder serving to heat the earth underneath. The atmosphere is then heated secondarily by the long-wave heat radiation emitted from the earth, of which it absorbs about 90 per cent. This is sometimes called the "greenhouse effect" of the atmosphere because of its similarity to the action of glass in a greenhouse. As a result the atmosphere, instead of being uniformly heated by the solar rays, takes on the local temperature pattern of the portion of the earth that it covers. Tropical air masses are usually hot; polar masses, cold.

The difference between *land* and *sea* is of importance in this connection. The absorptive properties of both are about equal, but they differ with respect to *conductivity* and *thermal capacity*. The conductivity of water is much better than that of rock; hence water takes on heat faster by exposure to the sun. The thermal capacity of one cubic centimeter of substance is the product of specific heat and density. For rock, these are approximately 0.15 cal gm^{-1} deg C^{-1} and 2 or 3 gm/cm^3; for water, both values are 1. Hence water, while conducting heat more readily, also requires more heat to be in thermal equilibrium with land; once heated, it must lose more calories than land to have its temperature reduced. Water therefore acts like a heat reservoir, and the sea has a tempering effect on coastal climates.

The vertical height of our atmosphere cannot be specified since it shades gradually into the interplanetary vacuum. Nevertheless it can be said that the air column standing above every square inch of the earth's surface weighs 14.7 lb. If its density were constant and equal to its value at sea level, the whole atmosphere would extend only to a height of about 5 miles. Actually the pressure (and hence the density) of air decreases in a manner roughly described in Sec. 13.10. Near the earth's surface the pressure drops by about 1 in. of mercury for each 1,000 ft increase in elevation.

The vertical variation in temperature is somewhat peculiar. As we proceed upward from sea level, T drops by about 3°F per 1000 ft. A decrease at this approximate "lapse rate" continues until an elevation of about 7 miles is reached. The temperature has here attained a value of −60°F, and it stays at about this value for several miles.

The part of the atmosphere below a height of 7 miles, where the temperature varies rapidly and the air is agitated—in short, the weather layer—is called the *troposphere;* above it lies the *stratosphere*, a quiescent body of air at low though constant temperature. All clouds are confined to the troposphere, and our interest will henceforth be directed mainly to this lowest stratum of air. But it should be remarked that other layers of the atmosphere with unusual electrical properties have been found to exist above the stratosphere. To these the name *ionosphere* has been given; they are important in explaining the propagation of radio waves around the curved surface of the earth.

***23.2. Meteorological Measurements.** Weather forecasting is based on two things—accurate knowledge of the condition of the atmosphere to as large an extent as possible, and an understanding of physical principles. The condition of the atmosphere is determined by making suitable measurements, some of which will now be discussed. Until the last decade or two, these measurements were made at the earth's surface, so that knowledge of atmospheric conditions was confined largely to the very lowest stratum. With the conquest of the air by man has come the possibility of an analysis of the troposphere at all heights: weather forecasting has become three-dimensional and has thus attained an unprecedented accuracy. The use of balloons and airplanes to carry measuring devices aloft is now very widespread.

Perhaps of greatest interest to the meteorologist is the *pressure* at the various points in the atmosphere. Everyone knows that a falling barometer indicates the coming of rain, and this chiefly for two reasons: (1) Water vapor saturated at ordinary temperatures is lighter than air; hence a low pressure may indicate its presence and so the probability of precipitation: (2) Air masses move toward regions of low pressure, and their coming invites turbulent, unsettled conditions promoting rain. Instruments used for measuring pressure are the mercury and the aneroid barometer, which were described in Sec. 13.6.

The *quantity of water vapor* present in the air, or the humidity, is of great importance, and its determination will be explained in the next section. The *temperature* is, of course, measured by means of thermometers and need not be discussed further, except to point out that the maximum-minimum thermometer, described in Sec. 16.8, finds much application in weather forecasting. Continuously recording thermometers are in constant use. The "mean temperature," as the term is used by the U.S. Weather Bureau, is not the time average over a day but simply one-half the sum of maximum and minimum temperature for the day.

Much attention has recently been paid to the vertical temperature gradient in the lower atmosphere. In certain parts of the world, notably in the near-desert regions of California, there occur abnormalities and local inversions of this gradient. These have been found to have a remarkable effect upon the climate of that region, and incidentally upon the propagation of radio waves.

The *direction of the wind* is determined by weather vanes, its *velocity* by a device known as a "cup anemometer." This instrument consists of several (usually three) hollow aluminum hemispheres mounted on arms extending from a vertical axis, so that they will revolve in the wind, their speed depending on the wind velocity. They operate a set of gears, which directly indicates wind velocity in miles per hour. The customary scale of wind velocities is one devised by the British admiral, Sir Francis Beaufort, in 1805; it is given in Table 23.1.

Table 23.1. *Beaufort Scale of Wind Velocities*

Beaufort number	Name	miles/hr	Description
0	Calm	1	Smoke rises vertically
1–2	Light breeze	1–7	Leaves rustle
3	Gentle breeze	8–12	Twigs in motion
4	Moderate breeze	13–18	Small branches move; dust is raised
5	Fresh breeze	19–24	Small trees sway
6–7	Strong wind	25–38	Large branches, whole trees move
8–9	Gale	39–54	Tree limbs break
10–11	Whole gale	55–75	Trees uprooted
12	Hurricane	>75	(Must be seen to be appreciated)

Rainfall is recorded by means of a rain gauge, a tiny bucket that tips each time $\frac{1}{100}$ in. of rain has fallen into it, the number of tips being mechanically counted. There is also an instrument known as a *sunshine recorder*. It is essentially a black-bulb thermometer, the mercury column of which closes an electric contact whenever it reaches a height which it attains only in sunshine.

23.3. Atmospheric Humidity and its Measurement. The maximum amount of water vapor that can be present in air corresponds to saturation. At saturation the pressure of the vapor is an increasing function of the temperature, as was shown in Sec. 18.5. In Fig. 23.1 we have redrawn the so-called "vapor-pressure curve" for water.

Under ordinary atmospheric conditions the amount of water vapor present is less than that which would be present at saturation, and it

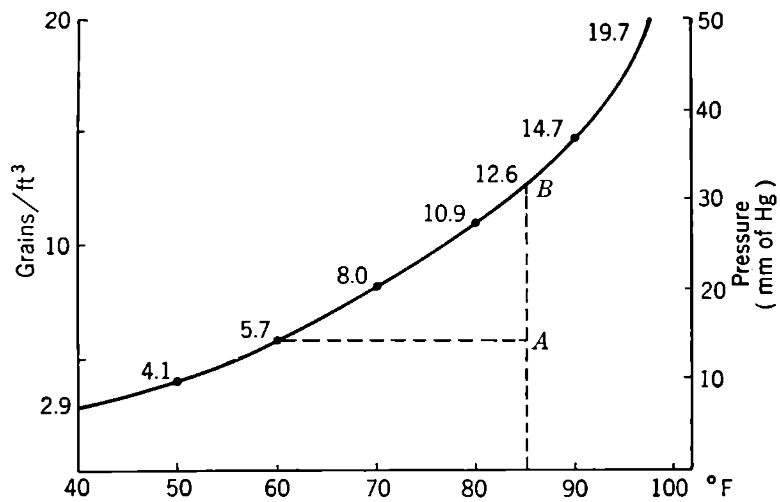


FIG. 23.1. Density and pressure of saturated water vapor.

becomes necessary to define a suitable measure, called "degree of humidity," for the water content of the air. Two such measures are in use. One is spoken of as *absolute humidity;* it is the actual mass of water vapor in a given volume of air and is expressed in gm/m³, or more commonly in the United States, in grains/ft³. Absolute humidity varies greatly from place to place and is not a very significant meteorological index. More often used is *relative humidity*, defined as the ratio of the mass of water actually present in a given volume of air to that present at saturation. This ratio is expressed in per cent. An example illustrating the meaning of these terms is given in Table 23.2.

Since at a given temperature the density of water vapor is very nearly proportional to its pressure, relative humidity may also be taken to be the ratio of actual vapor pressure to saturation pressure at the temperature in question. Thus, in Fig. 23.1, if at the temperature 85°F the actual water-vapor pressure were represented by A while saturation corresponds to B, the relative humidity would be the ordinate of A divided by the ordinate of B.

There is a simple experiment by which this ratio can be obtained. Assume that the point A represents the humidity, the temperature being 85°F. Place a thermometer in a metal cup polished to a high luster on the outside, and filled with water. Now drop a few pieces of ice into the water, so that its temperature will fall. Suddenly a haze indicating condensation of vapor will form on the outside of the cup. Note the temperature at which this happens. Let us say it occurred at 60°F. This temperature is known as the *dew point* of the air; it is the temperature at which the air would be saturated under the circumstances of humidity actually prevailing.

Table 23.2. *Relative Humidities*

Temperature, °F	Abs humidity, grains/ft^3	Rel humidity, %
40	2.9	100
60	2.9	51
80	2.9	27
90	2.9	19

Translated into terms of Fig. 23.1 our experiment indicates that a horizontal line, traced from point A to the left, intersects the curve at 60°, the dew point. But we see that at this temperature the saturated vapor has a density of 5.7 grains/ft^3, and this must therefore be the actual density of the vapor. Hence the relative humidity is 5.7/12.6, or 45.2 per cent.

In practice the method just described for finding atmospheric humidity is neither convenient nor accurate. Two devices are in use. The more important one is the *wet- and dry-bulb hygrometer* (Geek *hygros*, moisture). This is an ordinary thermometer, put to double use. First it serves to measure the air temperature in the ordinary way. Then a piece of loosely woven muslin is tied around the bulb and moistened, and the instrument is whirled vigorously so that rapid evaporation from the cloth takes place. Thus the temperature will fall, the fall being greater the smaller the relative humidity of the air. The latter can then be read from charts accompanying the instrument.

The second type of hygrometer that enjoys some popularity is the *hair hygrometer*, whose action is based on the circumstance that human hair, deprived of its oil content, lengthens as the relative humidity increases. By winding hair around a spindle, changes in length can be made obvious in numerous ways, such as the emergence of different figures from a toy weather house.

★23.4. Visibility and Ceiling. The prevalence of air transportation has brought into common use two meteorological terms, visibility and ceiling. *Visibility* is the greatest distance in miles at which large objects on the ground, such as buildings, can be seen. In aviation weather reports the number for visibility is usually followed by a letter indicating the obstruction to vision. Thus ¼ F implies a horizontal range of vision of ¼ mile owing to fog.

By *ceiling* is meant the distance between the ground and the lower surface of the lowest cloud layer covering more than half the sky. It is expressed in numbers indicating the ceiling in hundreds of feet. A ceiling greater than 100 (that is, 10,000 ft) is regarded as unlimited; ceiling zero means 50 ft or less.

A crude estimate of the ceiling can often be obtained from the type of clouds that are found lowest in the sky, since most clouds have their characteristic altitudes.

Accurate determinations are made during the day by means of *ceiling balloons*, small rubber craft containing a quantity of hydrogen, which causes them to ascend at a rate of 400 ft/min. The time required by the balloons to disappear in the cloud then gives the ceiling. At night a vertical-beam searchlight is directed against the cloud; it forms a bright spot, visible for a considerable distance, upon the base of the cloud. At a distance of about 1,000 ft from the projector an observer measures the angle at which the spot appears, and the ceiling is found by triangulation.

★23.5. General Distribution of Winds. If the earth were smooth and did not rotate, the unequal heating of the various portions of its surface would give rise to a fairly simple pattern of air currents. Hot air would rise at the equator, cold air would move from the poles to the equator, hot air in the upper atmosphere from the equator to the poles. If, therefore, we can disregard local irregularities, such as mountains and bodies of water, the winds on a stationary earth would, near its surface, blow south in the Northern and north in the Southern Hemisphere.

This simple pattern is considerably disturbed by the earth's rotation. Its effect is not easily explained in elementary terms. It is due to a force, called *Coriolis* force, which is brought into being by the circumstance that the earth's surface is not in uniform motion (cf. Sec. 10.5). But a simple consideration shows qualitatively what happens.

Suppose that a mass m moves northward from the equator, starting out along a meridian (circle of longitude) and remaining in frictionless contact with the earth's surface. At the equator its angular momentum is $mR^2\omega$, R being the radius and ω the angular velocity of the earth. As it moves north, it approaches the axis of rotation; hence R becomes smaller. But the principle of conservation of angular momentum requires $mR^2\omega$ to remain unaltered. The consequence is that ω must increase, m must be deflected to the right as the mass moves northward.

This is also what happens to a mass of air. An application of the same principle to masses moving south in the Northern, north in the Southern Hemisphere leads to the following two propositions, valid in both hemispheres:

1. Air traveling away from the equator is deflected eastward.
2. Air traveling toward the equator is deflected westward.

This gives rise to the system of winds represented schematically in Fig. 23.2. Confining our attention to the Northern Hemisphere, we note that air rises in the equatorial calms, moves northeastward in the upper troposphere, and reaches the surface at a latitude of about 30°. From there up to about 60° it

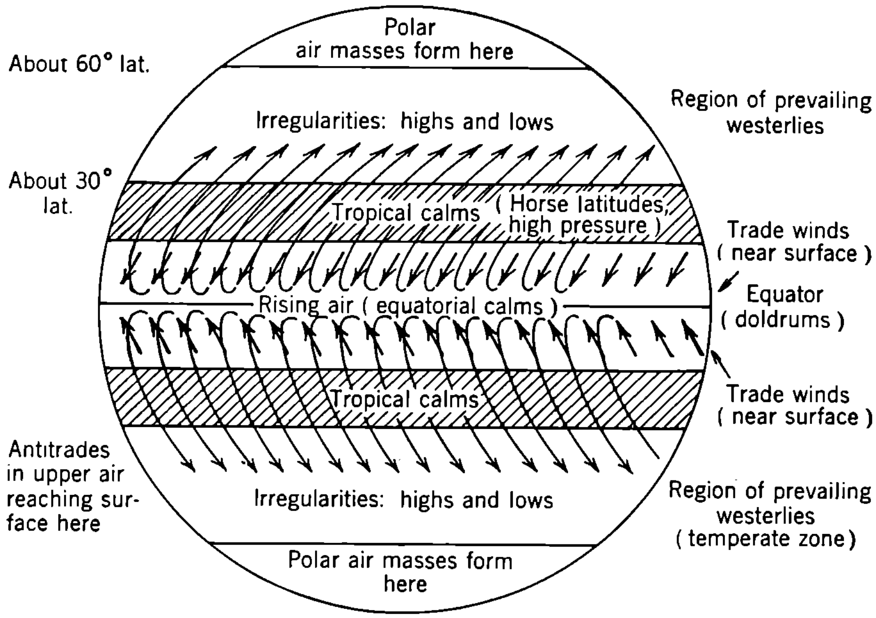


FIG. 23.2. Prevailing winds over the earth's surface.

sweeps over the surface in an easterly direction, giving rise to winds called the "prevailing westerlies" in the temperate zone. Below 30° there lies a belt in which surface motion over the earth is not yet clearly developed, while the upper air currents leave it relatively unaffected. It is known as the belt of the *tropical calm*, or the "horse latitudes." Between it and the equator the winds come predominantly from the northeast and are called *trade winds*.

The main part of North America lies in the region of the prevailing westerlies. Hence our weather comes to us from the west. This region is also marked by the occurrence of interesting irregularities called "highs" and "lows," which will be described in the next section.

***23.6. High- and Low-pressure Areas.** One of the outstanding features on any weather map is the distribution of high- and low-pressure areas. A low, sometimes called a *cyclone*, attracts the air from the surrounding regions of higher pressure. Because of the deflections of air currents studied in the preceding section the low will take on a counterclockwise whirl (cf. Fig. 23.3). Also, since it is the meeting place of frequently incompatible air masses, it represents an area of unsettled conditions. The significance of a falling barometer, indicating the approach of a cyclone, is thus clearly apparent.

For the reasons already stated a high revolves in a clockwise manner in the Northern Hemisphere. Winds disperse from it, and hence its internal condition is seldom unsettled. It contains little water vapor and usually brings good weather.

A remarkable property of highs and lows is that they persist as entities for a considerable time. Some have been traced around the globe, and "weathermen" are accustomed to thinking of them as slowly disappearing objects to which they occasionally give names. These vast air masses travel at a fairly definite rate, lows at about 20 miles/hr in summer and 30 miles/hr in winter. Highs travel somewhat more slowly. Their course is easterly in the United States, as noted; they have a tendency to move along the isotherm (line along which the temperature remains the same) drawn through their center. Observation has shown that

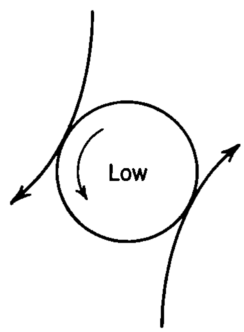


FIG. 23.3. Rotation of a cyclone in the northern hemisphere.

cyclones travel in general along one of several well-marked paths across the continent, and it is chiefly because of this behavior that the ingenious meteorologist can predict the weather for several days.

The regular eastward movement of the air in the temperate zone is disturbed by the passage of a low. Because of its counterclockwise vorticity, a south wind at a given locality usually means the approach from the west of a low, and an east wind indicates that a low is passing to the south; north and west winds follow the passage of a low.

For a high the relations are roughly reversed. Highs are sometimes to be feared in the winter, for they may forbode the coming of a cold polar air mass from the arctic regions. Thus a northwest wind during winter often indicates the approach of a high from western Canada, which will bring a severe cold wave to the central and eastern parts of the United States.

A hurricane, or tropical storm, is a violent cyclone originating mostly in the vicinity of the West Indies. It occurs in the autumn and takes its course toward the Gulf states. If thereafter it travels inland, it becomes extremely destructive and may leave in its wake a devastated countryside. Winds reach velocities of 75 miles/hr and drive high waves into coastal towns. Fortunately, many hurricanes swerve off northeastward into the Atlantic, causing damage only at sea.

***23.7. Local Climatic Conditions.** Local peculiarities of weather are too numerous for consideration here. There are, however, a few facts of interest that illustrate important physical principles. Among them is the formation of hot winds, like the *chinook* on the eastern slopes of the Rockies and the *foehn* of the European Alps. Their production is illustrated in Fig. 23.4.

Suppose that the surface winds have their "normal" course and blow from the west. To cross a mountain they have to ascend. During this process the air mass lowers its pressure, expands more or less adiabatically, and is thus cooled. Measurements show this cooling to take place at a rate of about 5.5°F per 1,000 ft of ascent.[1]

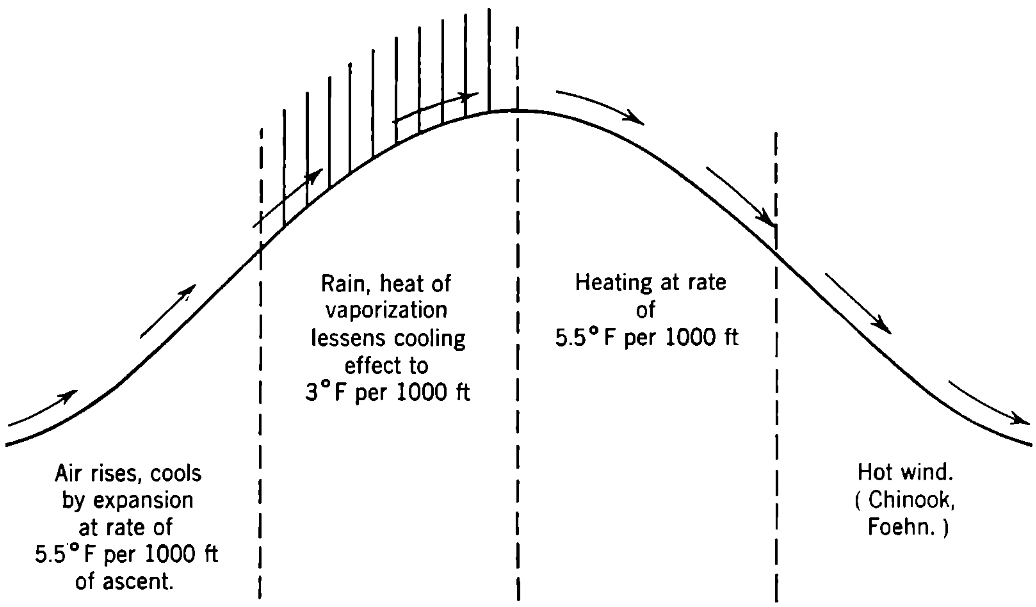


FIG. 23.4. Formation of hot mountain winds.

Now let us assume that the air has a high relative humidity. As soon as it cools below the dew point, precipitation will form and the *latent heat of vaporization* will be given to the air. Thus, in its further ascent, the cooling rate is reduced to a value of about 3°F per 1,000 ft. At the summit the air is relatively dry; on descending it is heated at the original rate, so that in the eastern valley the wind blows abnormally hot.

The same reasoning accounts for the fact that there is usually more precipitation on the windward side of mountains than on the leeward slopes. This effect is very noticeable in the Hawaiian Islands.

***23.8. Polar-front Theory of Cyclones.** The work of two Norwegians, the physicist Vilhelm Bjerknes (1862–) and his son Jacob (1897–) has contributed greatly to a deepened understanding of meteorological phenomena and has revolutionized the art of weather forecasting. Their theory has led to

[1] Notice that the rate of cooling of adiabatically expanding air is *greater* than the lapse rate (3°F per 1,000 ft), which is the rate at which T decreases as a thermometer is carried aloft.

what is now called *air-mass analysis*, the study of large bodies of air in their three-dimensional behavior. We wish to discuss here briefly some of the central ideas of this theory.

It has already been noted that great masses of air take on the characteristics of the region over which they form and that they are objects of considerable permanence. Thus, when two such masses meet, they do not mingle at once but form a definite boundary, a surface of discontinuity called a *front*. The air

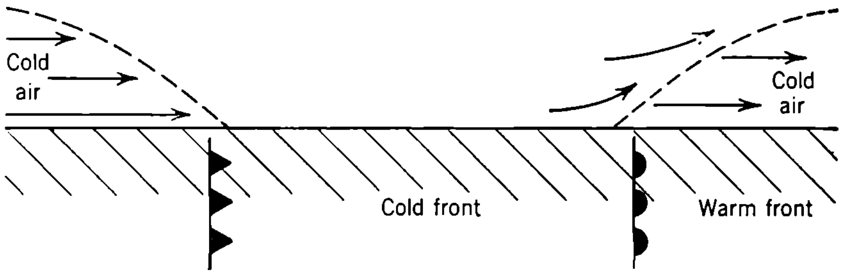


Fig. 23.5. Cold and warm fronts.

masses involved in such encounters may vary from 10 billion to 100 billion tons.

A *cold front* is formed when an advancing cold, and therefore denser, air mass wedges itself under a warm and less dense mass that it overtakes in its course. On modern weather maps a cold front is symbolized by  the teeth extending in the direction in which the front advances. On the other hand, when a warm air mass overtakes a retreating cold mass, it drives it back by flowing upon it, as is shown in Fig. 23.5; the surface of contact then becomes a *warm front* and is represented by .

Along both fronts warm air rises, expands, and tends to shed its water; hence clouds will form, and precipitation may develop along either front. An interest-

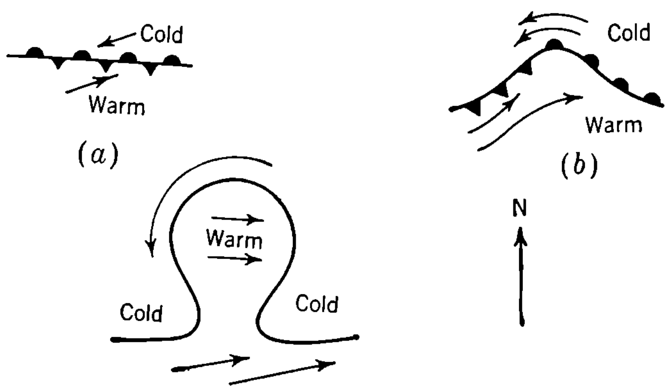
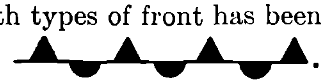


Fig. 23.6. Development of a low.

ing correlation of clouds with types of front has been made. A stationary front is indicated by the symbol .

The front theory gives satisfactory account not only of surface weather but also of conditions aloft and is of considerable benefit to the airplane pilot trained in meteorology.

To explain the formation of cyclones in the temperate zone it is assumed that, first of all, a stationary polar front is formed. That is, a cold air mass, coming from the arctic regions, is blocked in its progress by an opposing warm mass (Fig. 23.6*a*). The arrows indicate the direction of the winds likely to be set up under these conditions. The front does not remain stationary but suffers distortion, the warm air forming a tonguelike projection as shown in Fig. 23.6*b*. Finally (Fig. 23.6*c*) the cold air flows around the projection and severs it, and the trapped warm air becomes the nucleus of a low.

PROBLEMS

★1. Explain why the lapse rate of the atmosphere is smaller than the rate of cooling of ascending dry air.

★2. The temperature at sea level is 60°F. A mass of dry air rises 3,000 ft. What will be its temperature in relation to the temperature prevailing at that height?

★3. Assume that, in Fig. 23.4, condensation starts halfway up the mountain range and continues to the top. If the mountain is 6,000 ft high, what temperature increase has taken place when the air reaches the eastern valley?

4. The air in a room has a temperature of 70°. Its dew point is 45°. What is the relative humidity?

5. What is the absolute humidity of water vapor at 100°C?

★6. An air mass of 2×10^{10} tons has a relative humidity of 75 per cent at 70°F. It comes in contact with a polar mass and cools, losing one-half its moisture. Calculate the energy that is released in the form of latent heat of vaporization (about 600 cal/gm). What would this energy cost at the rate of 5 cents per kilowatt hour? (Condensation provides the energy of hurricanes.)

CHAPTER 24

ELECTRIC CHARGES AND COULOMB'S LAW

24.1. Electrification by Friction. It has been known since the time of the ancient Greeks (600 B.C.) that certain substances when rubbed acquire the ability to attract strongly bits of paper and other light objects. The body is said to have become "electrified" by rubbing. Amber is outstanding in the ease with which it can be electrified, and the word "electricity" and its many derivatives originated from the Greek word for amber, *elektron*. An electrified body is said to possess an electric charge, and the properties of attraction are to be ascribed to this charge. A small pith ball suspended by a light silk thread is a good object to use in testing the electrification of a body. When a hard-rubber rod charged by rubbing with felt or flannel is brought near a suspended pith ball, it attracts the ball strongly. If now the rod is touched to the pith ball, some of the charge on the rod is transferred to the pith ball. It is found that the ball is strongly repelled from the rod. If then a glass rod that has been lightly rubbed with fur or silk is brought near the pith ball, the force is attractive. The hard-rubber rod, however, still repels the ball. Thus it is apparent that the charge on the rubber rod is different in some way from the charge on the glass rod. These two kinds of charges were originally called "resinous" and "vitreous" but are now designated as "negative" and "positive" charges, respectively. Simple experiments with a pith ball can easily show that like charges repel and unlike charges attract one another. The study of electricity and the companion phenomena of magnetism is a study of the interaction of one charge upon another, the charges being either at rest or in motion. The science of charges at rest is called "electrostatics," that of charges in motion, "electrodynamics" or "electromagnetism."

24.2. The Conservation of Charge. Charge is a fundamental property of all matter: the electrons that are contained in all bodies are negatively charged, and the atomic nuclei are positively charged. Normally, equal amounts of positive and negative charge are present, and the net charge is zero. When a glass rod is rubbed, some of the electrons are scraped off the surface of the glass and the rod possesses a net positive charge as a result. The electrons do not disappear, however; they remain on the fur with which the rod was rubbed. This can be easily seen by testing the fur with a negatively charged pith ball. Indeed, experience teaches us that *charge cannot be created or destroyed:* positive and negative charges can only be separated or recombined. This statement may be called the *law of conservation of charge.* It is the first fundamental law of nature that we encounter in the study of electricity. The law ranks in importance with the other conservation laws of physics that have been studied—those of energy and momentum.

It is not possible to give a short definition of charge. Charge can be defined only by a description of the properties and behavior of charges, and this we shall attempt to do in succeeding sections.

24.3. Insulators and Conductors. The excess number of electrons that reside on a charged hard-rubber rod will stay there for a long time. There is very little tendency for them to move to another place even though the individual electrons are repelled by their neighbors. This ability of the rubber rod to hold the electrons in place is shared by a number of other common substances such as glass, silk, dry wood, and most synthetic plastics. Such substances are said to be "nonconductors" of electricity or to be "insulators." In other materials the electrons are relatively free to move from one place in the material to another, and therefore the material is a good conductor. The best conductors are the metals silver and copper. In Chap. 22 the high values of the conductivity of metals for heat were discussed and were said to arise from the transfer of energy by the free electrons within the metal. This same freedom of the electrons allows charge to be easily distributed or conducted in a metal.

Water and solutions of polar salts are also conductors of electricity, although poor ones. The charge in these cases is carried, not by free electrons, but by the charged ions into which the water or the salt dissociates. In electrostatic phenomena, even the poor conductivity of moist objects is sufficiently large to allow charge to leak away before the effects of the charge can be detected. Thus, if the attempt is made to charge a metal rod by rubbing, it will be found that the charge leaks away over the moist skin of the hand as fast as it is produced. If a metal rod is furnished with an insulating handle, however, it can be charged by rubbing exactly as can a glass or rubber rod.

In a metal in which the net charge is zero the electrons and positive ions are uniformly distributed throughout. If some extra electrons are added to the metal, there will be forces of mutual repulsion between the electrons and they will move as far apart as possible. Thus all the excess charge on a conductor resides on its surface, and none is contained within the body of the metal. Insulating materials, on the other hand, may be charged on the surface (by rubbing, for example) or may have charges distributed throughout their volume.

24.4. Electroscopes. Although a suspended pith ball is useful for testing charges, more convenient and precise instruments can be devised easily. All electroscopes employ the motion of a light object acted upon by forces of attraction or repulsion between charges. A common form of electroscope is the leaf electroscope shown in Fig. 24.1. A light metal leaf L of gold or aluminum is attached at one end to a metal strip S that is insulated from the metal case of the instrument. When the metal strip and leaf are charged by touching a charged body to the terminal T at the top of the instrument, the charge distributes itself over the surfaces of the strip and the metal leaf. The mutual repulsion of the charges on the leaf and the strip force the leaf to stand out at an angle from the vertical. The position of the leaf can be measured by a scale on the window of the instrument or by a measuring microscope. It is evident that the sensitivity of a leaf electrometer would be increased by the use of as light and thin a leaf as possible. Such metal leaves are fragile, of course, and modern electrometers employ, for the moving elements, quartz fibers that have been coated with gold by evaporation. Electroscopes also often have a pair of leaves without a fixed metal strip.

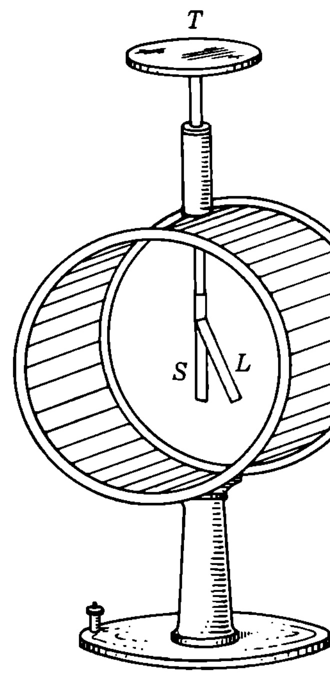


Fig. 24.1. Leaf electroscope.

24.5. Coulomb's Law.[1] A quantitative measurement of the force of repulsion between two charges could be made if the restoring force on the moving leaf of an electroscope were known. A more suitable instrument, however, is a torsion balance such as Cavendish used to measure the gravitational attraction between two masses (Chap. 11). The first investigation of this kind was made by Coulomb (1736–1806), who found that the force between two charges is inversely proportional to the square of the distance between them. Later investigations showed that the force is also proportional to the product of the amounts of the charges. Thus the force may be written as

$$F = k\frac{qq'}{r^2} \tag{24.1}$$

where q and q' are the amounts of the two charges, r is the distance between them, and k is the constant of proportionality. The constant k is positive, since if the charges are alike, *i.e.*, have the same sign, the

[1] Review Sec. 11.3.

force is repulsive. A repulsive force is taken to be positive since it tends to increase r. The law expressed by Eq. (24.1) is known as *Coulomb's law*. Though not written as a vector equation, it is understood that the force is always directed along the line joining the charges; it is a "central" force.

Coulomb's law at once recalls to mind the law of universal gravitation,

$$F = -G\frac{mm'}{r^2} \tag{11.1}$$

which is identical in form with Coulomb's law with the exception of the sign. This is as it should be since two masses attract one another but two like charges repel. The identity of the form of the laws allows us to use immediately a result derived for the gravitational case. It was proved (Sec. 11.3) that the gravitational force exerted by a spherical shell of total mass m is exactly the same as if all the mass were concentrated at the center of the spherical shell. We have already seen that the charge resides on the surface of a conductor, and hence the force between two *uniformly* charged spheres is given by Eq. (24.1), if r represents the distance between the centers of the charged spheres. It is important to note the qualification "uniformly"; an actual charged spherical conductor may under certain conditions not be charged in this manner. Moreover, since the gravitational force is zero inside a spherical shell of matter, the force on a charge inside a spherical shell of charge is also zero.

24.6. The Units of Charge. To give Eq. (24.1) complete quantitative significance it is necessary to define a unit in terms of which charge is to be measured and to assign a value for the constant k. A natural unit of charge does exist, *viz.*, the charge on one electron. Unfortunately a charge of this magnitude is too small for convenient use as a unit charge, and, moreover, the existence of such a natural unit has been substantiated only within the present century. A unit of charge can, however, be defined from Eq. (24.1). The units for the force F and the distance r have already been chosen as one dyne and one centimeter, respectively, in the cgs system. If the constant k is arbitrarily chosen to be unity, then the unit of charge defined by Eq. (24.1) is called the "cgs electrostatic unit (esu) of charge." Thus *one electrostatic unit of charge is that charge which repels an equal charge with a force of one dyne when the charges are separated by one centimeter.* This charge is very large compared with that of one electron; in fact the charge e on an electron is -4.80×10^{-10} esu. The electrostatic unit of charge is sometimes called the *statcoulomb*.

If k is arbitrarily assumed to have no dimensions, then Eq. (24.1) can be used to find the dimensions $[Q]$ of charge in the electrostatic system of units,

$$[\text{dyne}] = \frac{[Q^2]}{[\text{cm}^2]}$$

or
$$[Q] = [\text{cm}][\text{dyne}^{1/2}] = [M^{1/2}L^{3/2}T^{-1}] \tag{24.2}$$

The dimensions of charge in this system of units are thus very complicated ones.

For many practical applications the electrostatic unit of charge, although much larger than the charge on the electron, is still too small. Consequently a practical unit of charge must also be defined. The practical unit of charge is called the *coulomb* and is equal to approximately 3×10^9 statcoulombs. More exactly,

$$1 \text{ coulomb} = 2.996 \times 10^9 \text{ statcoulombs} \tag{24.3}$$

Strangely enough the ratio of these two units is numerically equal to one-tenth the velocity of light, as will be explained later. For many purposes a fraction of this unit is convenient, and the unit microcoulomb, or 10^{-6} coulomb is in common use. The charge of the electron in the practical system is 1.60×10^{-19} coulomb.

The practical system of electrical units is an extension of the mks system that was introduced in the study of mechanics (see Secs. 1.5 and 6.3). When the quantities F and r of Eq. (24.1) are expressed in mks units and q and q' are in coulombs, the value of the constant k is no longer unity but is

$$k = 8.99 \times 10^9$$

For convenience in later work, it is customary to define a new constant ϵ_0 by

$$\epsilon_0 = \frac{1}{4\pi k} \qquad k = \frac{1}{4\pi\epsilon_0} \tag{24.4}$$

Coulomb's law in practical units becomes

$$F = \frac{1}{4\pi\epsilon_0}\frac{qq'}{r^2} \tag{24.5}$$

where
$$\epsilon_0 = 8.85 \times 10^{-12} \frac{(\text{coulombs})^2}{(\text{joule})(\text{m})} \tag{24.6}$$

The factor 4π is explicitly included in Eqs. (24.4) and (24.5), so that a similar factor will disappear from equations to be derived later. The system of units so defined is called the *rationalized* mks system. If Eq. (24.4) were written without the factor 4π, an unrationalized system would result. Only the rationalized system will be considered in this book. The quantity ϵ_0 is usually given special units, but a discussion of this question must be deferred until later.

Summary. Coulomb's law reads

$$F = \frac{qq'}{r^2} \qquad \text{esu}$$

where q and q' are measured in esu and r in cm. It reads

$$F = \frac{1}{4\pi\epsilon_0}\frac{qq'}{r^2} \qquad \text{mks}$$

where q and q' are in coulombs, r is in meters, and ϵ_0 has the value given by Eq. (24.6).

Worked Example. Compare the electric force of repulsion and the gravitational force of attraction of two electrons a distance of 10^{-8} cm apart. Let us use esu, and hence $k = 1$.

The electric force is

$$\begin{aligned} F_e &= \frac{(4.80 \times 10^{-10})^2}{(10^{-8})^2} \text{ dyne} \\ &= 2.3 \times 10^{-3} \text{ dyne} \end{aligned}$$

The gravitational constant G is

$$G = 6.7 \times 10^{-8} \text{ cgs units}$$

and the mass of the electron is 9.1×10^{-28} gm. The gravitational force is, therefore,

$$\begin{aligned} F_g &= -6.7 \times 10^{-8} \frac{(9.1 \times 10^{-28})^2}{(10^{-8})^2} \text{ dyne} \\ &= -5.5 \times 10^{-46} \text{ dyne} \end{aligned}$$

The electric force thus far exceeds the gravitational attraction.

24.7. The Electrical Structure of Matter. Since all substances show electrical properties, it is reasonable to suppose that electrons are one of the constituents of their atoms. Matter must contain also some positively charged parts so that the net charge may be zero. We note that the mass of the hydrogen atom as derived in Sec. 20.7 is nearly 2,000 times the mass of the electron. Therefore, either an atom of hydrogen must contain many electrons, or else the mass must be associated chiefly with the positive charges. The second possibility happens to be the correct one. One electron can be removed from a hydrogen atom, but no more. Hydrogen is composed, then, of one light electron and a heavy positively charged part. Heavier atoms contain proportionally more electrons.

How, then, are the electrons and the heavy positively charged parts arranged within the atomic volume? It is evident that there are two extreme situations possible. First, the mass of the atom and therefore the positive charge might be more or less uniformly distributed within the atom, with the light electrons scattered throughout this mass; or, second, the positive massive part of the atom might be concentrated at the center and the electrons might form an atmosphere around it. To decide between these possibilities, it is necessary to appeal to experiment. The decisive experiment was made in 1911 by Rutherford (1871–1937), who measured the scattering of α-particles by metal foils.

Gamow[1] has compared Rutherford's experiment with the procedure of an (imaginary) South American customs officer who had to determine whether any arms were being smuggled into his country in bales of cotton. There were a number of bales and no facilities for opening them. The customs officer therefore fired his revolver into the bales, arguing that, if only cotton were present, the bullets would stick in the bales or pass nearly straight through. If there were any heavy metal pieces, however, the bullets would ricochet and emerge with large changes in direction. Rutherford's bullets were α-particles, the fast-moving helium atoms with two electrons removed from each, that are emitted in the decay of certain radioactive substances. Many particles were observed to be scattered at large angles. It must therefore be concluded that the mass and positive charges of an atom are concentrated in a small region. This region is called the "nucleus" of the atom, and the negative electrons move around the nucleus.

By a quantitative analysis of his observations Rutherford was also able to show that the force between the α-particles and the nucleus was just the force between the charges given by Coulomb's law. In addition he could determine the charge on the nucleus and consequently the number of satellite electrons. This nuclear model of the atom has already been employed in Sec. 11.9, where the motion of the electrons was discussed.

Many later experiments have abundantly confirmed the nuclear model of the atom and have furnished information about the arrangement of the electrons surrounding the nucleus. A shell structure is found to exist, each shell containing a definite number of electrons in rapid motion. The shells of charge are stable units and are not easily disturbed by external forces acting on the atom. The outermost shell is not so stable as the others, and nearly all the interatomic or chemical forces between atoms are the result of the behavior of the electrons in this shell. To understand the electronic behavior, it was found necessary to introduce quantum theories (see Chap. 49).

[1] G. Gamow, "The Birth and Death of the Sun," New York, 1945.

24.8. Metals. When the atoms of an element are grouped together to form a crystal, the outer electrons in each atom move under the action of forces arising not only from the positively charged atomic nucleus and from the inner electrons but also from the neighboring nuclei and the electrons in the other atoms. Thus the structure of the outer shell of electrons is considerably distorted. In fact, in many elements and particularly in those elements with one or two electrons in the outer shell, the force from the neighboring atoms is so strong that it can no longer be said that a given outer electron "belongs" to any particular nucleus. Rather the outer electrons are shared by all the atoms of the crystal. In this case the crystal is said to be a metal. The outer electrons can move with relative ease from one part of the metal to another. If external charges are brought near to the metal, the electrons move under the influence of the forces they exert: the metal is a good conductor of electricity. Since the electrons have kinetic energy, this energy is also transported as the electrons move and the metal is also a good conductor of heat. Typical metals that have one "free" electron per atom are the alkali metals, sodium, potassium, and so forth, and the metals copper, silver, and gold. Metals that are poorer conductors of heat and electricity, such as iron, zinc, and mercury, have two outer electrons in each atom. These electrons, however, do not become completely free when a crystal is formed.

The motion of the electrons in a metal can often be described quite exactly by treating the electrons as a perfect gas. The electrons of this gas "collide" with the positive ions as they move through the crystal, and the situation is similar to that in an actual gas. In crystals that are not metals the motion of the electrons is restricted. It is evident, however, from this description that it is impossible to make a sharp distinction between a metal and a nonmetal. Since free electrons behave like a gas, the kinetic theory predicts a value for the rms velocity $\sqrt{\overline{v^2}}$. In Chap. 20 it was shown that

$$\overline{v^2} = \frac{3kT}{m}$$

Since the mass m of an electron is approximately 9×10^{-31} kg, at $T =$ 300°K, $\sqrt{\overline{v^2}}$ has the value of 1.17×10^5 m/sec.

24.9. Induced Charges. If a positively charged rod is brought near an uncharged conductor, forces are exerted on the charged particles of which the body is composed; and the electrons that are free to move in the conductor concentrate themselves near the charged rod. The opposite side of the conductor then has a deficiency of electrons and is positively charged. This separation of the positive and negative charges is called *electrostatic induction.* Since the negative charges on the conductor are closer to the charged rod than the positive ones, there is a force of *attraction* between the two bodies, even though one body has a zero net charge. This is the explanation of the initial attraction of the pith ball to the charged rod described in Sec. 24.1. If the rod were charged negatively instead of positively, the force would again be attractive, since the electrons in the conductor would be forced away from the rod and would leave a positive charge behind them.

The process of induction can be used to give a conductor an excess of charge. Suppose that we have two spherical conductors, A and B in Fig. 24.2*a*, in contact and that a third body C charged positively is

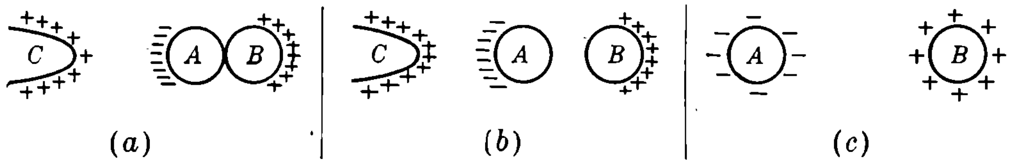


FIG. 24.2. The charging of conductors by induction.

brought near them. By induction the left side of body A acquires a negative charge, and an equal and opposite charge appears on B. The spheres A and B can now be separated a little as in Fig. 24.2*b*; the charges remain in nearly the same distribution. If the charged body C is now removed and the spheres separated, as in Fig. 24.2*c*, the mutual repulsion of the charges on A and B force a uniform distribution of charge to occur over each sphere.

The method of inductive charging is a convenient one to use for charging an electroscope. If the central system of the electroscope is initially uncharged, the approach of a charged body causes the leaves to diverge. The charge distribution on the electroscope leaves and case is indicated schematically in Fig. 24.3*a*. If the case of the electroscope is then connected by a movable contact to the terminal of the central system, as in Fig. 24.3*b*, negative charge flows from the case to the electroscope leaves and they collapse at once. The contact between the electroscope leaves and the case can be made usually by touching the electroscope terminal with the finger. Unless special precautions are taken, charge can flow from the electroscope case resting on a table across the floor and through the body of the experimenter to the leaves. If the case were well insulated, the connection would have to be made with a conductor. The charge distribution in Fig. 24.3*b* is analogous to that shown in Fig. 24.2*a*, and the leaf system and the electroscope case are labeled A and B correspondingly. If the contact between A and B is broken but the charged rod is still present, the charge distribution is not altered and the leaves are still not charged. Figures 24.3*c* and 24.2*b* thus correspond to each other. The positively charged rod is then removed. The forces on the charges on the leaf system are changed, and the charges move and redistribute themselves as indicated in Fig. 24.3*d* corresponding to Fig. 24.2*c*. The electroscope leaves are charged negatively, a charge opposite in sign to that on the rod being produced.

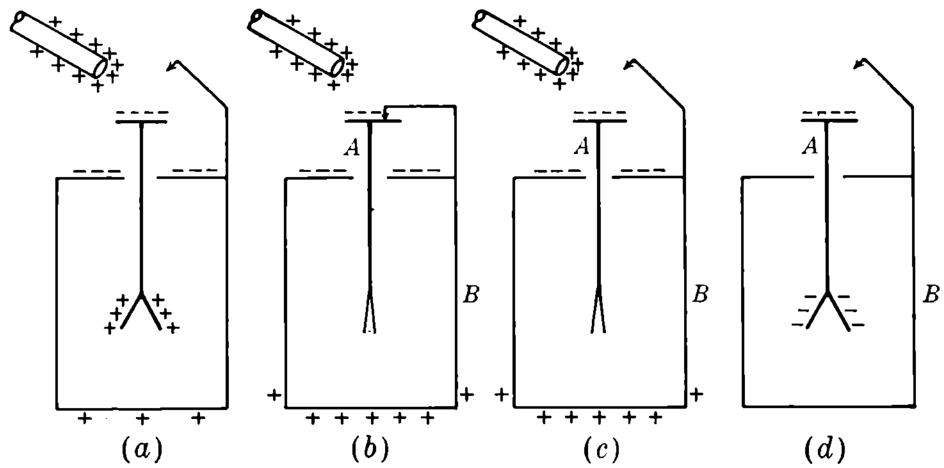


FIG. 24.3. Electroscope charging by induction.

PROBLEMS

1. Two positive charges of magnitudes 5 and 10 statcoulombs are situated 50 cm apart. At what point between them does a force on a unit positive charge vanish? At what point is the force on a negative charge of 2 statcoulombs equal to zero?

2. A negative charge of 1 esu at the origin of a system of rectangular coordinates is acted upon by 1 esu of positive charge at $x = 0$, $y = 10$ cm and a positive charge of 2 esu at $x = 10$ cm, $y = 0$. What are the direction and magnitude of the force on the negative charge?

3. Compare the electric and gravitational forces between electrons as in the worked example of Sec. 24.6, using practical mks units.

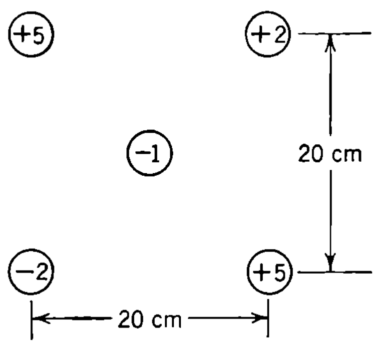


FIG. 24.4. Arrangement of charges in Prob. 6.

4. What is the number of electronic charges in one electrostatic unit of charge? In 1 coulomb?

5. What is the total charge of one mole of a gas composed of free electrons? The mass of the electron is 9.1×10^{-28} gm; the mass of a hydrogen atom is 1,840 times as large.

6. Four charges are placed at corners of a square with sides 20 cm long as shown in Fig. 24.4. The magnitudes of the charges in coulombs are indicated. What is the force (a vector!) on a negative charge of 1 coulomb at the center of the square?

7. Two pith balls, each of mass 2 mg, are suspended by very light threads, 5 cm long, from a common point. The balls are charged so that the threads form angles of 30° with the vertical. What is the charge on each ball?

8. If the charge on one of the pith balls in Prob. 7 is twice that on the other ball, what are the angles of the threads from the vertical?

★9. Show that the force on a charge q' at a distance r from a long wire of negligible diameter charged with p units per unit length is $2kpq/r$.

CHAPTER 25

THE ELECTRIC FIELD AND POTENTIAL

25.1. The Electric Field. Since the law of force between two point charges is known from Eq. (24.1), it is possible to find the force exerted by one charged body on another by taking a vector sum of the forces between small elements of charge on each body. Such a procedure would be a complicated one, even for simple cases, since the distribution of charge on one body is affected by the presence of the other body. It is useful, therefore, to describe the state by an *electric field*, just as it was useful to speak of a gravitational field in the neighborhood of a massive object. If a very small charge q' is placed near the charged body, it experiences a force proportional to q' and this force is the sum of the forces exerted by all elements in the distribution of charge acting on q'. For every position of the charge q' there is a corresponding value of the force F. *The strength of the electric field is then defined as the force per unit charge.* The strength of the field or, as it is often abbreviated, the field $\mathbf{E}$ is a vector quantity and may be expressed by the equation

$$\mathbf{E} = \mathbf{F}/q' \tag{25.1}$$

Thus to every point near a distribution of charges there is assigned a vector $\mathbf{E}$, and this collection of vectors forms a description of the electrified state of the body itself and is called the electric field.

The smallness of the charge q' is necessary to ensure that the presence of q' does not alter the distribution of charge by the process of induction. To be most precise, the electric field is best defined as a derivative

$$\mathbf{E} = d\mathbf{F}/dq' \tag{25.2}$$

In the simplest case of a single point charge of magnitude q the field strength produced by it is

$$E = kq/r^2 \tag{25.3}$$

and is directed at every point radially away from the point charge. If many point charges are present, the field is the vector sum of the fields of the individual charges. If charges $q_1, q_2, \ldots, q_n$ are located at points $1, 2, \ldots, n$ as indicated in Fig. 25.1, then the x component of the resultant field at the origin O is

$$E_x = -k\sum_{i=1}^{n}\frac{q_i}{r_i^2}\cos\theta_i \tag{25.4}$$

where the angle θ_i is shown in the figure.

It is evident from Eq. (25.1) that the unit of electric-field strength in the cgs system is one dyne per statcoulomb, or in the practical system one newton per coulomb. Another, more customary unit will be introduced later.

25.2. Lines of Force. Since the electric field is a vector quantity, there is associated a direction with every point near a charge distribution. A small test charge q' tends to move in the field always in the direction of the field. The path in the field that is everywhere tangent to the field is called a *line of force*. Thus, if AB in Fig. 25.2 is a line of force, the directions of **E** are tangents to the curve AB as indicated. For a single positive point charge the field is radially outward, and the lines of force are consequently radial lines.

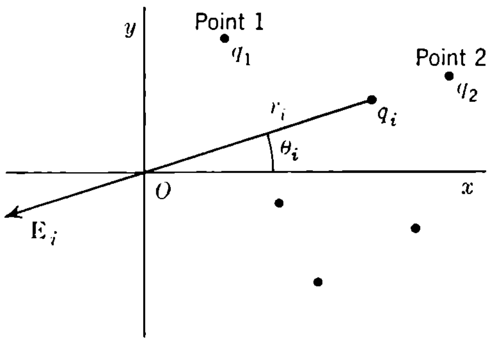


FIG. 25.1. The field of a number of point charges is the sum of the vectors $\mathbf{E}_i$.

Lines of force immediately give the *direction* of the electric field. With an additional convention, lines of force can also be used to represent the *strength* of the field. If the number of lines of force crossing a unit area perpendicular to the lines is chosen to be proportional to the field strength, then the density of lines is a measure of the strength of the field. It is customary to take the number of lines per unit area numerically equal to the field. When this is done, the lines are called *unit lines of force.*

The number of unit lines of force radiating from a unit charge depends on whether the cgs or the practical system of electrical units is used. The number of unit lines crossing a sphere of radius r with a charge of q units at its center is the area of the sphere times the field strength, or

$$N = 4\pi r^2 kq/r^2 = 4\pi kq \tag{25.5}$$

In the cgs electrostatic system, $k = 1$ and $N = 4\pi q$. In the practical system, $k = 1/4\pi\epsilon_0$, and $N = q/\epsilon_0$. The number of lines of force radiating from one coulomb of positive charge is therefore 1.13×10^{11}. It should be kept in mind that the unit area in the practical system is the square *meter*.

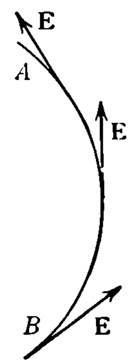


FIG. 25.2. The direction of **E** is tangent to a line of force.

The electric field at any point must have a definite direction as long as the field is not zero. Consequently lines of force cannot cross one another and must be continuous. A line of force must start on a positive charge and terminate on a negative one. Lines of force have many properties in common with the streamlines that are used to represent the flow of a liquid; indeed; the analogy between these phenomena is extremely close.

25.3. The Field of a Dipole. Suppose that two charges $+q$ and $-q$ are separated by a small distance l; then the field strength can be found by the application of Eq. (25.4), where the sum now consists of two terms. Such a configuration of charges is of special interest and is called a *dipole*. In Fig. 25.3 the field $\mathbf{E}_1$ on the axis of the dipole at the point A is in the x direction with a positive term from the charge $+q$ and a negative one from $-q$. If the origin is taken at the center of the dipole, then the magnitude of $\mathbf{E}_1$ is

$$E_1 = \frac{kq}{[x - (l/2)]^2} - \frac{kq}{[x + (l/2)]^2}$$

or

$$E_1 = (kq \times 2xl)/[x^2 - (l^2/4)]^2$$

Since the distance l is small compared with x, the term $l^2/4$ can be omitted from the denominator and

$$E_1 = 2kql/x^3 \qquad (25.6)$$

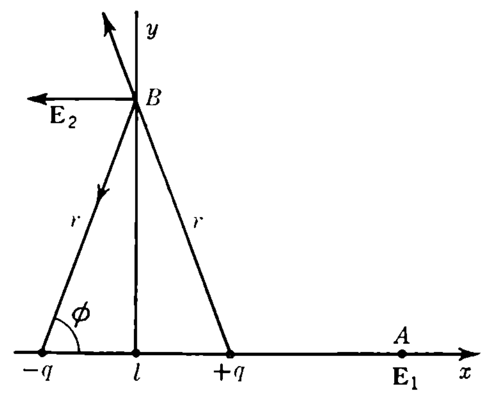


FIG. 25.3. The electric field of a dipole.

The quantity ql is defined as the *strength of the dipole*, or the *dipole moment.* If the dipole moment is denoted by p,

$$E_1 = 2kp/x^3 \qquad (25.7)$$

For a point B on the perpendicular bisector of the dipole axis the field $\mathbf{E}_2$ is the resultant of two fields, as indicated in Fig. 25.3. The components in the y direction evidently are equal and opposite, and the components in the x direction add.

From Eq. (25.4), $E_2 = -2k\frac{q}{r^2}\cos\phi$

But $\cos\phi = l/2r$

and $r = [y^2 + (l^2/4)]^{1/2}$

Hence $E_2 = -kql/[y^2 + (l^2/4)]^{3/2}$

Since $l \ll y$, the term $l^2/4$ can be neglected, and

$$E_2 = -kql/y^3 = -kp/y^3 \qquad (25.8)$$

The field is therefore again directly proportional to the strength of the dipole and inversely proportional to the cube of the distance.

The field at other points in the neighborhood of the dipole can be found in a similar manner. The algebraic manipulations become somewhat complicated, and the explicit expressions will not be derived. The field can, however, be well represented by a picture of the lines of force as shown in Fig. 25.4.

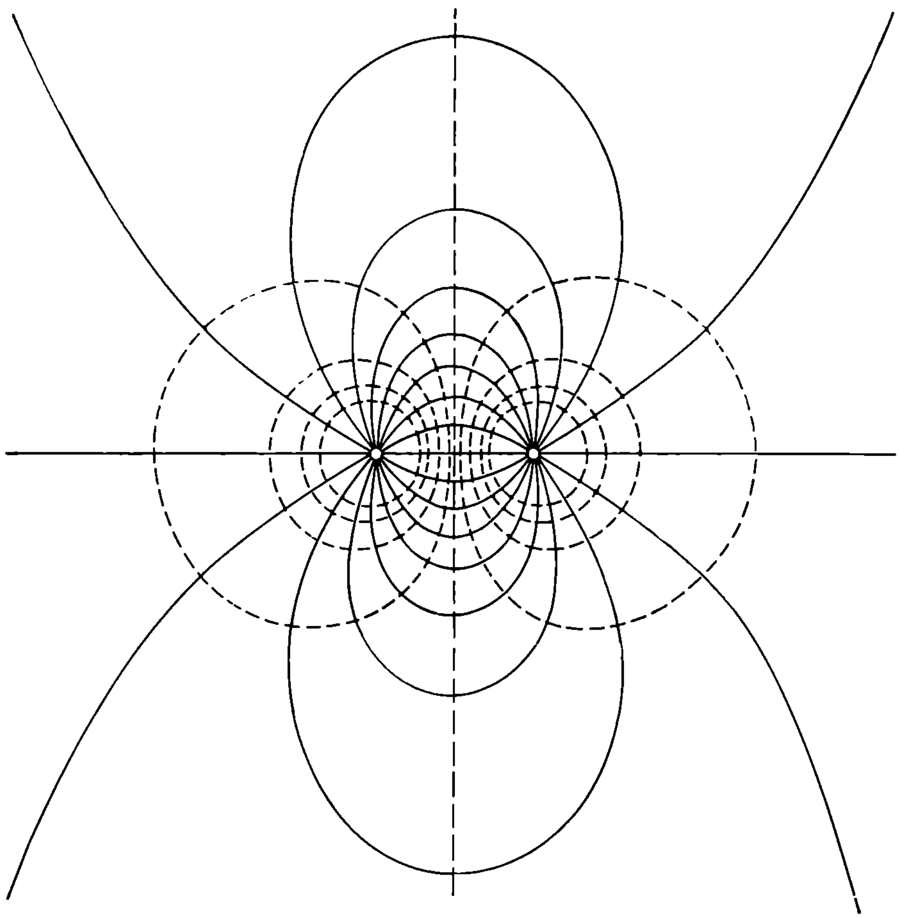


FIG. 25.4. The lines of force (solid lines) about an electric dipole.

25.4. Gauss's Law.[1] Since lines of force can begin or end only on charges, it follows that the total number of lines of force passing through any surface is proportional to the charge q within the surface. This is certainly true for a point charge when the surface is chosen as a sphere about the charge, as shown in Eq. (25.5). If the surface is not a sphere, the field is no longer constant over the surface or the lines of force perpendicular to the surface but the number of lines crossing the surface must remain the same. This property of lines of force is known as Gauss's law. To obtain a mathematical expression for this law, let us consider a small element of area dA, and let the lines of force make an

[1] K. F. Gauss (1777–1855).

angle θ with the normal to the surface, as indicated in Fig. 25.5. The *normal flux through the surface dA* is defined as $E \cos \theta \, dA$. It should be noted that for a line of force passing out of the surface, as shown in the figure, $\cos \theta$ has a positive value; for a line passing into the surface, $\cos \theta$ is negative. The total normal flux through the closed surface is obtained by integration over all elements dA. Gauss's law in mathematical form is found to be

$$\int_{\text{surface}} E \cos \theta \, dA = 4\pi k q \qquad (25.9)$$

where q is the total charge within the surface. The surface over which the integral in Eq. (25.9) is taken is called a *Gaussian surface.*

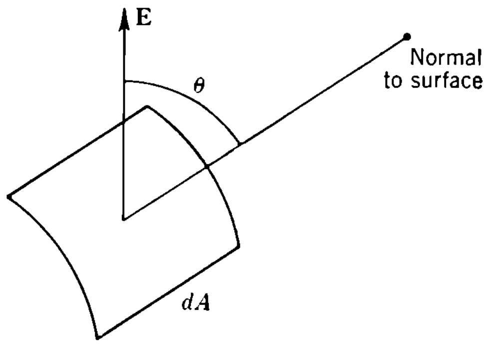


FIG. 25.5. The normal flux through the surface dA is $E \cos \theta \, dA$.

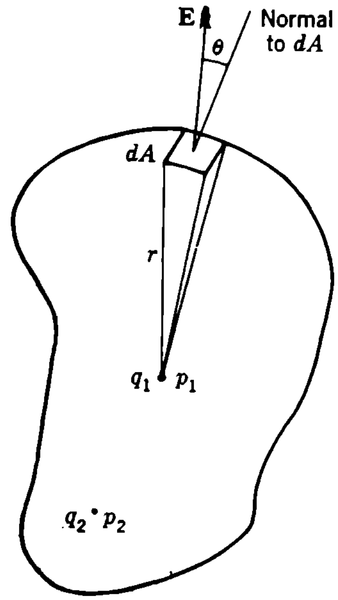


FIG. 25.6. Demonstration of Gauss's law.

Equation (25.9) can be derived in the following manner: Consider a point charge q_1 at the point P_1 within an arbitrary surface, as shown in Fig. 25.6. The normal flux across the element of area dA is

$$E_1 \cos \theta \, dA = \frac{kq_1}{r^2} \cos \theta \, dA$$

The area $\cos \theta \, dA$ is, however, the projection of dA on a sphere of radius r centered at P_1. Hence $\cos \theta \, dA/r^2$ is the element of solid angle $d\Omega$ about the point P_1 (cf. Sec. 1.4).

Hence $$\int_{\text{surface}} E_1 \cos \theta \, dA = \int_{\text{surface}} kq_1 d\Omega = 4\pi k q_1$$

since the integral of $d\Omega$ over the enclosing surface is 4π.

If another charge q_2 is located at P_2, then also

$$\int_{\text{surface}} E_2 \cos \theta \, dA = 4\pi k q_2$$

If both charges are present at the same time, the total normal flux through dA is $(E_1 \cos \theta_1 + E_2 \cos \theta_2)dA$ and the surface integral of the total normal flux is therefore $4\pi k(q_1 + q_2)$. Thus we can immediately generalize the result for a point charge, and for any distribution of charge within the surface we obtain Eq. (25.9).

As a simple example consider a spherical surface drawn around the dipole of Fig. 25.4. The net charge within the sphere is zero, and it is evident from the symmetry of the figure that equal numbers of lines of force enter the surface and leave it.

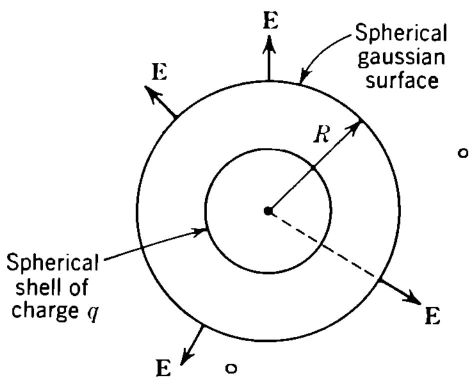


FIG. 25.7. Application of Gauss's law to a spherical shell of charge.

Gauss's law is useful in many other cases where the field is symmetrical. It is possible to show directly, for example, that the field from a uniformly charged spherical shell is the same outside the shell as though the charge were concentrated at the center of the shell (cf. Secs. 11.3 and 24.5). A very simple proof of this fact can be obtained from Gauss's law. Consider a spherical surface of radius R drawn outside the shell of charge, as in Fig. 25.7. From symmetry, **E** is everywhere perpendicular to the surface, and θ is zero. Equation (25.9) can then be written

$$E \int dA = E \times 4\pi R^2 = 4\pi k q$$

or

$$E = kq/R^2$$

as was obtained before.

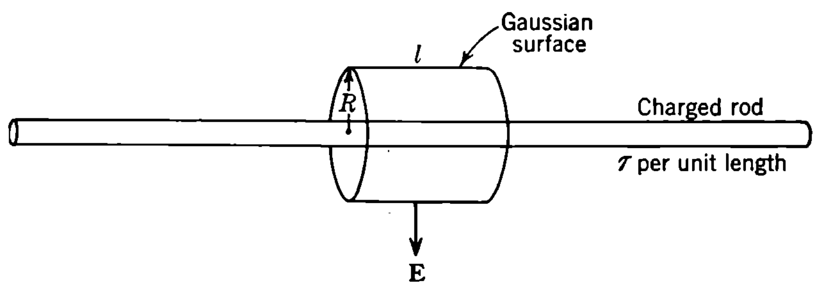


FIG. 25.8. The Gaussian surface about a long charged rod.

As another example, consider a long cylindrical conductor charged to an amount τ per unit length. Except near the ends of the rod, the field is everywhere perpendicular to the axis of the rod and the same all around the rod (see Fig. 25.8). Thus if the Gaussian surface is taken to be a small cylindrical box of length l and radius R coaxial with the rod, there is no flux through the flat end surfaces and $\cos\theta$ is unity over the curved surface. From Eq. (25.9), therefore,

$$E \int dA = E \times 2\pi Rl = 4\pi k\tau l$$

and

$$E = k \times 2\tau/R \tag{25.10}$$

It should be remembered that in the practical system of units, τ is measured in coulombs per meter, R in meters, and k has the value $1/4\pi\epsilon_0$. In the electrostatic system, k equals 1, and τ is measured in statcoulombs per centimeter.

25.5. The Field near the Surface of a Conductor. It will be recalled that in a conductor of electricity the electrons are free to flow under the action of the electric forces acting upon them. The electric field acting on each electron is, of course, the sum of the fields set up by all other charges within the conductor plus the field due to any external charges. The charges will therefore move under the action of these forces until they reach a position of equilibrium. In general, then, the charge is not distributed uniformly over the surface of the conductor, but the charge of an element of surface, the *surface charge density* on the body, varies from place to place. The equilibrium distribution of the charges is perhaps a complicated one, but one property is immediately obvious. *The electric field at the surface of the conductor must be everywhere perpendicular to the surface.* This condition must obtain; for if it were not true, there would be a component of the field along the surface and a force acting to make the electrons move; equilibrium would not exist. Furthermore within the conductor the electric field must be zero under static conditions.

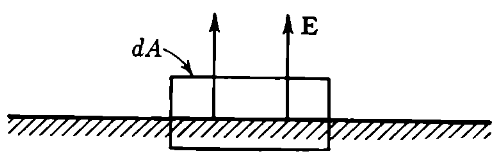


FIG. 25.9. The field near the surface of a conductor.

By means of Gauss's law we can easily derive the relation between the charge density and the field normal to the conductor. For the Gaussian surface take the surface of a pillbox of areas dA for the flat sides and of very small height, as shown in Fig. 25.9. The normal flux of the field will be $E\,dA$ through the top surface and zero over the other surfaces. Hence Gauss's law requires that

$$E\,dA = 4\pi k\sigma\,dA$$

where σ is the charge density. The area of the box cancels and the field is

$$\underline{\overline{E = 4\pi k\sigma}} \tag{25.11}$$

For the electrostatic system of units,

$$k = 1 \qquad \text{and} \qquad E = 4\pi\sigma \text{ esu}$$

For the practical system,

$$k = \frac{1}{4\pi\epsilon_0} \qquad \text{and} \qquad E = \frac{\sigma}{\epsilon_0} \text{ practical units}$$

The relations just proved hold, of course, for a curved surface of a conductor as well as for a flat one, provided that we mean by E the field in a region so close to the conductor that the surface is practically flat, or more precisely that the electric flux through the sides of the pillbox can be neglected.

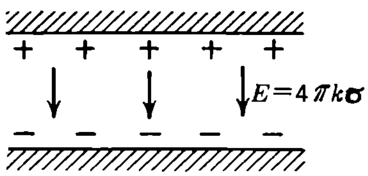


FIG. 25.10. The electric field between parallel plates.

The same relation [Eq. (25.11)] holds for the electric field between two parallel plates, one charged with a density of σ and the other a density of $-\sigma$, as shown in Fig. 25.10. The proof of this fact is left as an exercise for the student.

25.6. Fields within Conductors. Another important result can be obtained by reversing the reasoning just followed. Since the field inside a conductor must be zero for equilibrium, *no net charge can exist within a conductor*. For, suppose that the Gaussian surface is chosen to be just within the surface of the conductor, as in Fig. 25.11*a*; then the field is

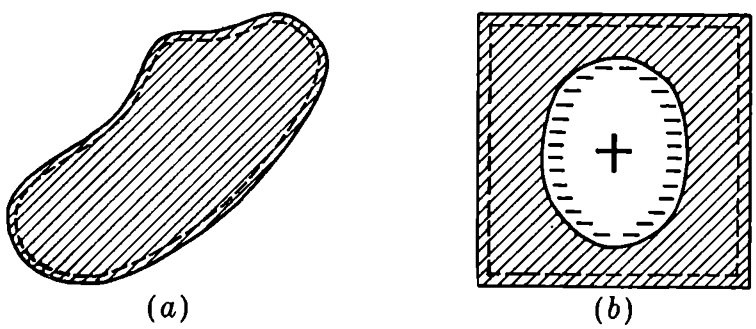


FIG. 25.11. A Gaussian surface just within the conductor is shown by the dashed line.

everywhere zero on the Gaussian surface, and the net charge enclosed by the surface must be zero. The net charge must also vanish for a hollow conductor such as is shown in Fig. 25.11*b*. If a positive charge is somehow introduced into the cavity, an equal negative charge must appear on the inside of the cavity walls. Conversely, if there is no net charge within the cavity, the net charge on the cavity walls must be zero.

An experimental test of the behavior of charges within a hollow conductor is the famous "ice-pail" experiment of Faraday (1791–1867).

Faraday is responsible for the useful concept of lines of force and did much to further our understanding of electric and magnetic phenomena. In this experiment a metal ice pail is insulated and connected to a sensitive electroscope, as indicated in Fig. 25.12. If a charged body is brought near the pail, the leaves of the electroscope become charged by induction and diverge. As the charged body is brought closer, the divergence increases. When the body is within the conducting pail, the divergence is a maximum and remains the same no matter where the charge is located within the pail. If the body is finally touched to the walls, no change in the divergence of the electroscope leaves takes place. The effect is explained in terms of induced charges. When the body is brought within the conducting pail, there must be an induced charge on the inner walls of the pail; touching the inside walls with the charged body merely neutralizes these charges, and the charge on the outer surface and on the electroscope is unaltered.

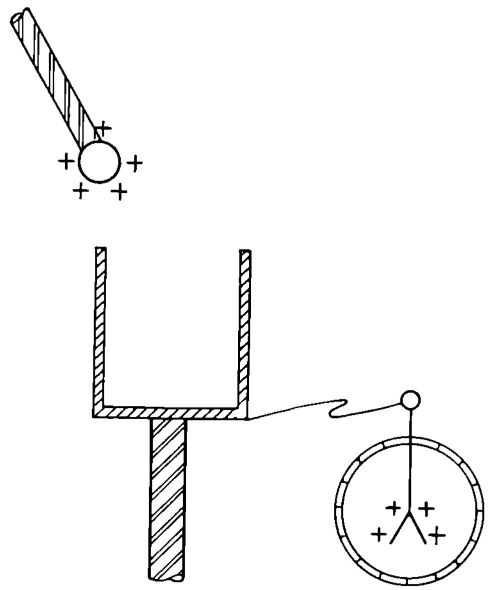


FIG. 25.12. Faraday's ice-pail experiment.

This important property of a conductor is extensively used to produce an enclosure that cannot be influenced by external electric forces. Since no electrostatic line of force can penetrate a conductor, such a cavity is effectively shielded from external effects and electric fields within the cavity cannot reach outside the conductor.

25.7. The Electric Potential. If a small electric charge q' is brought from a very large distance to a distance r from a point charge q by an external agency which exerts a force F, that external agency does an amount of work

$$\int_{\infty}^{r} F\,dr$$

But the force F is the negative of the Coulomb force between the charges, $F = -k(qq'/r^2)$. Hence the work is

$$W = -\int_{\infty}^{r} k\frac{qq'}{r^2}\,dr = k\frac{qq'}{r} \tag{25.12}$$

The external agency has supplied energy to the charge q', and the charge q' therefore possesses potential energy in the electric field of the charge q.

This potential energy is equal to $k(qq'/r)$. The situation is exactly analogous to the gravitational attraction of two masses except that the force between two like charges is repulsion and not attraction. By an argument similar to that used in Chap. 11 it can be shown that the potential energy is independent of the path by which the charge moves in from infinity and that the negative derivative of the potential energy is equal to the force. It is evident that the work done is proportional to the charge q'. In the study of electricity it is more convenient to speak of *the potential energy per unit charge, and this quantity is called the electrostatic potential.* The potential is usually denoted by V and hence

$$V = W/q' \tag{25.13}$$

For a point charge q the potential is, from Eq. (25.12),

$$V = k\frac{q}{r} \tag{25.14}$$

If a number of point charges produce the field, the potential V can be obtained by summing over many equations like Eq. (25.14), and

$$V = \sum_i k\frac{q_i}{r_i} \tag{25.15}$$

The sum expressed by Eq. (25.15) is correct because the potential is independent of the path by which the point in question is reached. The potential is, of course, a scalar quantity, and therefore it is much easier to add potentials than to add forces or electric fields. The field can always be obtained from the potential by differentiation.

The units in which potential is measured are evident from Eq. (25.14). In the electrostatic system of units, $k = 1$, and V is measured in statcoulombs per centimeter. This unit is a common one and is given the special name of *electrostatic volt* or *statvolt.* Since potential is energy per unit charge, one electrostatic volt is also equal to one erg per statcoulomb. In the practical system of units the unit of potential is called the *volt,* and it is equal to one joule per coulomb. In the practical system, Eq. (25.14) is written as

$$V\text{ (volts)} = \frac{1}{4\pi\epsilon_0}\frac{q\text{(coulombs)}}{r\text{(m)}}$$

where the units in which the various quantities are expressed are indicated within the parentheses. If the value of ϵ_0 is used, it is readily found that

$$299.6 \text{ volts} = 1 \text{ statvolt} \tag{25.16}$$

In Sec. 25.1 the practical unit of the electric field E was given as one newton per coulomb. It is more usual to use another unit derived from the unit of potential. Since the field is the negative derivative of potential with respect to distance, the unit of field can be taken to be one volt per meter. In the cgs system, fields are correspondingly expressed in statvolts per centimeter.

As an example of the calculation of potential, consider point A on the axis of the dipole in Fig. 25.3. The potential is, by Eq. (25.15),

$$V = \frac{kq}{x - (l/2)} - \frac{kq}{x + (l/2)} = \frac{kql}{x^2 - (l^2/4)}$$

The field in the x direction is

$$E_1 = -\frac{dV}{dx} = \frac{2kqlx}{[x^2 - (l^2/4)]^2}$$

which reduces for small values of l to Eq. (25.6).

The potential can also be found, of course, by calculating in general the work done by an external force in bringing a unit charge into the field. The external force per unit charge is $-\mathbf{E}$ and hence the vector $\mathbf{E}$ must be known for all points along the path. Since the amount of work done depends on the component of $\mathbf{E}$ along the path (see Sec. 8.4), the potential V at the point a can be written as

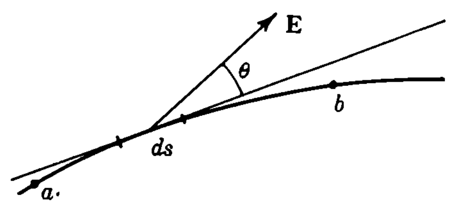


FIG. 25.13. Definitions of θ and ds.

$$V = -\int_{\infty}^{a} E \cos \theta \, ds \tag{25.17}$$

where θ is the angle between $\mathbf{E}$ and a tangent to the path and ds is an element of the path, as shown in Fig. 25.13. By differentiation of Eq. (25.17) the inverse relation is obtained.

$$E \cos \theta = -\frac{dV}{ds} \tag{25.18}$$

If $\theta = 0$, as in the case of a field due to a point charge or a charge of spherical symmetry, this relation takes the simple form

$$E = -\frac{dV}{dr}$$

It is often convenient to speak of the *potential difference* between two points in an electric field. The potential difference between points a and b is defined as

$$\begin{aligned} V_a - V_b &= -\int_\infty^a E \cos\theta \, ds + \int_\infty^b E \cos\theta \, ds \\ &= \int_a^b E \cos\theta \, ds \end{aligned} \tag{25.19}$$

If a charge q' is moved from one point in a field to another, the work done is given by

$$W = q'(V_a - V_b) \tag{25.20}$$

as is evident from Eq. (25.13). If V_a is greater than V_b, the charge q' loses potential energy when it is moved from point a to point b and a positive amount of energy W is produced.

Since the earth is a relatively good conductor and can usually be considered to be very far away from the charges and conductors under consideration, the potential of the earth is taken to be zero. Nearly always the walls of the laboratory are sufficiently good conductors and far enough away so that they can be assumed to have a potential of zero, or earth potential.

***25.8. Equipotential Surfaces.** Lines of electric force provide a good method for the graphical representation of electrostatic fields. The electric potential affords another equally useful method. Suppose we construct the surface in a field that is the locus of all points with the same value of V. Such a surface is called an *equipotential surface*. For a point charge the equipotential surfaces are obtained by putting V in Eq. (25.14) equal to a constant and are therefore represented by the equation r = const. Hence the surfaces are concentric spheres about the point charge.

If a charge is moved along an equipotential surface, no work is done. The component of the field along the surface must therefore be zero, and *the field is everywhere perpendicular to the equipotential surface.* The radial lines of force from a point charge are thus perpendicular to the spherical equipotential surfaces. If the lines of force in a field are parallel straight lines, the equipotential surfaces are planes normal to the lines. The equipotential surfaces for a dipole are indicated by the dotted lines in Fig. 25.4.

For a small motion of a charge along an equipotential surface, Eq. (25.18) can be applied; V is constant, and $-dV/ds$ must be zero. Consequently $\cos\theta$ is zero, θ is $\pi/2$, and the field is perpendicular to the surface. If a charge is moved along a line of force, θ is zero, and $\cos\theta$ is 1. Consequently the field is the rate of change of V in a direction perpendicular to the equipotential surface. Equipotential surfaces are contour lines of the electric potential, analogous to contours of equal elevation on a map. A particle moving along a contour on the side of a hill experiences no gravitational force along its direction of motion; a particle moving downhill from one contour to another is accelerated by gravity.

Since no work is done if a charge is moved along the surface of a conductor, *a conductor must be an equipotential surface.* In any arrangement of fixed charges and conductors the induced charges on the conductors must distribute themselves to make the surface of the conductor an equipotential surface with the lines of force, as we have already seen, perpendicular to the surface of the conductor. It is possible, therefore, to guess the configuration of the lines of force and of the equipotential surfaces and approximate closely the values of the field strength or potential in many cases where the evaluation by means of the general equations would be impractical.

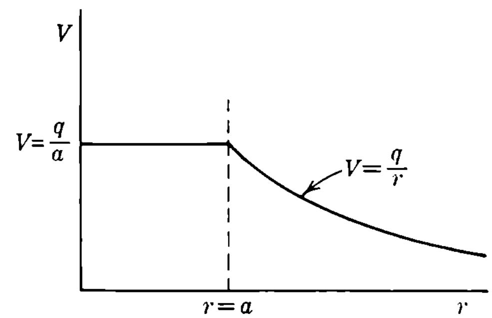


FIG. 25.14. The potential near a sphere of radius a and charge q ($k = 1$).

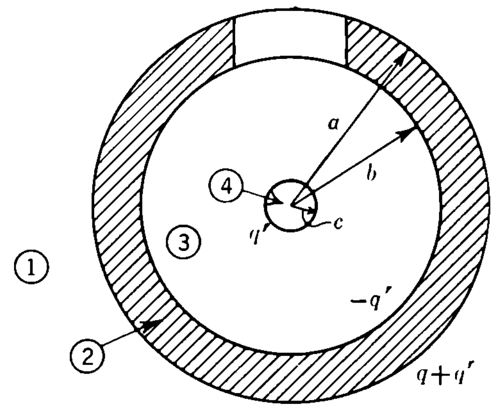


FIG. 25.15. One charged sphere inside another.

***25.9. The Potential of Charged Spheres.** It has already been noted that the field outside a spherical conductor with a charge q is equal to

$$E = k\frac{q}{r^2}$$

where r is the distance from the point under consideration to the center of the sphere. The potential is found from this relation and Eq. (25.17) to be

$$V = k\frac{q}{r}$$

If the radius of the conductor is a, the potential at the surface is $V = kq/a$. Within the conductor, whether it be solid or hollow, the field is zero if there is no net charge there. Consequently, for all values of $r < a$, the potential V is constant and equal to the value at the surface of the conductor. The potential therefore varies according to the curve of Fig. 25.14.

Let us suppose now that the sphere is hollow, the cavity has a radius b, and the sphere still carries a charge q. A second spherical conductor of radius c and charge q' is introduced into the cavity through a hole in the first conductor. We thus have the configuration of conductors and charges shown in Fig. 25.15, which is similar to the arrangement of Faraday's ice-pail experiment. A charge $-q'$ must flow to the inner walls of the first sphere, and a charge of $q + q'$ therefore resides on the outer surface of the outer conductor. The potential distribution near these spheres can be calculated if the effects of the hole in the first conductor are neglected. Let us divide the field into four regions, region (1) outside the outer conductor, region (2) inside the outer spherical shell, region (3) the space between the two spheres, and region (4) within the inner sphere. The electric fields in these regions will be

$$\left.\begin{aligned} &\text{Region (1)} \qquad && E_1 = k\frac{q+q'}{r^2} \\ &\text{Region (2)} && E_2 = 0 \\ &\text{Region (3)} && E_3 = k\frac{q'}{r^2} \\ &\text{Region (4)} && E_4 = 0 \end{aligned}\right\} \tag{25.21}$$

The potentials in the four regions can be found by the application of Eqs. (25.17) and (25.19) with $\cos\theta$ put equal to 1.

$$\left.\begin{aligned} &\text{Region (1)} \qquad && V_1 = -\int_\infty^r E_1 dr = k\frac{q+q'}{r} \\ &\text{Region (2)} && V_2 = -\int_\infty^a E_1 dr = k\frac{q+q'}{a} \\ &\text{Region (3)} && V_3 = V_2 - \int_b^r E_3 dr = k\left(\frac{q+q'}{a} - \frac{q'}{b} + \frac{q'}{r}\right) \\ &\text{Region (4)} && V_4 = V_3(\text{at } r = c) = k\left(\frac{q+q'}{a} - \frac{q'}{b} + \frac{q'}{c}\right) \end{aligned}\right\} \tag{25.22}$$

This variation of potential is plotted in Fig. 25.16.

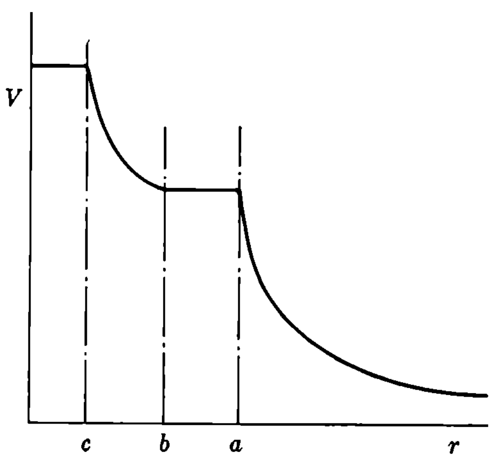


FIG. 25.16. The potential variation for the charged spheres of Fig. (25.15).

From these results several important conclusions can be drawn. If the outer shell were far removed from the inner shell, so that b and a became very large, the potential V_4 would be simply kq'/c. Thus in the presence of the outer shell, the potential V_4 of the inner sphere is different from the potential without the outer shell although the charge on the inner sphere remains the same. In particular, if the net charge q on the outer sphere is zero, the potential of the inner sphere is decreased when it is placed within the larger sphere. In a similar manner the presence of the inner sphere alters the potential V_2 of the outer sphere from kq/a to $k(q + q')/a$, an increase in potential. The charge on the outer shell is unchanged. It is evident, therefore, that the potential of a conductor depends not only on its charge but also on its surroundings. To emphasize this further, it should be noted that, even if the charge q on the outer sphere is zero, the potential of that sphere is q'/a. Thus the potential need not be zero even if the charge is zero. Likewise, if $q' = -q$, the potential of the outer sphere is zero, even though the conductor has a net charge q.

25.10. The Van der Graaff Electrostatic Generator. As we have seen in the preceding section, the potential of the outer sphere was increased when a charge was introduced into the interior of the sphere. This principle is employed to charge a large spherical electrode to a high potential in the electrostatic generator invented by Van der Graaff. The arrangement is represented schematically in Fig. 25.17. Charge is carried to a sphere by means of a moving belt. The charge is removed from the portion of the belt within the electrode, and the belt is charged again with charge of the opposite sign, which is then carried down by the belt. The charging mechanism C and the discharging mechanism D add or take away charges from the belt. These mechanisms usually consist of sharp rows of points like a comb charged with the proper kind of electricity, and the charges jump across through the air from the points to the belt. A set of brushes would perform the same function, but brushes do not work so well at high belt speeds. In this way as much as 1,000 μcoulombs of charge per second can be transported by the belt, and the spherical electrode is rapidly charged to several million volts of potential. In fact the potential that can be attained in this way is limited only by the charge conducted away along the insulating column that supports the sphere or through the surrounding air. To reduce the possibility of conduction through the air the generator is often enclosed in a tank, which can be filled with air or other gas to pressures above atmospheric pressure. Figure 25.18 is a photograph of a generator of this type. Both the exterior tank and interior are shown by a double exposure. The tank is about 7 ft high.

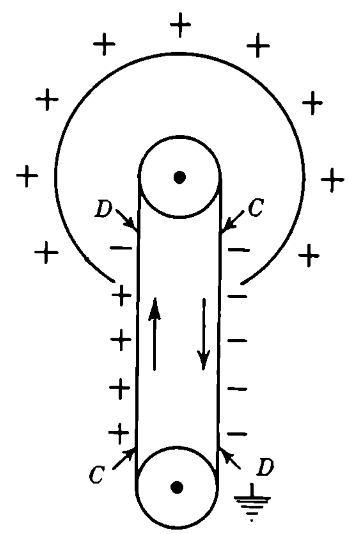


FIG. 25.17. Schematic diagram of a Van der Graaff generator.

Electrostatic generators of this type are used to impart large velocities to electrons or protons, which travel from the electrode at high potential down an evacuated tube and strike some target at the lower end of the equipment. A discussion of what happens upon impact must be postponed to Chap. 50.

25.11. The Motion of Charged Particles in Electric Fields. Let us consider a particle of charge e and mass m having a velocity v in the x direction and an electric field of strength E in the same direction.

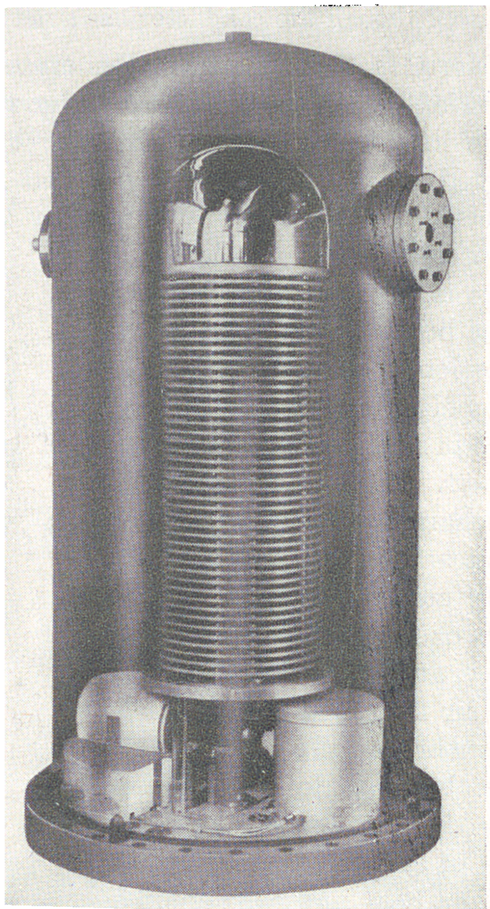


FIG. 25.18. A "ghost" photograph (double exposure) of a Van der Graaf generator. The series of rings is used to equalize the electric field up the supporting column. (*Courtesy High Voltage Engineering Corporation, Cambridge, Mass.*)

Newton's second law of motion then states

$$eE = m\frac{dv}{dt} \tag{25.23}$$

This equation can be integrated with respect to x, after transformation as in Sec. 8.7,

$$\frac{dv}{dt}\,dx = \frac{dx}{dt}\,dv = v\,dv$$

It yields

$$\int_{x_0}^{x_1} eE\,dx = \int_{v_0}^{v_1} mv\,dv = \tfrac{1}{2}mv_1^2 - \tfrac{1}{2}mv_0^2$$

The integral on the left is just what has been defined in Eq. (25.19) as $e(V_0 - V_1)$; hence

$$\tfrac{1}{2}mv_1^2 - \tfrac{1}{2}mv_0^2 = e(V_0 - V_1) \tag{25.24}$$

The change in kinetic energy is therefore equal to the negative of the change in potential energy, as it should be in a conservative system. The quantity $e(V_0 - V_1)$ of energy is that acquired by the particle in moving through a difference of potential equal to $V_0 - V_1$. The energy

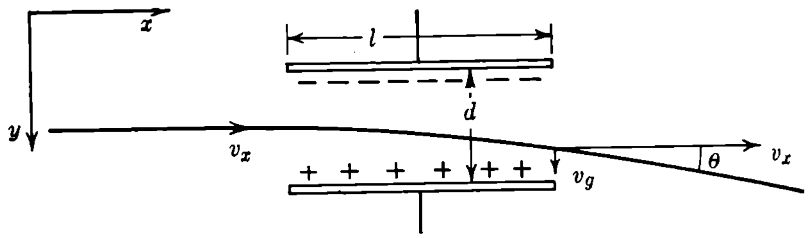


FIG. 25.19. The path of an electron between charged plates.

acquired by an electron in "falling through" a potential difference of one volt is often used as a *unit* of energy, called the *electron volt* (ev). We have

$$\begin{aligned} 1 \text{ ev} &= 1.6 \times 10^{-19} \text{ coulomb} \times 1 \text{ volt} \\ &= 1.6 \times 10^{-19} \text{ joule} = 1.6 \times 10^{-12} \text{ erg} \end{aligned}$$

Since the mass of the electron is 9.1×10^{-28} gm, the velocity with which an electron of one electron volt of kinetic energy moves is 6×10^7 cm/sec

$$\tfrac{1}{2}mv^2 = 1.6 \times 10^{-12} \text{ erg}$$

for A proton with one electron volt of energy has a smaller velocity corresponding to the larger mass.

If the field is uniform, *i.e.*, if **E** is constant, Eq. (25.23) can be immediately integrated twice with respect to t in the usual manner, with the result

$$x = \frac{1}{2}\frac{eEt^2}{m} + v_0t + x_0$$

Worked Example. An electron that has moved through a potential difference of 100 volts passes between two parallel plates 1 cm apart and 2 cm long and charged to a potential difference of 10 volts, as is shown in Fig. 25.19. What is the angle θ of deflection of the path of the electron? The field between the plates is $E = 10$ volts/cm. The velocity which the electron acquires in the y direction is

$$v_y = \frac{e}{m}Et$$

where t is the time necessary to traverse the plates. If the velocity corresponding to 100 volts is v_x and the length of the plates $l = 2$ cm, then

$$t = \frac{l}{v_x}$$

The angle θ of deflection is given by

$$\tan \theta = \frac{v_y}{v_x} = \frac{eEl}{mv_x^2}$$

This quantity can, however, be expressed in terms of the potential difference V' between the plates and the potential V that corresponds to the electron velocity. If d is the separation of the plates,

$$E = \frac{V'}{d}$$

$$\tfrac{1}{2}mv_x^2 = eV'$$

Therefore $$\tan \theta = \frac{l}{2d}\frac{V'}{V} = \frac{2}{2 \times 1}\frac{10}{100} = 0.1$$

and $$\theta = 5°42'$$

The deflection of an electron or of a beam of electrons by a field is employed in the cathode-ray oscilloscope described in Chap. 35.

25.12. The Oil-drop Experiment. By the study of the motion of charged particles in an electric field the value of the electronic charge can be determined experimentally. The motion of an electron is difficult to observe, but it is easily possible to observe the motion of small particles or droplets that are charged with one or two electronic charges. Observations of the rate of fall or rise of a small oil droplet between two charged parallel plates have been used by R. A. Millikan (1868–) to make a very precise determination of the smallest unit of electricity.[1] The experimental arrangement is shown diagrammatically in Fig. 25.20. Oil is used so that the mass of the drop will not change by evaporation. A droplet from the oil spray at the top of the apparatus falls through the tiny hole in the upper charged plate and is observed in the measuring microscope to the right. There are two forces acting on the droplet if it is charged: the force of the field upward and the force of gravity acting downward. These two forces can be balanced and the drop held stationary by adjusting the value of the field strength E between the plates. Let the charge on a drop of radius r be ne, where n is a small integer, either positive or negative, and e is the electronic charge. The total force F on the drop consists of the force of the electric field and the force of gravity,

[1] R. A. Millikan, "Electrons (+ and −)," University of Chicago Press, 1946.

It is

$$F = neE - \tfrac{4}{3}\pi r^3 dg$$

where d is the density of the drop. Thus if r and n are known, e can be determined. A charged drop can be balanced in this way by a field not too large only if r is very small—too small, in fact, to be measured directly. The drop size can be found, however, by observing the rate of fall of the drop when E is zero. As we saw in Chap. 15, the velocity of the drop soon attains a constant value v, as the drop falls, since the

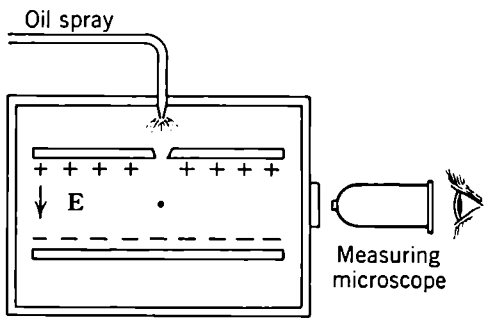


FIG. 25.20. Millikan's oil-drop experiment.

resistance of the air results in a retarding force that is proportional to the velocity and is given by the expression

$$F_a = 6\pi r\eta v$$

where η is the coefficient of viscosity of the air. The radius r can therefore be found from the relation

$$\tfrac{4}{3}\pi r^3 dg = 6\pi r\eta v$$

when E is absent. The small integer n can be found by making successive observations on the same droplet. It is observed that a drop, particularly when ultraviolet light, X rays, or the radiations from radioactive substances are allowed to fall upon it, is continually gaining or losing electrons to the surrounding air. Thus a whole series of values of ne can be observed, and the smallest value must be that for which $n = 1$. Millikan has recorded observations in which as many as 15 different values of n have been observed with a single drop. Each observation was of high precision, and all led to the same value of e. More direct or convincing evidence of the atomic nature of electricity can hardly be imagined. The value of e has been demonstrated to be independent of the nature of the drop, the density and pressure of the gas, the field, and all the other variables of the experiment.

PROBLEMS

1. What is the value of the field at the origin, Fig. 25.3?

2. Sketch the lines of force near two equal positive charges separated by a small distance. What is the field midway between the two charges? How would the lines of force near two equal negative charges differ from this?

3. Two dipoles, parallel but oppositely directed and separated a small distance, are known as an "electric quadripole." Sketch the lines of force near such a quadripole.

4. What is the electric field at a point with coordinates (x,y) due to the dipole of Fig. 25.3?

5. The dipole moment of a water molecule is 1.8×10^{-18} esu; what is its value in practical mks units? By what distance must two electronic charges be separated to produce a moment of this value?

6. Find the value of the electric field within a long cylindrical shell of charge.

7. Obtain the value of the potential at point A of Fig. 25.3 by integration of the field strength.

8. Sketch the lines of force and the equipotential surfaces near a right-angle corner of a long charged conductor.

9. An electron moves under the action of a uniform field in the x direction of 100 volts/cm, starting from the origin with zero velocity. What are its position, velocity, and acceleration after 0.01 μsec?

10. What is the average kinetic energy in electron volts of a hydrogen molecule at 0°C?

11. What is the velocity of a proton with 1 ev of energy; with 10^6 ev?

12. An oil drop is suspended motionless in a field of 1,000 volts/cm. If the density of the oil is 0.9 gm/cm^3 and the drop has one electron of charge, what is its radius?

13. The earth, in temperate latitudes, has an electric field at its surface of approximately 100 volts/m directed downward. What charge density must exist on the surface? If this field were the same over the whole earth, what would be the total charge on the earth? What would be its potential?

14. A positive charge Q is uniformly distributed throughout a spherical volume of radius R. What is the field at a distance $r < R$ from the center?

15. Calculate the time taken by the electron in the worked example on page 349 to traverse the plates.

CHAPTER 26

DIELECTRICS AND CAPACITANCE

26.1. Polarization. When a metal is placed in an electric field, the free electrons in it move, as we have seen, always distributing themselves in such a way that the metal becomes an equipotential surface. Let us now investigate what happens when a nonconductor is placed in an electric field. A nonconductor, or *dielectric material*, contains no charges that are free to move about within it, but rather all the electrons are tightly bound to the atomic nuclei. These electrons do, however, experience a force when the electric field is applied. Consequently the field produces a change in the distribution of electricity within the atoms of the dielectric, since the motion of each electron is confined to the immediate vicinity of one individual nucleus. One of two things may happen.

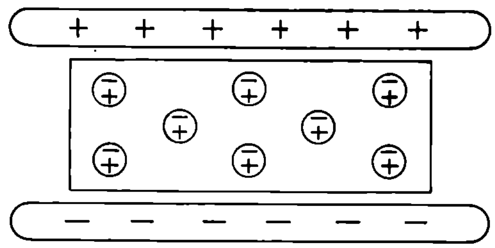


FIG. 26.1. The polarization of atoms in a dielectric.

1. If the charge in each atom of the material is symmetrically distributed, as it is in crystalline sodium chloride, the electric field produces a concentration of negative charge on one side of the atom and a small electric dipole is produced. Each atom acts as a small spherical conductor and becomes polarized. Under the influence of a uniform field the atoms may be schematically represented as in Fig. 26.1. The dielectric becomes composed of little electric dipoles, all having the same strength and all oriented in the same direction.

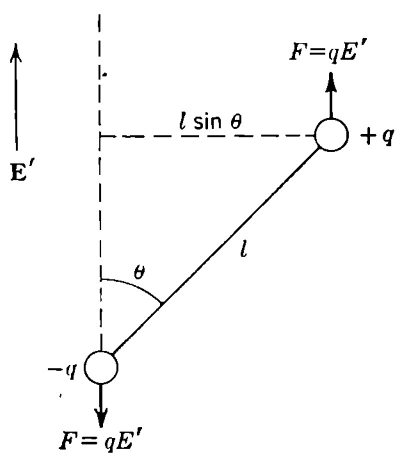


FIG. 26.2. The torque on a dipole in the field **E'**.

2. It may happen that the charges on the atoms of the dielectric are not uniformly distributed even in the absence of an electric field. The molecules have a permanent dipole moment. Water is such a substance, for example. The application of an electric field results in a torque on each molecule that tends to align the dipoles in the same direction. The student can easily show, using Fig. 26.2, that this torque

L is given by

$$L = qlE' \sin \theta = pE' \sin \theta \tag{26.1}$$

where ql is the strength p of the dipole, θ the angle between the field and the direction of the dipole, and E' the total electric field tending to rotate the dipole. The motion that this torque tends to produce is resisted by the random heat motions of the atoms and perhaps is also partly counteracted by torques from the field of neighboring atoms. The net result, however, is that on the average there is some alignment with the applied field. The situation is schematically represented in Fig. 26.3.

The effect of the rotation of dipoles is evidently very much the same as that of polarization by the displacement of charges. In both cases the dielectric material placed between charged metal plates effectively acquires a surface charge, as indicated in Fig. 26.4. The electric field within a dielectric is therefore that produced by the net effect of the free charges on the metal plates and the bound charges on the surface of the dielectric. The strength of the resultant field thus depends upon the strength of the electric dipoles in the dielectric. *The dipole moment per unit volume of the dielectric material is called the polarization P.* The polarization is a vector quantity and has the direction of the dipole moment. If there are n dipoles per unit volume, each of moment p, and all of them are aligned with the field then

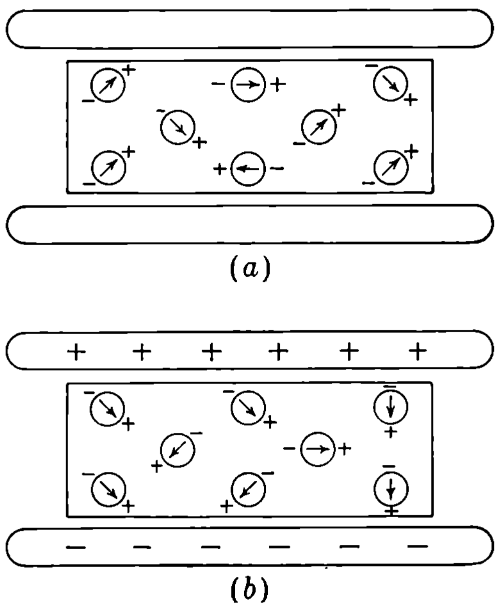


FIG. 26.3. The alignment of molecules with permanent electric dipoles in a field.

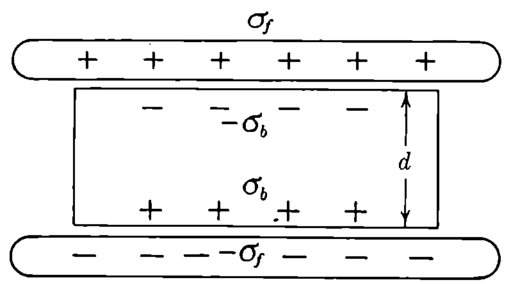


FIG. 26.4. The free and bound charges at the surface of a dielectric.

$$P = np \tag{26.2}$$

Some values of the electric moment p are given in Table 26.1.

Since the units of p are those of charge times distance, the units of P are those of charge divided by the square of the distance. In practical

Table 26.1. *Values of the Electric Moments of Molecules*

Substance	Formula	Electric moment p, esu
Water	H_2O	1.8×10^{-18}
Ammonia	NH_3	1.5×10^{-18}
Hydrogen chloride	HCl	1.03×10^{-18}
Hydrogen bromide	HBr	0.78×10^{-18}
Hydrogen iodide	HI	0.38×10^{-18}

units, therefore, P is measured in coulombs per square meter; in esu, in statcoulombs per square centimeter.

For the dielectric slab of material of Fig. 26.4 the polarization can be expressed in terms of the magnitude of the bound charges. If the surface density of these charges is σ_b, then the total dipole moment is $\sigma_b Ad$, where A is the area and d the thickness of the slab. The moment per unit volume is then

$$P = \frac{\sigma_b Ad}{Ad} = \sigma_b \tag{26.3}$$

It is certainly expected, and indeed found to be accurately true for most substances, that P is proportional to the applied field E. Thus

$$\boxed{P = \chi_e E} \tag{26.4}$$

where χ_e, the constant of proportionality, is called the *electric susceptibility*. Since the practical units of the polarization P are coulombs/m^2, and the units of E are volts/m, the practical units of χ_e are evidently coulombs per (volt times meter). In the electrostatic system, χ_e is a dimensionless quantity. One statvolt is the potential 1 cm from a charge of 1 statcoulomb, and hence

$$[\chi_e] = \left[\frac{\text{statcoulomb}}{\text{statvo t} \times \text{cm}}\right] = \left[\frac{\text{statcoulomb}}{\text{statcoulomb}}\right]$$

26.2. Specific Inductive Capacity and Dielectric Constant. In addition to the susceptibility, other quantities are used to characterize the behavior of a dielectric material. These can best be understood from a calculation of the field within the dielectric of Fig. 26.4. If the density of free charge on the metal plates is denoted by σ_f, then the total field-producing charge is $\sigma_f - \sigma_b$ and the resultant field is

$$E = 4\pi k(\sigma_f - \sigma_b) \tag{26.5}$$

where k is the constant whose value depends on the units used. If Eqs. (26.3) and (26.4) are used, the field can be expressed in terms of the free charges alone as $E = 4\pi k\,(\sigma_f - \chi_e\,E)$

$$E = \frac{4\pi k\sigma_f}{1 + 4\pi k\chi_e} \tag{26.6}$$

The denominator of this expression occurs often, and we shall denote it by the quantity K.

$$\boxed{K = 1 + 4\pi k\chi_e} \tag{26.7}$$

The quantity K is called the *specific inductive capacity* of the material. It should be noticed that, if no dielectric is present, Eq. (26.6) becomes simply

$$E = 4\pi k\sigma_f$$

With a dielectric it is

$$E = \frac{4\pi k\sigma_f}{K} \tag{26.8}$$

Thus K may be defined as the ratio of the field strength without a dielectric to that when a dielectric is present. Although this relationship has been demonstrated only for a uniform field, it is true in general.

$$E(\text{with dielectric}) = \frac{E}{K}\,(\text{without dielectric})$$

Since K is a ratio, it can have no dimensions and its numerical value is independent of the units used. If esu are used, K is often called the *dielectric constant* of the material.

Units. When practical mks units are employed, a third constant is defined. If k is given the value $1/4\pi\epsilon_0$, then Eq. (26.7) can be written

$$K = 1 + \frac{\chi_e}{\epsilon_0} \qquad \text{mks units}$$

If this equation is multiplied by ϵ_0, the left-hand side becomes $\epsilon_0 K$ and this combination is called the *permittivity* of the dielectric and is denoted by ϵ. Thus ϵ_0 can also be spoken of as the "permittivity of free space." Permittivity is rarely used in other systems of units than the practical one.

The relations between these various constants are easily summarized. For practical units,

$$\left.\begin{aligned} K &= 1 + \frac{\chi_e}{\epsilon_0} = \frac{\epsilon}{\epsilon_0} \\ \epsilon &= K\epsilon_0 = \epsilon_0 + \chi_e \\ \chi_e &= \epsilon - \epsilon_0 = (K - 1)\epsilon_0 \end{aligned}\right\} \quad \text{mks} \tag{26.9}$$

In the electrostatic system of units we have

$$\left.\begin{aligned} K &= 1 + 4\pi\chi_e \\ \chi_e &= \frac{K-1}{4\pi} \end{aligned}\right\} \quad \text{esu} \tag{26.10}$$

The relation between the susceptibilities in the two systems of units is

$$\chi_e\ (\text{mks}) = \frac{1}{k}\,\chi_e\ (\text{esu}) = 4\pi\epsilon_0\chi_e\ (\text{esu}) \tag{26.11}$$

Some values of these parameters for several common substances are given in Table 26.2.

Table 26.2. *Values of Permittivity and Specific Inductive Capacity*

Substance	Specific inductive capacity K	Permittivity ϵ	Electric susceptibility χ_e, esu	Electric susceptibility χ_e, mks
Vacuum..........	1	$\epsilon_0 = 8.85 \times 10^{-12}$	0	0
Air at NTP*........	1.00059	8.85×10^{-12}	4.77×10^{-5}	5.22×10^{-16}
Sulfur..............	4.0	3.5×10^{-11}	0.24	2.6×10^{-12}
Glass (mean).......	6	5×10^{-11}	0.4	4.4×10^{-11}
Mica	5.7	5.0×10^{-11}	0.37	4.2×10^{-11}
Polystyrene.........	2.55	2.26×10^{-11}	0.12	1.37×10^{-11}
Water..............	81.1	7.18×10^{-10}	6.4	7.1×10^{-10}

* Normal temperature and pressure.

It is evident that for all but the most accurate calculations K can be taken as unity for air as well as for vacuum.

From the first of Eqs. (26.9), since K has no dimensions, it is evident that both ϵ and ϵ_0 have the same dimensions as χ_e, which are coulombs per (volt times meter).

***26.3. The Force between Charges Immersed in a Dielectric.** Two small conductors in empty space with charges q and q' repel one another with a force given by Coulomb's law,

$$F = k\,\frac{qq'}{r^2} \tag{24.1}$$

where r is the distance between them. If these conductors are immersed in a dielectric medium, the force is altered. The field of each charge polarizes the dielectric near it, and hence the field of one charge acting on the other is reduced. Let us suppose that the charge q induces a charge $-q_b$ in the dielectric near it. The field at the location of q' is, then,

$$E = k\,\frac{q - q_b}{r^2} \tag{26.12}$$

and the force on q' is

$$F = Eq' = k\frac{(q - q_b)q'}{r^2} \tag{26.13}$$

The force is thus smaller when the dielectric is present. In the preceding section it was shown that the presence of the dielectric reduced the field between parallel plates by a factor of K. The same rule must apply here also, and Eqs. (26.12) and (26.13) become

$$E = k\frac{q}{Kr^2} \tag{26.14}$$

$$F = k\frac{qq'}{Kr^2} \tag{26.15}$$

Equation (26.14) is *Coulomb's law for dielectric materials.*

A comparison of Eqs. (26.12) and (26.14) or Eqs. (26.13) and (26.15) shows that q_b is proportional to q and in fact

$$q_b = \frac{K-1}{K}q = \frac{4\pi k\chi_e}{1 + 4\pi k\chi_e}q \tag{26.16}$$

A similar relation must hold, of course, between the charge q' and the bound charge $-q'_b$ that is induced in the dielectric near it.

***26.4. Gauss's Law for Dielectric Materials.** In Sec. 25.4 Gauss's law was stated as the fundamental relation between charge and the electric field,

$$\int_{\text{surface}} E\cos\theta\,dA = 4\pi kq_t \tag{25.9}$$

This relation is valid also in the presence of dielectric material, but it must be remembered that q_t refers to the total charge within the Gaussian surface. Thus q_t must be made up of the free charge q_f that is present on the conductor and the bound charge $-q_b$ that the field of q_f induces in the dielectric,

$$q_t = q_f - q_b$$

It is desirable to have Gauss's law expressed in terms of the free charge on the conductors rather than the total charge. The bound and the free charges are related by Eq. (26.16), and hence Eq. (25.9) can be written,

$$\int_S E\cos\theta\,dA = 4\pi k(q_f - q_b) = 4\pi kq_f\left(1 - \frac{K-1}{K}\right) = \frac{4\pi kq_f}{K} \tag{26.17}$$

Equation (26.17) is thus in a form that contains explicitly the effect of the medium through the constant K.

For some purposes it is convenient to define a new vector quantity, the electric displacement $\mathbf{D}$, associated with the electric field in a dielectric. In the practical mks system the displacement $\mathbf{D}$ is defined as

$$\mathbf{D} = \frac{K}{4\pi k}\mathbf{E} = K\epsilon_0\mathbf{E} = \epsilon\mathbf{E} \qquad \text{mks units} \tag{26.18}$$

and hence Gauss's law becomes

$$\int_S D \cos \theta \, dA = q_f \qquad \text{mks units} \tag{26.19}$$

The displacement **D** is measured in coulombs per square meter. By the use of the relations in Eqs. (26.9), **D** can be expressed in terms of the other vectors **E** and **P**,

$$\mathbf{D} = \epsilon\mathbf{E} = (\epsilon_0 + \chi_e)\mathbf{E} = \epsilon_0\mathbf{E} + \mathbf{P} \qquad \text{mks units} \tag{26.20}$$

This relation between **D** and **E** and **P** should be considered more fundamental than Eq. (26.18). Since the electric displacement has a value at every point in a field, it can be represented graphically by means of lines of displacement, similar to the representation of **E** by lines of electric force.

26.5. The Capacitance of an Isolated Conductor. The potential of an isolated sphere of radius a was shown in Eq. (25.20) to be proportional to the charge q on the sphere,

$$V = \frac{kq}{a} \tag{25.20}$$

For a body of another shape the potential is still proportional to the charge, but with a different constant. This proportionality is often expressed in terms of the *capacitance* C of the body defined by

$$\boxed{C = \frac{q}{V}} \tag{26.21}$$

The capacitance of a sphere is therefore

$$C \text{ (sphere)} = \frac{a}{k} \tag{26.22}$$

If the charged sphere is immersed in a dielectric medium, the electric field everywhere, and also its integral over the distance, or the potential, is reduced by a factor of K. Consequently the capacitance is increased by this same factor,

$$C \text{ (sphere)} = K\frac{a}{k} \tag{26.23}$$

Units. Equation (26.22) takes the two alternate forms

$$\begin{aligned} C(\text{sphere}) &= a \qquad \text{esu} \\ C(\text{sphere}) &= 4\pi\epsilon_0 a \qquad \text{mks units} \end{aligned} \tag{26.24}$$

The capacitance of a conducting body depends only on the geometrical shape of the body. In esu the capacitance is measured in centimeters.

In the mks system, however, the unit of capacitance is one coulomb per volt and is called the *farad*. This unit is too large for convenient use, since ϵ_0 is so small, and the submultiples microfarad (μf) and micromicrofarad ($\mu\mu$f) are customarily employed. The units of the quantity ϵ_0 are usually expressed in terms of the farad. It is evident from Eq. (26.24) that ϵ_0 has the dimensions of farads per meter. The permittivity ϵ and the electric susceptibility χ_e must also have these same dimensions.

In mks units Eq. (26.23) is written

$$C(\text{sphere}) = 4\pi\epsilon_0 Ka = 4\pi\epsilon a$$

Worked Example. What is the capacitance in $\mu\mu$f of a sphere 1 cm in radius? From Eq. (26.24)

$$C = 4\pi\epsilon_0 a = 4\pi \times 8.85 \times 10^{-12} \times 10^{-2} \text{ farad} = 1.1 \ \mu\mu\text{f}$$

This result gives, of course, the relation between the electrostatic and the practical units of capacitance.

26.6. Capacitors and Capacitance. If several charged conductors are present, the potential of any one conductor depends not only upon its charge but also upon the charges on all the others. The capacitance as defined by Eq. (26.21) no longer has a unique meaning. If two of the conductors are close together and far away from all the others, then capacitance may be defined by

$$\boxed{C = \frac{Q}{V_{ab}}} \tag{26.25}$$

where Q is the charge on either conductor and V_{ab} is the difference in potential between them. If the two conductors are close together, all the lines of force that leave one conductor must end upon the other and the charges are consequently equal. Such a pair of conductors is called a *capacitor;* an older name still often used is *condenser*.

Suppose, for example, that there are two parallel charged plates with a dielectric material between them as in Fig. 26.3. The charge Q on one plate is $\sigma_f A$, where A is the area of the plate. The potential difference between the plates can be found by integrating the field given in Eq. (26.8),

$$V_{ab} = \int E\,ds = \frac{4\pi k\sigma_f}{K}\,d$$

where d is the distance between the plates. The capacitance is the quotient of the charge divided by this potential difference, and hence

$$C = \frac{KA}{4\pi k d} \tag{26.26}$$

The capacitance in esu is obtained by setting $k = 1$, or

$$C = \frac{KA}{4\pi d} \qquad \text{cms}$$

In practical units, $k = 1/4\pi\epsilon_0$,

$$C = \epsilon_0 K \frac{A}{d} = \epsilon \frac{A}{d} \qquad \text{farads}$$

Capacitors are often used as elements in electric circuits. To indicate a capacitor a standard symbol is used, which is shown in Fig. 26.5.

A capacitor can be constructed of two concentric spheres. If the radii of the inner and outer spheres are a and b, respectively, as in Fig. 26.6, the potential difference between them is

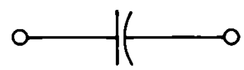


FIG. 26.5. The circuit symbol for a capacitor.

$$V_{ab} = \frac{kQ}{K}\left(\frac{1}{a} - \frac{1}{b}\right)$$

as shown by Eqs. (25.22), where Q is the charge on the outside of the inner sphere, and $-Q$ the charge on the inside of the outer sphere. The capacitance is therefore

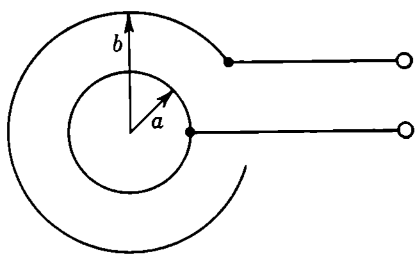


FIG. 26.6. A spherical capacitor.

$$C = \frac{K}{k\left(\frac{1}{a} - \frac{1}{b}\right)} = \frac{Kab}{k(b-a)} \tag{26.27}$$

It should be noted that an additional charge may exist on the outer surface of the outer sphere. This additional charge may increase or decrease the potential of both spheres, but the potential difference between them, and therefore the capacitance, is independent of this charge.

The capacitance C_1 per unit area of the outer sphere of a spherical capacitor is

$$C_1 = \frac{Kab}{k(b-a)4\pi b^2}$$

and the capacitance of a small area A is

$$C = \frac{KabA}{k(b-a)4\pi b^2} \tag{26.28}$$

If the radii a and b are assumed to become very large while the difference

$(b - a)$ stays constant, ab is nearly equal to b^2 and, the capacitance given by Eq. (26.28) approaches that for two parallel plates [Eq. (25.26)],

$$C = \frac{KA}{k(b - a)4\pi}$$

26.7. Practical Capacitors. The capacitors used in practical work take many forms, nearly all of which are modifications of parallel plates. A stack of plates can be used, separated by thin sheets of a dielectric, usually mica, with alternate plates connected together. The whole arrangement is often molded together with an insulating plastic material, with two terminal wires protruding. Ceramic materials or glass are also used as the dielectric. Mica capacitors and some ceramics result in capacitances that are very constant with temperature, often an important consideration.

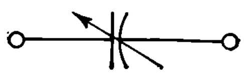


FIG. 26.7. A variable capacitor.

Another form of capacitor is constructed by winding layers of metal foil and insulating paper dielectric into a small cylinder, the whole being impregnated with wax to keep out moisture. Such wound-paper capacitors are very inexpensive and are extensively used. For better insulation a capacitor of this type is often impregnated with oil and sealed into a can with porcelain bushings for the terminals.

If large values of capacitance are desired, the area can be made large, a dielectric with a large value of K can be chosen, and the separation d between the plates can be made very small. Small values of d are obtained in the so-called "electrolytic" capacitors. These capacitors are constructed with plates of aluminum, and the dielectric is an extremely thin coating of aluminum oxide on the surface of the plates, which is made by an electrolytic process. Although such capacitors cannot withstand high potential differences, very large capacitances can be obtained. An old form of capacitor is the Leyden jar, which is simply a glass bottle coated inside and out with metal foil. Leyden jars are now found only as antiquated equipment in physics laboratories.

If only small capacitances are desired, air can be used as the dielectric material. Stacks of plates alternately connected are frequently used for this purpose. A variable capacitor can be constructed easily by mounting one set of plates on a shaft so that they can be rotated away from another set of plates. All radio sets contain capacitors of this type. The symbol for a variable capacitor is shown in Fig. 26.7.

26.8. Capacitors in Parallel and in Series. Several capacitors may be combined in an electric circuit to form the equivalent of a single capacitor. The connection shown in Fig. 26.8 is called a "parallel" connection. The potential difference across each capacitor is the same; let this be denoted by V. The total charge Q on the upper plates of the capacitors is the sum of the individual charges; hence

$$\frac{Q}{V} = \frac{Q_1}{V} + \frac{Q_2}{V} + \frac{Q_3}{V}$$

or, by Eq. (26.25)

$$C = C_1 + C_2 + C_3 \tag{26.29}$$

In Fig. 26.9 several capacitors are shown connected in series, or in cascade. The system is evidently composed of four conductors, numbered (1) to (4) in the figure. If a charge Q is initially placed on conductor (1) and there is no net charge on conductors (2) and (3), the

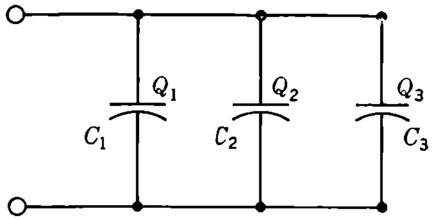


FIG. 26.8. Capacitors in parallel.

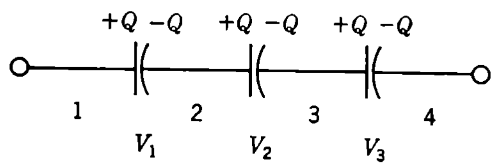


FIG. 26.9. Capacitors connected in series.

distribution of charges becomes as shown. The sum of the potential differences across the individual capacitors is equal to the total potential difference. Therefore we can write

$$\frac{V}{Q} = \frac{V_1}{Q} + \frac{V_2}{Q} + \frac{V_3}{Q}$$

or

$$\frac{1}{C} = \frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3} \tag{26.30}$$

It is evident that Eqs. (26.29) and (26.30) can be generalized to any number of capacitors.

***26.9. The Energy of a Charged Capacitor.** In order to charge a capacitor, work must be done in transferring charge from one plate to the other. This energy is stored in the capacitor and can be recovered when the capacitor is discharged. The stored energy can be calculated easily. Suppose the capacitor has a charge q and this charge is increased to $q + dq$. The amount of work that must be done is, by the definition of potential,

$$dW = v\,dq$$

where v is the potential difference across the plates when the charge is q. Since $v = q/C$, the total amount of work done for a total charge of Q is

$$W = \int dW = \int_0^Q \frac{q}{C}\,dq = \frac{1}{2}\frac{Q^2}{C} \tag{26.31}$$

Since the final potential $V = Q/C$, the energy can also be written

$$W = \tfrac{1}{2}CV^2 = \tfrac{1}{2}QV \qquad (26.32)$$

This stored energy may be regarded to reside in the electric field that is produced. In a parallel-plate capacitor the electric field is confined between the plates and has the same value at all points within this region. The energy per unit volume, or *energy density*, is therefore

$$\frac{W}{Ad} = \frac{1}{2}\frac{CV^2}{Ad}$$

If Eq. (26.26) is used for C, there results

$$\frac{W}{Ad} = \frac{1}{2}\frac{K}{4\pi k}\left(\frac{V}{d}\right)^2 = \frac{1}{2}\frac{K}{4\pi k}E^2$$

Units. The energy density in the electrostatic system where $k = 1$ is therefore

$$\frac{W}{Ad} = \frac{1}{8\pi}KE^2 \qquad (26.33)$$

In the practical system, $k = 1/4\pi\epsilon_0$,

$$\frac{W}{Ad} = \frac{1}{2}\epsilon_0 KE^2 = \frac{1}{2}\epsilon E^2 \qquad (26.34)$$

Although Eqs. (26.33) and (26.34) have been derived for a special case, they represent general expressions for the energy density in any field configuration.

PROBLEMS

1. Water molecules have permanent electric moments. How does the permittivity of water change with temperature? How should you expect the permittivity of ice to compare with that of water? Why?

2. Compare the polarization of water in a field of 1,000 volts/cm with the polarization that would result if the dipole moments of all the molecules were perfectly aligned.

3. An insulating oil has a specific inductive capacity of 2.5. What is the value of χ_e in practical units and in esu?

4. Find the radius of a sphere whose capacitance is 1 farad.

5. What is the capacitance of the earth in centimeters; in microfarads?

6. A capacitor is constructed of a pile of 10 parallel plates each of area 10 cm², alternate plates being connected together. The plates are separated from one another by 0.3 mm of mica. What is the capacitance?

7. What is the capacitance of a capacitor constructed of three concentric spheres of radii 1, 2, and 3 cm, with the outer and the innermost spheres connected together?

8. A glass plate 2 mm thick, of specific inductive capacity 8, is used as the dielectric in a parallel-plate capacitor with plates 10 cm² in area. If the potential difference is 20 statvolts, what is the polarization of the glass?

***9.** What is the stored energy in the capacitor of Prob. 8? What is the energy density?

10. What is the capacitance per unit length of two coaxial cylinders?

11. A 1-μf capacitor is charged by connecting it to another capacitor of 2 μf capacitance charged to 100 volts. What are the charges on each?

12. Three capacitors are connected as shown in Fig. 26.10. What is the capacitance of the combination?

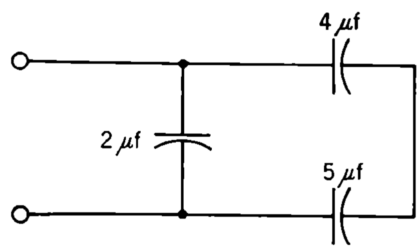


FIG. 26.10. Circuit for Problem 12.

***13.** Metals are often welded together by the energy dissipated in the discharge of a large capacitor. Suppose 1,000 μf charged to 100 volts is used; how much energy in calories is available?

14. A soap bubble receives a charge. Does it become larger or smaller? Explain.

15. A spherical capacitor with radii 6 and 10 cm has the outer sphere split in half so that the pieces can be removed from around the inner sphere. If the potential difference between the spheres is 100 volts, how much work must be done to remove the outer conductor and join the pieces together again far away from the inner sphere?

CHAPTER 27
ELECTRIC CURRENTS

27.1. Electric Current and Current Density. The motion of charges under the action of an electric field produces an electric current, just as the motion of the molecules of a fluid in a gravitational field produces a current in the fluid. If water flows in a pipe of cross-sectional area A, the number of water molecules crossing a transverse plane in the pipe per second is equal to NvA, where N is the number of molecules per unit volume and v is the velocity of flow. In a similar fashion the strength of an electric current is measured by the number of electrons crossing a given plane per second. If a current is flowing in a wire of area A, *the strength of the current is defined as the amount of charge crossing a transverse plane per second.* The current strength i, or more simply the current, is therefore

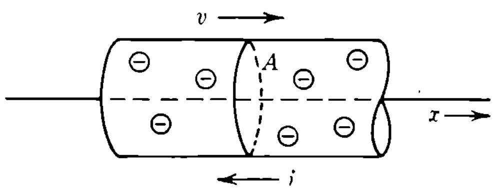


FIG. 27.1. A flow of electrons constitutes a current.

$$i = \frac{dq}{dt} = nevA \tag{27.1}$$

where n is the number of electrons per unit volume in the wire, v the average velocity of an electron, and e the charge on each electron.

It should be remembered that the charge e is a negative quantity. If the positive x direction is to the right and electrons are moving from left to right as indicated in Fig. 27.1, then v is a positive quantity. The current i is, by Eq. (27.1), negative and flows from right to left. Electrons therefore move in the *opposite* direction to the electric current. This rather strange circumstance is not difficult to understand. It arises from the arbitrary definitions of positive and negative electricity. The kinds of electricity were named before it was realized that negatively charged electrons move when an electric current flows.

The unit of current is defined as the flow of a unit charge per second. In practical units a unit current is thus one coulomb per second, and this current is called one ampere in honor of the discoverer, André Ampère (1775–1836), of the magnetic effects of electric currents. In esu a unit current is one statcoulomb per second, or one statampere. From Eq.

(24.3), the two units are connected by the relation

$$1 \text{ amp} = 2.996 \times 10^9 \text{ statamp} \tag{27.2}$$

It is of interest to calculate by Eq. (27.1) the value of the average velocity v for a unit current. Suppose that 1 amp is flowing in a copper wire having a cross-sectional area of 1 mm². Since the practical units are based on the meter as a unit of length, A is 10^{-6} m², the number n of electrons per cubic meter, which is equal to the number of atoms per cubic meter, is 8.4×10^{28}, and

$$v = \frac{1}{8.4 \times 10^{28} \times 1.6 \times 10^{-19} \times 10^{-6}} = 7.5 \times 10^{-5} \text{ m/sec}$$

The value of v is much smaller than the rms velocity of electrons in a metal as calculated in Sec. 24.8, just as the velocity of flow of a gas through a pipe is small compared with the rms velocity of the gas molecules.

The electric current flowing across a *unit area* is called the *current density* and is denoted by J. From Eq. (27.1)

$$J = \frac{i}{A} = nev \tag{27.3}$$

In practical units J is measured in amperes per square meter.

27.2. Conductivity and Resistance. The magnitude of the current density that flows when a given electric field is applied depends upon the material. The number n of electrons varies from one substance to another, and v also changes. Over a large range of substances and values of the electric field the current density is proportional to the field. In terms of symbols we may write

$$\boxed{J = \sigma E} \tag{27.4}$$

where the constant of proportionality, σ, is called the *conductivity* of the material. Equation (27.4) is known as *Ohm's law*. It is evident that the electric conductivity σ is defined in a similar manner to the heat conductivity k. The conductivity is characteristic of the material and contains the variations of both n and v. If Eqs. (27.3) and (27.4) are combined, there results

$$\sigma = ne\frac{v}{E} \tag{27.5}$$

The ratio v/E is called the "mobility" of the charges.

In a wire of cross-sectional area A, the current can be written as

$$i = JA = \sigma AE = -\sigma A\frac{dV}{dx}$$

where V is the electric potential and the current is flowing in the x direction. If the electric field is uniform over a length l of the wire, we can write

$$i = \sigma A \left(\frac{V_1 - V_2}{l} \right) \tag{27.6}$$

where V_1 and V_2 are the values of the potential at the ends of the length l of the wire, as indicated in Fig. 27.2. The figure has been drawn on the assumption that $V_1 > V_2$, and hence **E** has the direction indicated. The quantity dV/dx is therefore negative, and i is positive. The current thus flows from a place of high potential to a place of low potential. In Eq. (27.6) the combination $\sigma A/l$ is called the *conductance* G of the wire, or

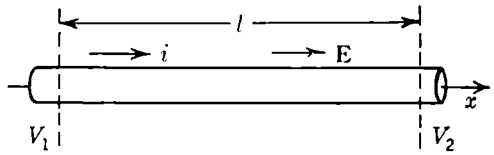


FIG. 27.2. Current flows from a high potential V_1 to a lower one V_2.

$$G = \frac{\sigma A}{l} \tag{27.7}$$

The reciprocal of the conductance G is known as the *resistance* R of the wire, and

$$R = \frac{l}{\sigma A} \tag{27.8}$$

Equation (27.6) can therefore be written also as

$$i = G(V_1 - V_2) = \frac{1}{R}(V_1 - V_2) \tag{27.9}$$

The reciprocal of the conductivity σ is named the *resistivity* ρ of a material. Hence

$$\rho = \frac{1}{\sigma} \tag{27.10}$$

$$R = \frac{\rho l}{A} \tag{27.11}$$

$$G = \frac{A}{\rho l} \tag{27.12}$$

It should be noted that Eqs. (27.4), (27.6), and (27.9) are all equivalent, and all are expressions of Ohm's law.

Units. It is necessary to define the units for the measurement of these new quantities. From the second part of Eq. (27.9), the unit of resistance would be one volt per ampere. This unit of resistance is called the *ohm*. Its reciprocal, the unit of conductance, is called the *mho*. From Eqs. (27.7) and (27.8), the unit of resistivity must be one ohm meter and the unit of conductivity one mho per meter. These units are, of course, all practical mks units. There are also, to be sure, esu for these quantities, but they are very little used. Sometimes a mixed unit of resistivity is encountered, the ohm centimeter. Table 27.1 gives some representative values of the resistivity ρ and the conductivity σ.

Table 27.1. *Values of* ρ, σ, *and* α

Substance	ρ ohm m	σ mhos/m	α per °C
Copper	1.72×10^{-8}	5.8×10^{7}	0.0039
Iron	10×10^{-8}	1.0×10^{7}	0.0050
Nichrome*	100×10^{-8}	0.1×10^{7}	0.0004
Carbon	$3{,}500 \times 10^{-8}$	2.9×10^{4}	−0.0005
Tungsten	5.5×10^{-8}	1.8×10^{7}	0.0045
Fused quartz	7.5×10^{17}	1.3×10^{-18}	

* Nichrome is an alloy of nickel, iron, and chromium often used for its high resistivity.

The resistivity ρ and conductivity σ vary with the temperature. This variation can be expressed empirically by

$$\rho = \rho_0(1 + \alpha T + \beta T^2 + \cdots) \tag{27.13}$$

where T is the temperature at which the resistivity has the value ρ and ρ_0 is the resistivity when $T = 0$. The quantity α is called the *temperature coefficient of resistivity*. For good conductors such as copper, α is approximately $1/273$ per degree centigrade. Values of α are given also in Table 27.1.

Coils of wire with a high resistivity are used as resistance elements, and they are called *resistors*. Resistors are often constructed of carbon in various forms and molded into plastic cylinders with copper leads for making connections.

27.3. Power Dissipation and Electromotive Force. When an electric current flows, charge is transported from a place of high potential to a place with a lower potential. Consequently energy is given up by the electrical system. This energy is dissipated in the wire and appears as heat. As the electrons flow, they collide with the relatively stationary ions of the conductor and transmit kinetic energy to them. The amount of energy transformed to heat per second, or the power dissipated, is, for example in the wire of Fig. 27.2, from Eq. (25 20),

$$P = \frac{dW}{dt} = \frac{d}{dt}[q(V_1 - V_2)] = i(V_1 - V_2) \tag{27.14}$$

If we solve Eq. (27.9) for $V_1 - V_2$, the power P becomes

$$P = \frac{i^2}{G} = i^2R \tag{27.15}$$

For a given current the power dissipated is proportional to the resistance. The resistance is thus a measure of the number and violence of the collisions between the electrons and the ions of the conductor.

Since the practical electrical units are mks units, in Eq. (27.15) i is measured in amperes, R in ohms, and P in watts. This relation is often employed to determine the mechanical equivalent of heat by an electrical method. If a coil of wire of resistance R is immersed in water and a current i passed through it, the temperature rise of the water in a given time can be measured and the mechanical or "electrical" equivalent of the heat gained by the water can be determined.

The energy relations of electric currents are often expressed in terms of the *electromotive force* (emf) $\mathcal{E}$. *The emf acting between two points a and b is defined as the work done when a unit charge is carried between the two points in question.* For the simple cases of steady currents that we have considered so far the emf is equal to the difference in potential between the two points, and consequently $\mathcal{E}$ is independent of the path chosen between a and b. In other cases, to be treated later, this is not true. In all cases, if W is the energy dissipated in the circuit between a and b, then

$$\mathcal{E} = \frac{dW}{dq} \tag{27.16}$$

Equation (27.14) can be expressed in terms of $\mathcal{E}$ since

$$P = \frac{dW}{dt} = \frac{dW}{dq}\frac{dq}{dt} = \mathcal{E}i \tag{27.17}$$

If Eqs. (27.15) and (27.17) are combined, we have

$$\mathcal{E} = Ri = \frac{i}{G} \tag{27.18}$$

This equation, rather than Eq. (27.4), is often called Ohm's law, and indeed the law was first stated for currents flowing in conducting wires.

Equation (27.4) expresses a more general relationship and applies whether or not wires are present. It will already have occurred to the student that the electromotive force is *not* a force; the term is a misnomer.

Electromotive forces can be produced by chemical actions in cells or by electric generators of various sorts. For the present we shall merely assume that such sources exist. If a wire of resistance R is connected to the two terminals of such a source, a current i flows whose magnitude is given by Eq. (27.18). Such an arrangement is the simplest electric circuit. Circuits are usually described by schematic diagrams, and the diagram for this simple case is shown in Fig. 27.3. The symbols used for the source of emf and for the resistor are the conventional ones.

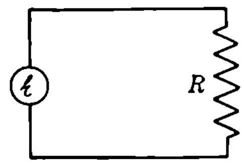


FIG. 27.3. A circuit containing a source of electromotive force $\mathcal{E}$ and a resistance R.

27.4. The Combination of Resistances. More complicated circuits can be formed, with several resistors. If the resistors are connected as indicated in Fig. 27.4, they are said to be connected in *series*, or in cascade. The same current flows through each resistor. The potential difference across the combination is the sum of the differences in potential across each one. The resistance R of the combination is consequently the sum of the individual resistances, thus,

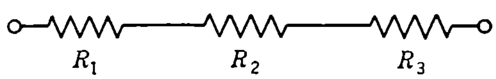


FIG. 27.4. Resistors connected in series.

$$R = R_1 + R_2 + R_3$$

or, in more general terms,

$$R = \sum_s R_s \qquad (27.19)$$

Resistors can also be connected in *parallel*, as indicated in Fig. 27.5. Here the potential difference across each resistor is the same, but the current through each one is different. The resistance R of the combination must be given by

$$Ri = R_1 i_1 = R_2 i_2 = R_3 i_3 \qquad (27.20)$$

where i_1, i_2, and i_3 are the currents through the individual resistors and

$$i = i_1 + i_2 + i_3 \qquad (27.21)$$

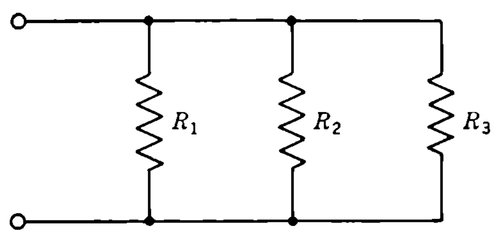


FIG. 27.5. Resistors connected in parallel.

If the four currents i, i_1, i_2, and i_3 are eliminated from Eq. (27.21) and the three equations (27.20), the desired expression for R can be found. It is simpler, however, to use the conductances G_s of the individual resistors. If V is the potential difference across the combination of resistors, then

$$i = VG = \sum_s VG_s = V\sum_s G_s$$

$$G = \sum_s G_s \tag{27.22}$$

The conductance of the parallel combination is therefore the sum of the conductances of the individual resistors. In terms of the resistances, Eq. (27.22) becomes

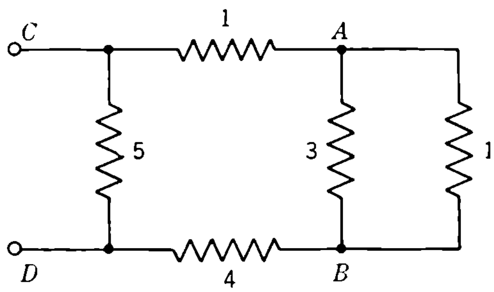


FIG. 27.6. Network of resistors with resistance values in ohms.

$$\frac{1}{R} = \sum_s \frac{1}{R_s} \tag{27.23}$$

For two resistors,

$$R = \frac{R_1R_2}{R_1 + R_2} \tag{27.24}$$

By the use of Eqs. (27.19) and (27.23) successively, the resistance of more complicated networks can be found. The resistance between the points A and B in Fig. 27.6 is found from Eq. (27.24) to be ¾ ohm. This ¾-ohm resistance is in series with resistances of 1 and 4 ohms, and hence, by Eq. (27.19), the total is 5¾ ohms. The total resistance R between C and D is thus 5¾ ohms in parallel with 5 ohms, or, by Eq. (27.24),

$$R = \frac{5 \times 5\frac{3}{4}}{5 + 5\frac{3}{4}} = \frac{115}{43} \text{ ohms}$$

27.5. Kirchhoff's Laws. In complex combinations of resistors and perhaps several sources of emf it is often desired to find the currents that flow through each resistor. The procedure to be followed in such cases can be formulated concisely in terms of two laws known as Kirchhoff's laws. Consider the simple network shown in Fig. 27.7. Let the currents through the resistors R_1 and R_2 be i_1 and i_2. The first of Kirchhoff's laws states that *the algebraic sum of all the currents flowing into any point of the network must be zero.* For a point such as C in Fig. 27.7 the law states the trivial fact that the current flowing into C from B is equal to the current flowing out of C through the resistor

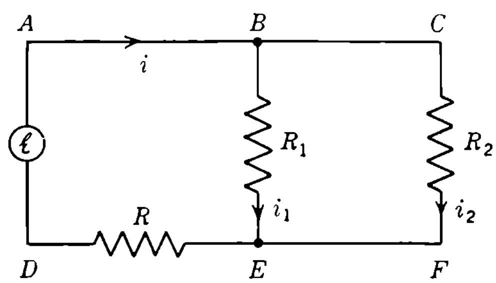


FIG. 27.7. Application of Kirchhoff's laws.

R_2. When the law is applied to point B, the result is

$$i = i_1 + i_2 \tag{27.25}$$

This is obviously true since charge does not accumulate at B. Kirchhoff's first law is thus another expression of the law of conservation of charge (Sec. 24.2).

The network of Fig. 27.7 is made up of loops. Thus one loop is $ABED$. This loop contains the resistor R_1 and the source of emf. Kirchhoff's second law is a generalization of Eq. (27.18) and states that, *around any loop, the sum of the emfs must be equal to the sum of the products of the resistances and the currents flowing through them.* Application of this law to the loop $ABED$ gives

$$\mathcal{E} = R_1 i_1 + Ri \tag{27.26}$$

Another equation can be obtained by applying the law to another loop, such as $BEFC$. This results in

$$R_1 i_1 - R_2 i_2 = 0 \tag{27.27}$$

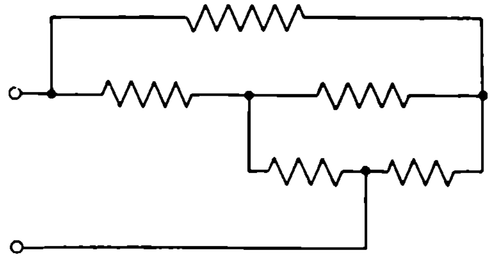


FIG. 27.8. Network of resistors.

Equations (27.25), (27.26), and (27.27) are three independent relations that can be used to determine the three unknown currents, i, i_1, and i_2. Another relation can be obtained from the loop $ACFD$, but this leads to no new result.

Kirchhoff's laws can be used to obtain the resistance of combinations of resistors that cannot be resolved into series and parallel combinations. Such a network is shown in Fig. 27.8. Assume currents i_1, i_2, . . . flowing through the resistors, and apply Kirchhoff's first law,

$$\sum_s i_s = 0 \tag{27.28}$$

to several points in the network. The second Kirchhoff law,

$$\sum_s \mathcal{E}_s = \sum_j R_j i_j \tag{27.29}$$

can be applied to as many loops as are needed to obtain a sufficient number of independent equations.

Worked Example. A current of 5 amp divides between two resistors connected in parallel. If the resistance values are 2 ohms and 3 ohms, what are the currents through each resistor? Let us call these currents i_2 and i_3, respectively. Kirchhoff's first law takes the form

$$i_2 + i_3 = 5 \text{ amp}$$

Application of the second Kirchhoff law to the loop formed by the two resistors results in

$$2i_2 - 3i_3 = 0$$

If the first equation is multiplied by 2 and the second subtracted from it, it is found that $i_3 = 2$ amp and hence $i_2 = 3$ amp.

27.6. The Wheatstone Bridge. An electric network that is commonly used to compare the resistance values of two resistors is known as a Wheatstone bridge. The arrangement is shown schematically in Fig. 27.9. Between points A and B a long wire of high resistance and length l is connected. Between the junction C of a standard resistor R_s with an unknown resistor R_x and a variable point of contact D on the wire, a sensitive current-detecting instrument or galvanometer is connected. The distance x from A to the variable contact is changed until a position is found in which no current flows through the galvanometer. In this condition the bridge is said to be balanced. The same current i_1 then flows through R_s and R_x, and a current i_2 flows through both parts of the wire. Kirchhoff's second law can then be applied to the circuits ACD and CBD. The circuits contain no emfs, and hence

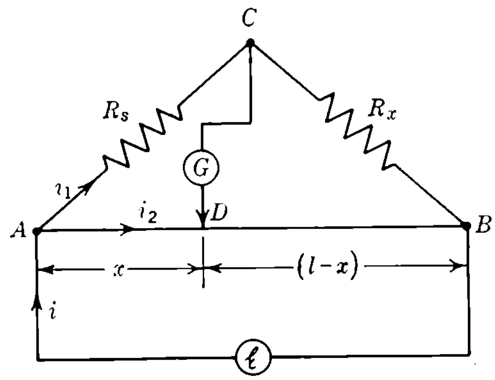


FIG. 27.9. A Wheatstone bridge.

$$i_1R_s - i_2xR/l = 0$$
$$i_1R_x - i_2(l - x)R/l = 0$$

where R is the resistance of the wire. These equations can be combined to give

$$\frac{R_x}{R_s} = \frac{l - x}{x} \tag{27.30}$$

The unknown resistance is thus determined in terms of the standard resistance and easily measurable lengths. It is fairly easy to achieve good accuracy in a measurement of the ratio of resistances by this method. The condition for balance [Eq. (27.30)] is independent of $\mathcal{E}$ and the currents, as well as independent of the properties of the detecting device.

Practical Wheatstone bridges take many forms. The wire may be replaced by sets of coils tapped at lengths to give resistance ratios of 1, 10, 1/10, and so forth, and the standard resistor R_s may be variable and calibrated. It should be noted that the network of Fig. 27.8 has the form of a bridge circuit.

***27.7. A Circuit Containing Resistance and Capacitance.** In Fig. 27.10 is shown a simple circuit containing both resistance and capacitance. If the capacitor C is initially uncharged and the switch S is closed, a current flows through the resistor to charge the capacitor. At any instant the emf must be equal to the sum of the potential difference across the capacitor and that across the resistance,

$$\mathcal{E} = \frac{q}{C} + Ri$$

where q is the momentary charge on the capacitor. The current i flows to the capacitor, and

$$i = \frac{dq}{dt}$$

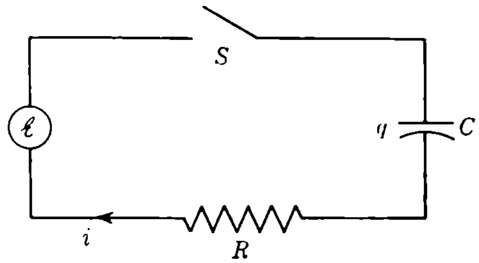


FIG. 27.10. A circuit containing resistance and capacitance.

Hence q is the solution of the differential equation

$$\mathcal{E} = \frac{q}{C} + R\frac{dq}{dt} \tag{27.31}$$

This equation can be solved by a separation of the variables. Transposition and division by R result in

$$\frac{dq}{dt} = \frac{\mathcal{E}}{R} - \frac{q}{RC} = \frac{\mathcal{E}C - q}{RC}$$

Hence

$$\frac{-dq}{\mathcal{E}C - q} = -\frac{dt}{RC}$$

which can be integrated at once. We obtain

$$\ln(\mathcal{E}C - q) = (-t/RC) + \ln A$$

where the constant of integration is written in the form ln A. This can be transformed to

$$\frac{\mathcal{E}C - q}{A} = e^{-t/RC} \tag{27.32}$$

To determine the parameter A it is necessary to know the value of q for one value of t. Suppose, for example, that $q = 0$ when $t = 0$; then

$$A = \mathcal{E}C$$

and

$$q = \mathcal{E}C(1 - e^{-t/RC}) \tag{27.33}$$

Equation (27.33) shows that q changes monotonically from the value 0 when $t = 0$ to the value $\mathcal{E}C$ for very large values of t. The variation is illustrated in Fig. 27.11.

The quantity RC in the exponent is known as the *time constant* of the circuit. At $t = RC$,

$$q = \mathcal{E}C\left(1 - \frac{1}{e}\right) = 0.631\mathcal{E}C \tag{27.34}$$

The time constant is a measure of the time taken for the capacitor to charge. During the time RC the charge reaches 63 per cent of its final value.

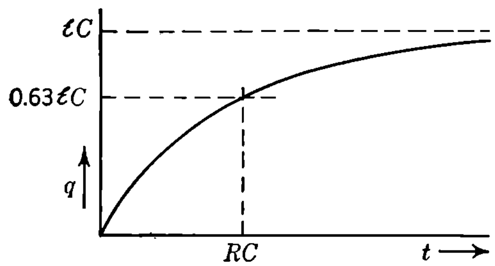


FIG. 27.11. The charge on the capacitor in Fig. 27.10.

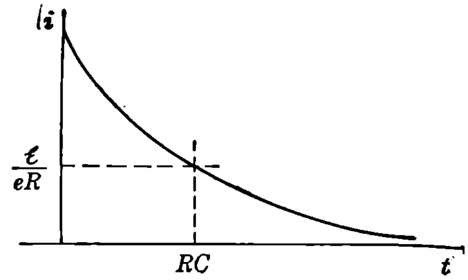


FIG. 27.12. The current in the circuit of Fig. 27.10.

The current as a function of time can be found by differentiation of Eq. (27.33), with the result

$$i = \frac{\mathcal{E}}{R} e^{-t/RC} \tag{27.35}$$

The current thus decays exponentially from the value $\mathcal{E}/R$ when $t = 0$. When $t = RC$, the current has $1/e$ of its initial value. Figure 27.12 shows the variation of i with t.

Worked Example. In the circuit of Fig. 27.10, the emf $\mathcal{E} = 100$ volts, $R = 10^4$ ohms, and $C = 1\ \mu$f. The switch S is closed at $t = 0$ with the capacitor uncharged. Find the initial value of the current in the circuit. From Eq. (27.35) at $t = 0$,

$$i = \frac{\mathcal{E}}{R} = \frac{100}{10^4} = 10^{-2} \text{ amp} = 10 \text{ ma}$$

The time constant is

$$RC = 10^4 \times 10^{-6} = 10^{-2} \text{ sec}$$

The current after 10^{-2} sec has decreased to

$$i = \frac{10^{-2}}{e} = 0.0037 \text{ amp} = 3.7 \text{ ma}$$

PROBLEMS

1. What is the resistance of 1 mile of copper wire 1 mm in diameter?

2. Prove that for two conductances in series $G = G_1G_2/(G_1 + G_2)$.

★3. The constant A in Eq. (27.32) is determined by the condition that $q = q_0$ at $t = 0$. What is q as a function of t?

4. What is the conductance of four resistors in parallel having resistances of 5, 10, 15, and 20 ohms, respectively?

5. It is desired to make a resistor of an iron element and a carbon element in series so that the combination has a zero temperature coefficient. What must be the ratio of the resistances of the two elements? Neglect the changes in volume of the elements.

6. Derive the temperature coefficient of resistance of a wire in terms of the temperature coefficients of resistivity and linear expansion.

7. What are the current and the resistance of a 100-watt lamp bulb operating at 115 volts?

8. A family uses 200 liters of hot water, at a temperature of 50°C, in a day. Water is heated by electricity and has a temperature of 15°C in the supply mains. If half the heat energy is lost, how much electric energy would be required? What is the cost at 5 cents per kilowatt hour?

9. If each of the resistors in Fig. 27.8 has the value of 1 ohm, what is the resistance across the terminals?

10. A 1-μf capacitor is charged at the rate of 10^6 volts/sec. What is the current flowing into the capacitor?

***11.** Complete the solution of Eqs. (27.25) to (27.27).

***12.** Show by using Kirchhoff's laws that the current i_g through the galvanometer in the Wheatstone bridge when it is not balanced is

$$i_g = \frac{(R_x Q - R_s P)\mathcal{E}}{R_s R_x P + R_x PQ + R_s PQ + R_s R_x Q + R_g(R_s + R_x)(P + Q)}$$

where R_g is the resistance of the galvanometer, $P = R - Q$ and $Q = Rx/l$.

13. Calculate the mobility of electrons in copper.

14. What resistance R must be connected between the terminals C and D in Fig. 27.13 in order to have the resistance between terminals A and B equal to R?

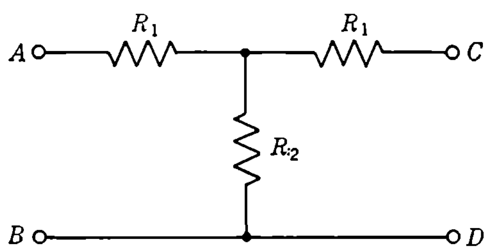


FIG. 27.13. Circuit for Problem 14.

CHAPTER 28
CHEMICAL AND THERMAL ELECTROMOTIVE FORCES

28.1. Conduction in Electrolytes. In metals an electric current is the result of the motion of electrons; in liquids and solutions the conduction mechanism is different. In water, for example, the electrons are tightly bound to the atoms and cannot move about. Some of the water molecules are *dissociated* according to the equation

$$H_2O \rightleftharpoons H^+ + OH^- \tag{28.1}$$

By H^+ is meant a hydrogen atom with one extra positive charge (one electron missing) or a bare hydrogen nucleus; OH^- is a combination of an oxygen and a hydrogen atom with a net negative charge of one electron. These charged pieces of molecules are known as *ions* (Greek *ion*, wandering). The ions can move through the liquid and give rise to an electric current. This current can be expressed by an equation similar to Eq. (27.3), but with two terms, one for the positive and the other for the negative charges. Thus

$$\frac{i}{A} = J = n^+e^+v^+ + n^-e^-v^- \tag{28.2}$$

where A is the area, n^+ and n^- are the numbers of positive and negative ions per unit volume, v^+ and v^- their velocities, and e^+ and e^- their charges. If the liquid has no net charge, $n^+ = n^-$. The number of ions per unit volume depends on the temperature, since Eq. (28.1) is a reversible reaction. The conductivity σ is defined as the ratio of the current density to the electric field, and

$$\sigma = n^+e^+m^+ + n^-e^-m^- \tag{28.3}$$

where m^+ and m^- are the mobilities of the positive and negative ions,

$$m^+ = \frac{v^+}{E} \qquad m^- = \frac{v^-}{E} \tag{28.4}$$

The mobilities have values that are characteristic of the kind of ions but that depend also upon temperature. For example, a sodium ion in water has a mobility of 5.3×10^{-4} cm/sec per volt/cm or 5.3×10^{-4} cm^2 sec^{-1} volt^{-1} at a temperature of 25°C.

Other liquids also dissociate into ions and therefore are conductors of electricity. For example, ethyl alcohol splits up according to

$$C_2H_5OH \rightleftharpoons (C_2H_5)^+ + OH^- \tag{28.5}$$

Liquids that do not dissociate are insulators, *e.g.*, most oils. Small amounts of water dissolved in the oil do dissociate to some extent and produce a small conductivity. It is therefore important to keep insulating oils pure and dry.

If a salt is dissolved in water, the molecules of the salt also dissociate. Thus sodium chloride in solution becomes almost entirely split up into sodium and chlorine ions,

$$NaCl \rightleftharpoons Na^+ + Cl^- \tag{28.6}$$

Some salts, in bivalent compounds, produce doubly charged ions, for example,

$$Cu(NO_3)_2 \rightleftharpoons Cu^{++} + 2(NO_3^-) \tag{28.7}$$

where the two plus signs indicate that the copper ion has two electrons less than a normal copper atom. Another example is

$$CuSO_4 \rightleftharpoons Cu^{++} + SO_4^{--} \tag{28.8}$$

where both ions are doubly charged. A solution containing ions of a salt is called an *electrolyte* (Greek *lyein*, to dissolve). Acids and bases behave similarly in water solutions.

Salts dissociate to a much greater extent than does water, and electrolytes are better conductors of electricity than pure liquids. The values of n^+ and n^- in Eq. (28.3) are larger, but the mobilities of the ions of salts and liquids do not differ greatly. As an example, a normal solution of sodium chloride has a specific conductance of 7.5 mhos/m at 18°C. Such a solution contains 1.2×10^{27} ions per cubic meter. In copper there are approximately 70 times as many carriers of electricity per unit volume, but the conductance is nearly 10^7 times larger. The ionic mobilities are therefore much smaller than the mobility of an electron in copper.

One reason for the ease of dissociation of salts in water is the high value of the specific inductive capacity of water. We have shown [Eq. (26.15)] that the force between charges is inversely proportional to the specific inductive capacity of the surrounding medium. The electric forces that hold a salt molecule together are therefore weakened by the water until they are unable to maintain the molecular stability.

28.2. Electrolysis. In order to produce a current in an electrolyte it is necessary, of course, to apply an electric field. This is most conveniently done by immersing in the electrolyte two pieces of metal connected to the terminals of some external source of an emf. These metal pieces are called *electrodes;* the positive electrode is called the *anode* (Greek *ana*, upwards; *hodos*, path), the negative one the *cathode* (Greek *kata*, downward). In general, a pair of electrodes produces, not a uniform electric field, but a field that varies from point to point within the electrolyte.

Suppose that two electrodes of an inert metal such as platinum are immersed in water and connected to a source of emf. The hydrogen ions in the water are positively charged and hence flow to the negatively charged cathode. Upon reaching the cathode, each ion is neutralized and becomes a hydrogen atom. Two such atoms combine, and hydrogen gas is released and bubbles off. The hydroxyl ions, OH^-, flow to the anode, and each one loses an electron there. At the anode the following reaction takes place,

$$2OH^- \rightarrow 2e + H_2O + O \tag{28.9}$$

where e denotes an electron. When two oxygen atoms combine, one molecule of oxygen is formed. Thus four hydroxyl ions must reach the anode for each oxygen molecule formed. On the other hand, one hydrogen molecule is formed for every two hydrogen ions. Two volumes of hydrogen gas are therefore evolved for each volume of oxygen. The net result of the passage of an electric current through water is thus the decomposition of the water into hydrogen and oxygen. This process is known as *electrolysis*. Electric energy is transformed into chemical energy in electrolysis.

If other materials are used as electrodes and electrolyte, various other chemical reactions can be produced. If a copper anode is employed in a solution of copper sulfate, copper will go into the solution at the anode and be deposited out on the cathode. This process of *electroplating* has many commercial applications and can be used to produce many metallic finishes and coatings.

The amount of material transported in any electrolytic process can be calculated. Since the charge on an ion of a substance has a definite value, the amount of material transported is proportional to the charge Q transported. The mass m deposited is in fact equal to the mass of one ion, which is A/N_0 as shown in Sec. 20.7, times the number of ions Q/ev where v is the valence of the ion. Hence

$$m = \frac{AQ}{N_0 ev} \tag{28.10}$$

Equation (28.10) is often written

$$m = ZQ \tag{28.11}$$

where Z is called the *electrochemical equivalent* of the substance and $Z = A/vN_0e$. Thus, for hydrogen,

$$Z = \frac{1.008}{1 \times 6.0 \times 10^{23} \times 1.6 \times 10^{-19}} \text{ or } 10^{-5} \text{ gm/coulomb}$$

Equations (28.10) and (28.11) are known as *Faraday's laws of electrolysis,* after their discoverer. The quantity of electricity N_0e is a universal constant, and its value can be determined accurately from electrolytic experiments. The constant is called the *faraday* and is equal to 96,496 coulombs. Once the value of the faraday is known, currents can be determined accurately from the weight of the electrolytic deposit of a metal, usually silver. An instrument for this purpose is called a *voltameter*.

28.3. Cells. In electrolysis, electric energy is transformed into chemical energy. Inversely, chemical energy can be utilized to produce electric energy. If electrodes of any two metals are immersed in an electrolyte, it is found that a difference of potential exists between the metals and a current flows in a wire connected between the electrodes. Such a source of emf is called a *cell.* The flow of current through the wire is accompanied by chemical changes within the cell, the nature of these changes depending, of course, on the composition of the electrodes and the electrolyte.

If a copper electrode is immersed in water, some of the copper goes into solution as ions, and electrons are left on the electrode. The reaction is

$$\mathrm{Cu} \rightleftharpoons \mathrm{Cu}^{++} + 2e \tag{28.12}$$

The copper electrode thus becomes negatively charged and assumes a potential lower than that of the solution. An equilibrium state is reached in which the number of Cu^{++} ions formed per second is equal to the number per second attracted to the electrode and precipitated there. If additional Cu^{++} ions are present in the solution from some other source, the equilibrium is shifted and the potential of the electrode altered. The potential depends also upon the temperature and upon the electrode material.

If two electrodes of different metals are present, the potential difference between the metals is the difference between two metal-solution potentials, and a cell is formed. The potentials may be represented schematically in Fig. 28.1a. If the electrodes are connected by a resistance, a current flows and there will be an iR drop across the electrolyte in the cell, as schematically shown in Fig. 28.1b. If the electrodes are connected by a wire of negligible resistance, so that they are at the same potential,

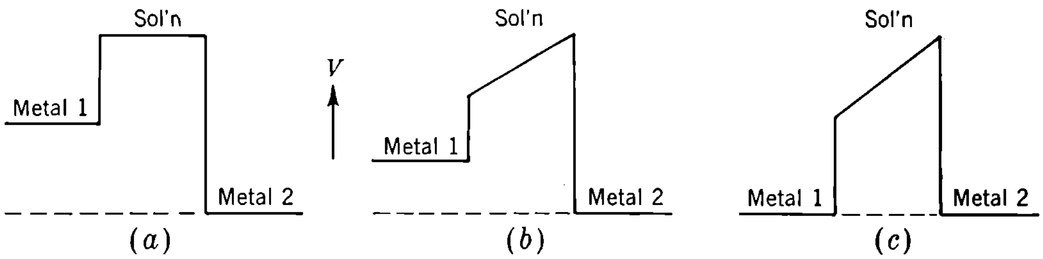


FIG. 28.1. The potential relations within a cell; (a) open circuit, (b) circuit closed through a resistance, (c) cell short-circuited.

a larger current flows and the potentials are shown in Fig. 28.1c. A cell therefore has a net emf and an internal resistance caused by the resistance of the electrolyte. The potential difference between the terminals of the cell depends, therefore, upon the current flowing through it.

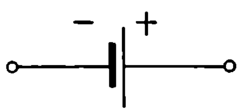


FIG. 28.2. The symbol for a cell.

In electrolysis the two electrodes are immersed in a solution and hence a potential difference exists between them; the arrangement is called an *electrolytic cell*. To produce the electrolysis an emf from an external source must be applied which is greater than, and in the opposite direction to, the emf of the electrolytic cell. The electrolytic cell is said to produce a *counter* or *back emf*.

A cell is represented in circuit diagrams by the symbol shown in Fig. 28.2. The long line is conventionally the positive terminal. Outside the cell, current flows away from the positive and toward the negative terminal; within the cell, current flows from the negative to the positive electrode. The power that can be drawn from a cell can be calculated for the circuit in Fig. 28.3. The internal resistance of the cell is r and the external load resistor has a resistance of R. The current i in the circuit is

$$i = \frac{\mathcal{E}}{r + R} \tag{28.13}$$

The power P dissipated in the load resistor is

$$P = i^2R = \frac{\mathcal{E}^2R}{(r + R)^2} \tag{28.14}$$

When $R = r$, the power has a maximum value given by

$$P_{max} = \frac{\mathcal{E}^2}{4R} = \frac{\mathcal{E}^2}{4r} \tag{28.15}$$

Equation (28.15) applies, of course, to all sources of emf. The value of P_{max} is called the *available power*. Groups of cells are often employed; such a group is called a *battery*. For a higher emf, cells are connected in series; for a larger current capacity, they are connected in parallel.

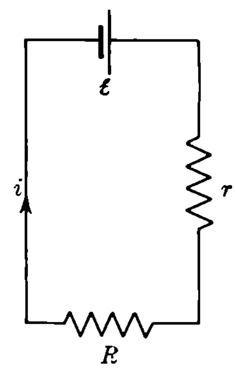


FIG. 28.3. A cell with internal resistance r and external load R.

***28.4. Practical Forms of Cells.** In practical cells the chemical reactions that take place are much more complicated than those just described. For example, some gas is often released at one or both electrodes. Many secondary reactions also occur. The net result is that the emf of the cell decreases after current has been flowing, but the cell recovers again when the current is stopped. These secondary actions are known as *polarization*. The phenomenon has nothing to do, of course, with the polarization of a dielectric. Additional substances are usually added to the cell to minimize the polarization, and the chemical effects are made still more complicated.

The *Daniell cell* (Fig. 28.4) is one of the simplest practical cells. It consists of a copper anode immersed in saturated copper sulfate solution and a zinc cathode in a solution of zinc sulfate. The two solutions have different specific gravities and mix only by diffusion. In some Daniell cells the two solutions are separated by a cup of porous ceramic material. In a freshly prepared cell the zinc sulfate solution is dilute, and the emf is about 1.06 volts. With the flow of current the zinc cathode gradually dissolves, and the solution becomes more concentrated. The emf therefore gradually decreases with age.

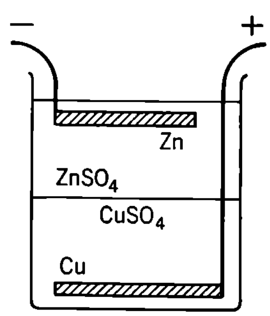


FIG. 28.4. The Daniell cell.

The *dry cell* is the most common form of portable cell. The positive electrode is a carbon rod and the negative electrode a zinc can, which encloses the carbon rod. The electrolyte is a paste containing zinc chloride, ammonium chloride, and manganese dioxide. The whole is sealed at the top with a layer of wax. As the cell is discharged, the zinc cathode dissolves and zinc ions are formed. At the positive terminal the ammonium ions react with the manganese dioxide according to

$$2(NH_4^+) + 2e + MnO_2 \rightarrow 2NH_3 + MnO + H_2O$$

The cell is worn out when the manganese dioxide becomes exhausted. The emf of a new cell is between 1.5 and 1.6 volts.

The *lead storage cell* can be recharged by passing a current through it and reversing the chemical reactions. In the fully charged condition the anode of this cell is a lead plate coated with the lead peroxide, PbO_2, and the cathode is metallic lead. The electrolyte is sulfuric acid. The emf of the cell is 2.1 volts. As the cell discharges, the anode reaction is

$$PbO_2 + SO_4^{--} + 4H^+ + 2e \rightarrow PbSO_4 + 2H_2O$$

The lead sulfate formed adheres to the plate. At the cathode the reaction is

$$Pb + SO_4^{--} \rightarrow PbSO_4 + 2e$$

and the lead sulfate again is deposited. Since sulfate ions are removed from the electrolyte, its specific gravity falls as the cell is discharged. When the cell is

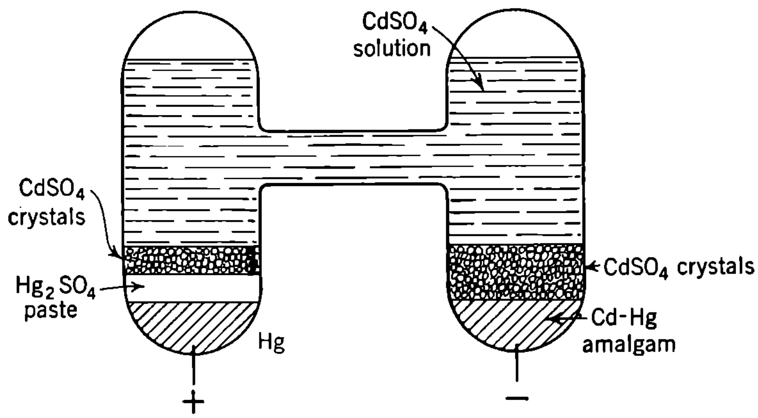


Fig. 28.5. The Weston standard cell.

charged, the reactions are effectively reversed and the metallic lead and lead peroxide are restored. Hydrogen gas is also evolved. A lead storage cell can be made with a very low internal resistance, and hence large currents can be drawn from it.

The *Weston standard cell* is useful for measurements since the emf remains unchanged over long periods of time. The construction is shown schematically in Fig. 28.5. The emf is 1.0183 volts at 20°C.

28.5. The Potentiometer. Since a cell has internal resistance, the potential difference between the terminals depends on the current flowing. In order to measure the emf of the cell a method must be employed that does not involve a flow of current. Such a device is known as a potentiometer. A schematic circuit diagram is given in Fig. 28.6. The cell C produces a steady current i through the resistance R and the resistance wire of resistance r per unit length. This current can be adjusted to a convenient value by the variable resistor R'. The potential of the point b is higher than that at a because of this current, and

$$V_b - V_a = Ri + rxi \tag{28.16}$$

If the double-pole double-throw switch S is used to connect cell A in the circuit, the distance x can be varied until $V_b - V_a$ is equal to the emf of cell A and no current flows through the sensitive galvanometer G. Let this value of x be x_A. The switch can now be thrown the other way, and the distance x_B, such that again no current flows through G, can be determined. The ratio of the emfs is then equal to the ratio of the resistances,

$$\frac{\mathcal{E}_A}{\mathcal{E}_B} = \frac{R + rx_A}{R + rx_B} \qquad (28.17)$$

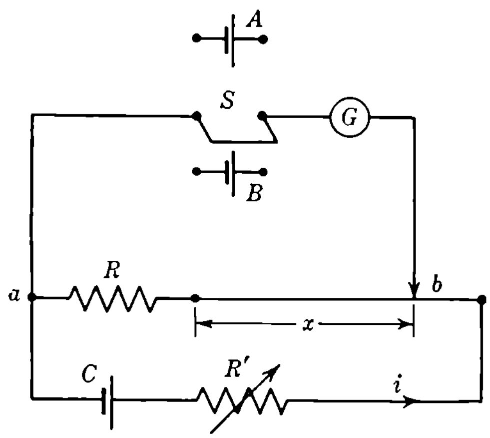


FIG. 28.6. The potentiometer.

If cell A is a standard cell of known emf, then the emf of cell B can be easily found.

In a practical potentiometer, C might be a lead storage battery of two cells. The wire can be calibrated to read directly in volts. Suppose the current i is chosen to be 100 ma, R is equal to 10 ohms, and the wire resistance also is 10 ohms. The range of the potentiometer is then from 1.0 to 2.0 volts. The current i is adjusted until no current passes through the galvanometer when the standard cell is connected and the slide-wire contact is set to the proper reading. The emf of an unknown cell can then be read directly from the point of balance on the wire.

Since a potentiometer "divides" a potential difference into adjustable proportions, a variable resistor with contacts at each end and a variable contact between the ends is also called a potentiometer, or potential divider. Such a device is indicated in circuit diagrams by the symbol in Fig. 28.7.

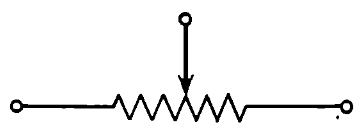


FIG. 28.7. Symbol for a potentiometer or potential divider.

★28.6. The Thomson and Peltier Emfs. When current passes through a resistor, heat is developed *irreversibly* at the rate i^2R. Heat can also be converted to electric energy by a reversible process, and this process gives rise to a *thermal emf*. If a metal bar is heated at one end, the temperature, and therefore the pressure of the electron gas within the metal, is increased at one end over that at the other. A difference in electron concentration results between the two ends of the bar of such an amount that the electrostatic field produced is just counterbalanced by the tendency for diffusion of electrons from one end of the bar to the other. The difference in temperature thus results in an emf called the *Thomson emf*. Thomson emfs are of the order of magnitude of millivolts.

If two dissimilar metals are joined, there is a diffusion of electrons from one metal to the other until a field is established of sufficient magnitude to maintain equilibrium. The junction is therefore the seat of an emf called a *Peltier emf*. This emf is also of the order of millivolts. Peltier emfs depend upon the two metals and upon the temperature of the junction. The existence of a Peltier emf can be demonstrated by the circuit of Fig. 28.8. If the current and the Peltier emf are in opposite directions, heat is developed at the junction in excess of the normal i^2R heating. If the current and emf are in the same direction, the heating is less than normal. With the metals antimony and bismuth, for which the Peltier effect is unusually large, an actual cooling of the junction may take place. The cooling effect is small and not of direct practical importance.

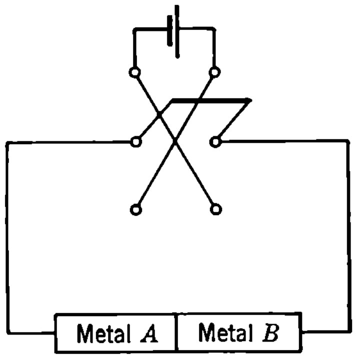


FIG. 28.8. Circuit to show the Peltier emf.

***28.7. Thermocouples.** If two dissimilar metals A and B form a closed circuit as in Fig. 28.9, with the junctions at different temperatures, the Peltier emfs at the two junctions are different. Since there must be a temperature gradient in each metal, Thomson emfs also exist. The net emf does not vanish, and consequently a current is set up in the circuit. This effect was discovered by Seebeck (1770–1831), and the net emf is known as a *Seebeck emf*. The Seebeck emf $\mathcal{E}$ is found by adding the Peltier and Thomson emfs around the circuit and thus obtaining

$$\mathcal{E} = P_1 + T_A - P_2 - T_B \tag{28.18}$$

A current flows in the circuit formed. This can be easily measured. The current i has the value $\mathcal{E}/R$, where R is the total resistance of the circuit. Such a device is known as a *thermocouple*.

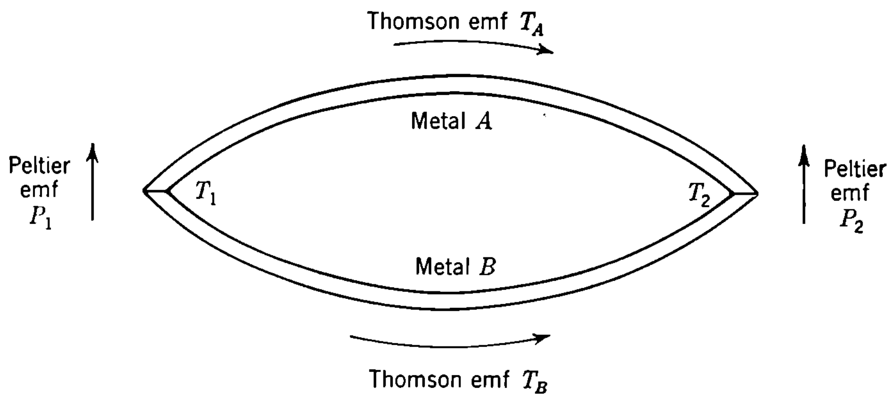


FIG. 28.9. The Seebeck emf in a thermocouple.

A thermocouple is an extremely important device for the measurement of temperatures. It is evident from Eq. (28.18) that an intermediate metal can be introduced in the thermocouple circuit, and provided that both new junctions are at the same temperature the Seebeck emf is unaltered. Consequently the current-measuring device can be inserted in the circuit without a change of $\mathcal{E}$, as in Fig. 28.10. If the temperature T_1 is fixed by immersing one junction in melting ice and the thermocouple calibrated, the temperature T_2 at the other junction can be determined. Since the current in the thermocouple circuit depends upon the resistance as well as upon the temperature difference of the junctions, a potentiometer is often used to measure $\mathcal{E}$.

Thermocouples can be connected in series to increase the emf generated. A group of them so connected is called a *thermopile*. A *thermopile* is used as a sensitive detector of radiant energy. A receiver R (Fig. 28.11) has one set of junctions attached to it in close thermal contact but electrically insulated from it. The other junctions are attached to the support S. The incident radiation, raising the temperature of R over that of S, produces an emf proportional to the energy absorbed.

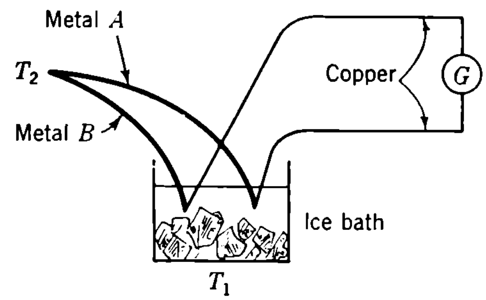


FIG. 28.10. Circuit for measurement of temperature with a thermocouple.

The thermoelectric force can be well represented, as a function of the absolute temperatures of the two junctions, by a second-degree equation,

$$\mathcal{E} = a(T_2 - T_1) + \tfrac{1}{2}b(T_2^2 - T_1^2) \tag{28.19}$$

If T_1 is kept constant, a plot of $\mathcal{E}$ as a function of T_2 is a parabola. The emf is zero for

$$T_2 = T_1 \qquad \text{and} \qquad T_2 = -\frac{2a}{b} - T_1 \tag{28.20}$$

The second of these temperatures is called the *inversion temperature* T_i. The temperature at which $d\mathcal{E}/dT_2$ is zero is called the *neutral temperature* T_n, and

$$T_n = -\frac{a}{b} \tag{28.21}$$

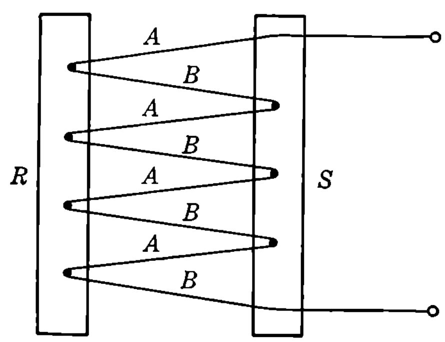


FIG. 28.11. Schematic diagram of a thermopile.

The derivative $d\mathcal{E}/dT_2$ is called the *thermoelectric power*. The variation of $\mathcal{E}$ with temperature is indicated in Fig. 28.12.

Table 28.1 gives the values of a and b relative to lead for several metals. These values are not very reliable, since they depend sensitively upon the impurities in the metals and the state of cold working. It should be noted that the emf of metal A relative to that of metal B is the difference between the emf of A relative to lead and that of B relative to lead.

Worked Example. Find the thermoelectric power and the neutral temperature of a thermocouple composed of antimony and bismuth. From Table 28.1, the thermoelectric power is

$$\frac{d\mathcal{E}}{dT_2} = [36 - (-74)] + (0.14 - 0.032)T_2$$
$$= 110 + 0.11\ T_2\ \mu\text{volts/deg C}$$

The inversion temperature is

$$T_i = -2\frac{110}{0.11} - T_1 = -2000°\text{C} - T_1$$

Such a temperature cannot be reached for any T_1.

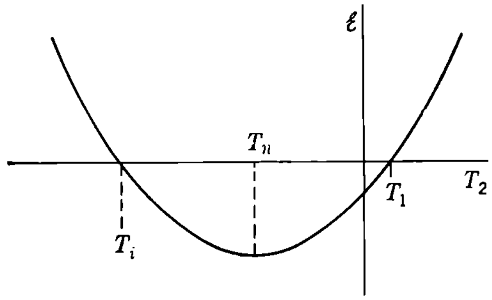


FIG. 28.12. The variation of the Seebeck emf with temperature.

Table 28.1. *Values of the Thermoelectric Constants a and b*
(in μvolts/deg C and μvolts deg C^{-2} relative to lead. When ℰ is positive, the direction of the current is from metal A to metal B at the hot junction)

Substance	a	b
Antimony	+36	+0.14
Bismuth	−74	+0.032
Constantan (60% Cu, 40% Ni)	−38	−0.089
Copper	+ 2.7	+0.0079
Iron	+17	−0.030
Nickel	−19	−3.0
Platinum	− 3.0	−3.25

PROBLEMS

1. From the conductivity of a normal solution of sodium chloride given in Sec. 28.1, compute the average mobility of an ion of the electrolyte.

2. A steel sheet of 4 ft² total area is to be electroplated with copper to a thickness of 0.005 in. If a current source of 100 amp is used, how long will the plating take?

3. The *legal* definition of one ampere is that "current, which, when passed through a solution of silver nitrate in water . . . deposits silver at the rate of 0.001118 gm/sec." Compute from this the value of the faraday.

4. A dry cell of emf. 1.60 volts causes a current of 1.45 amp to flow through a resistor of 1.10 ohms. What is the internal resistance of the cell? What is the maximum power that can be taken from it?

5. Two lead storage cells, each of emf 2.10 volts and internal resistance 0.10 ohm, are connected in parallel to the terminals of a resistor of 5 ohms resistance. What is the current in the resistor? How much power is dissipated?

6. Hydrogen is sometimes manufactured by the electrolysis of water. The *counter* emf of the electrolytic chamber is 1.5 volts, and the internal resistance is 0.1 ohm. What is the rate of generation of hydrogen if a power source having an emf of 1.8 volts and an internal resistance of 0.05 ohm is used?

7. With a potentiometer the emf of a cell is found to be 1.57 volts. When a voltmeter is connected to the terminals of the cell, it reads 1.56 volts. When a 5-ohm resistor is connected to the cell, the current is 0.31 amp. What is the resistance of the voltmeter?

★8. What is the inversion temperature for a thermocouple composed of iron and nickel wires? What is the value of $\mathcal{E}$ for a junction temperature difference of 100°C? Take $T_1 = 0°C$.

9. A dry cell momentarily short-circuited causes a current of 30 amp. The potential difference measured with a high-resistance voltmeter is 1.5 volts. What is the internal resistance?

10. An automobile storage battery of 6.3 volts can furnish a maximum power of 1 kw. What is the internal resistance?

11. A storage battery is charged at 5 amp for 20 hr. If 80 per cent of this charge can be utilized, how much silver can be plated with the battery?

12. A cell with an emf of 1.50 volt and an internal resistance of 1 ohm is connected in parallel with a cell of 1.60 volts emf and internal resistance 1.5 ohm. The combination is connected to a 5-ohm resistor. What is the current through each cell and through the resistor?

CHAPTER 29

MAGNETIC FORCES ON CURRENTS ELECTRICAL INSTRUMENTS

29.1. The Magnetic Field of Induction.[1] In addition to the electrostatic force between charges, it is found that a second force exists when the charges are in motion. This force is called a *magnetic* force. Two current-carrying coils attract or repel each other, depending on their relative orientation and upon the strengths of the currents and the number of turns in each coil. Although originally it was thought that magnetic forces were associated only with pieces of lodestone and other permanent magnets, it has since been established that the peculiar behavior of lodestone is the result of the electric currents arising from electronic motion within the material. As in the electrostatic case, it is convenient to separate the effect of one moving charge upon another into two parts. One moving charge is said to produce a *magnetic field*, and this magnetic field produces a force on the second moving charge. In the present chapter we shall confine our attention to the second half of the problem and shall investigate the forces on moving charges produced by a magnetic field. In Chap. 30 we shall treat the production of magnetic fields.

The field that acts upon moving charges is called the field of *magnetic induction*. It is analogous to the electric field **E**, which is a measure of the electrostatic force on a charge. The magnetic induction is a vector quantity, and the induction field can be represented by lines of induction similar to the lines of **E**. The strength of the magnetic induction can be *defined* by the force that is exerted on a charge. If the velocity **v** of the charge is in the x direction and the magnetic induction **B** is in the y direction, then a force **F** in the z direction is produced. The magnitudes of these vectors are related by the equation

$$F = qvB \tag{29.1}$$

where q is the electric charge. It is important to note the directions of the three vectors shown in Fig. 29.1. The force is not in the direction of the field, as in the electrostatic case, but at right angles both to the field and to the velocity of the charge. The positive direction of **F** is the direction in which a right-handed screw would advance if it were rotated from the vector **v** toward **B**. This apparently complicated convention

[1]At this point it is assumed that the student has a general familiarity with some method of producing a magnetic field, for example a bar magnet.

is necessary since **F** depends on both **v** and **B**. If **v** is reversed, **F** is reversed also. The direction of **B** does not define the direction of **F**, but it does define the plane in which **F** must lie.

If the velocity **v** of the charged particle is not perpendicular to **B**, as in Fig. 29.2, then Eq. (29.1) is written in terms of the component of **v** in the x direction. If ϕ is the angle between **v** and **B**, then the force becomes

$$F = q(v \sin \phi)B \qquad (29.2)$$

which reduces to Eq. (29.1) when $\phi = \pi/2$.

The integral of the normal component of **B** over a surface,

$$\Phi = \int B \cos \theta \, dA \qquad (29.3)$$

where θ is the angle between the normal to dA and **B**, is called the *magnetic flux of induction*. The magnetic flux is thus similar to the electric flux (see Sec. 25.4 and Fig. 25.5.

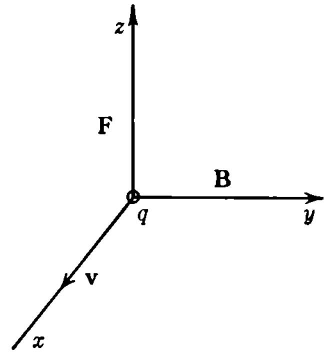


FIG. 29.1. The magnetic force **F** on a moving charge.

Equation (29.1) serves to define not only the induction **B** but also the units in which **B** is measured. In the practical mks system, **F** is measured in newtons, q in coulombs, and **v** in meters per second. The units of **B** are then newton sec per coulomb meter. A cgs system of units is also in use, and Eq. (29.1) is valid for this case also. This system of units is called the *electromagnetic system* of units (emu). In a cgs system, **F** is measured in dynes, and **v** in centimeters per second. The electromagnetic unit of charge q is called the *abcoulomb*, and

$$1 \text{ abcoulomb} = 10 \text{ coulombs} \qquad (29.4)$$

The emu of **B** is then one dyne sec per abcoulomb cm and is called one *gauss*. The units of the magnetic flux Φ are given special names. In the practical system the unit of magnetic flux is called the *weber* after W. E. Weber (1804–1891), a German physicist.

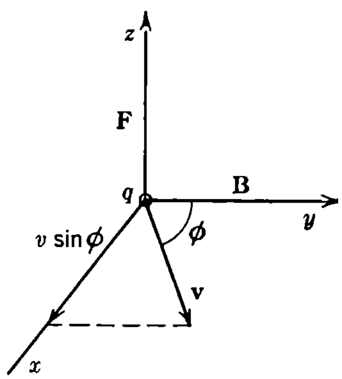


FIG. 29.2. The magnetic force when **v** is not perpendicular to **B**.

Using the weber as the primary unit, it is customary to derive a unit of induction which is equivalent to a newton second per coulomb meter. Since induction is flux per unit area, we obtain for this unit of B the *weber* *per square meter.* This is indeed the practical unit commonly employed. In the electromagnetic system, the unit of flux is called the *maxwell* named after J. C. Maxwell (1831–1879), one of the great mathematical physicists. It follows from the definition that

$$1 \text{ maxwell/cm}^2 = 1 \text{ gauss} = 10^{-4} \text{ weber/m}^2 \qquad (29.5)$$

Since the ratio of 1 gauss to 1 weber/m^2 is a power of 10, the gauss is sometimes used as an mks subunit.

29.2. The Motion of a Charged Particle in a Magnetic Field. If a particle of charge q moves in a uniform field **B**, perpendicular to the

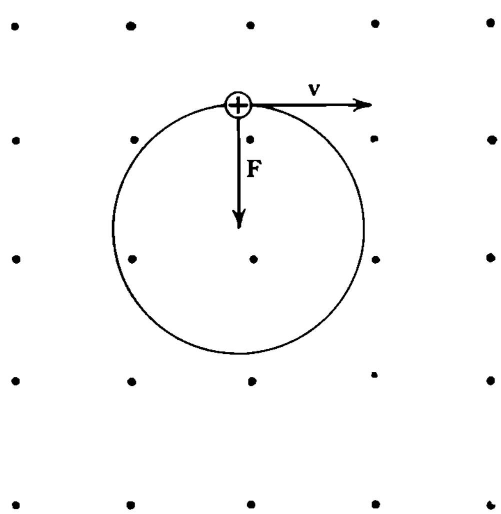


FIG. 29.3. A positive particle in a field outward from the paper.

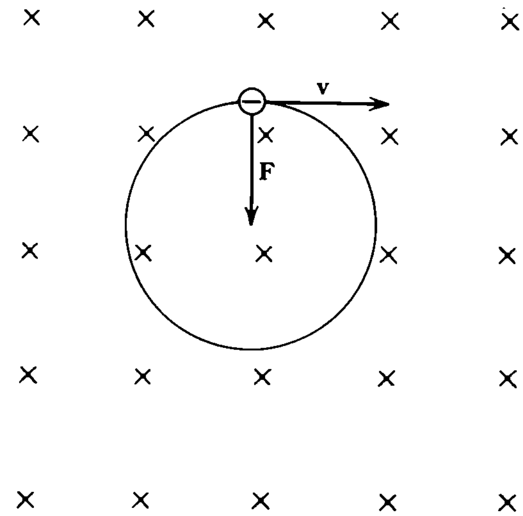


FIG. 29.4. A negative particle in a field inward to the paper.

velocity **v**, the force is also perpendicular to **v** and constant in magnitude. It produces an acceleration constant in magnitude and perpendicular to **v**, and we have seen in Sec. 5.9 that this is characteristic of uniform circular motion. The path of the particle is therefore a circle, and the magnetic force is a centripetal force. If the field is outward from the paper, as in Fig. 29.3, then for a positively charged particle **F** and **v** have the directions indicated. The field is represented in the figure by the series of dots. A magnetic field in the opposite direction is conventionally indicated by crosses (×). These conventions can be recalled by thinking of the dots as the heads of arrows traveling with the field and the crosses as the feathered tails of the field arrows. The motion of a negatively charged particle with the field **B** into the paper is shown in Fig. 29.4.

If the mass m times the centripetal acceleration is equated to the force of Eq. (29.1), there results

$$\frac{mv^2}{r} = qvB \tag{29.6}$$

where r is the radius of the circle. The radius r is thus proportional to the *momentum* of the particle and inversely proportional to B, since

$$r = \frac{mv}{qB} \tag{29.7}$$

If the velocity is not perpendicular to the field but makes an angle ϕ with it, then the particle moves in a helix advancing with the velocity $v \cos \phi$ in the direction of the field, and the radius of a turn equal to

$$r = \frac{mv \sin \phi}{qB}$$

29.3. The Measurement of Charge-to-mass Ratio. By measurement of the bending of the paths of charged particles in electric and magnetic

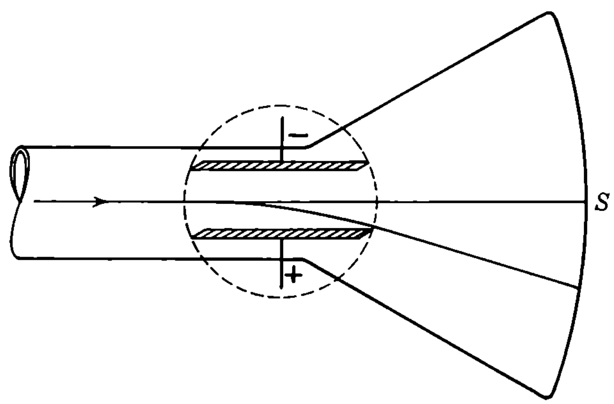


FIG. 29.5. J. J. Thomson's measurement of q/m.

fields the value of the ratio q/m can be found. This experiment was first performed by J. J. Thomson.[1] In the evacuated tube shown in Fig. 29.5, electrons were passed through an electric field between the parallel plates and a magnetic field perpendicular to the paper in the area enclosed by the dashed circle. By a proper adjustment of the fields a beam of electrons could be made to pass through the tube and fall upon a fluorescent screen S without deflection. Since the electric and magnetic forces balance, we can write, in practical units,

$$qE = qvB$$

[1] Sir J. J. Thomson (1856–1940) was for many years director of the Cavendish Laboratory at Cambridge University. He is credited with the discovery of the electron in 1897 and did the above experiment in 1910.

The velocity v of the particles in the beam is therefore

$$v = \frac{E}{B} \tag{29.8}$$

If the electric field is removed, the beam deflects to a measurable position and from this the radius of curvature of the path in the magnetic field can be found. When r is thus determined, elimination of the velocity from Eqs. (29.7) and (29.8) results in

$$\frac{q}{m} = \frac{E}{rB^2} \tag{29.9}$$

For an electron, experiments give

$$\frac{e}{m} = 1.76 \times 10^{11} \frac{\text{coulombs}}{\text{kg}}$$

and this can be checked by recalling that $e = 1.6 \times 10^{-19}$ coulomb and $m = 9 \times 10^{-31}$ kg.

29.4. Isotopes. Since most charged particles carry one or two electronic charges whose value q is accurately known, measurements of q/m are often used for the determination of the mass m of a particle. Thomson used his apparatus with charged atoms of several substances, as well as with electrons. In this way he discovered, for example, that neon atoms do not possess a unique mass but that there are two kinds, one with an atomic weight near 20 and a second with an atomic weight near 22. A third variety of neon with an atomic weight near 21 has been found by later experimenters. The atomic weight of neon, derived by density measurements, is 20.2, and this value is therefore the mean atomic weight of a mixture of components. These varieties of an element are called *isotopes* (Greek *isos*, equal; *topos*, place) since they all have the same chemical properties and hence all occupy the same place in the periodic table. Nearly all the chemical elements have been found since to consist of mixtures of isotopes in various proportions.

Table 29.1 shows the isotopes that have been found for the elements at the beginning of the periodic table. The integer nearest the atomic weight of an isotope is called the *mass number*. The approximate relative abundance of each isotope is also given. It is rather surprising that the relative abundances of the isotopes of an element are independent of the source from which the element is obtained. Careful investigation shows that there are indeed variations, although only small ones. A possible explanation is that all of a given element was formed under the same conditions and hence probably at the same time and that it then was distributed over the earth at a later period.

Table 29.1. *The Isotopes of the Elements*

Element	Atomic number	Mass numbers	Percentage abundances in order of mass number
H	1	1, 2	99.98, 0.02
He	2	3, 4	10^{-4}, 100
Li	3	6, 7	7.9, 92.1
Be	4	9	100
B	5	10, 11	18.4, 81.6
C	6	12, 13	98.9, 1.1
N	7	14, 15	99.62, 0.38
O	8	16, 17, 18	99.76, 0.04, 0.20
F	9	19	100
Ne	10	20, 21, 22	90.0, 0.27, 9.73
Na	11	23	100
Mg	12	24, 25, 26	77.4, 11.5, 11.1
Al	13	27	100
Si	14	28, 29, 30	89.6, 6.2, 4.2
P	15	31	100
S	16	32, 33, 34	95.0, 0.74, 4.2
Cl	17	35, 37	75.4, 24.6
A	18	36, 38, 40	0.307, 0.061, 99.632
K	19	39, 40, 41	93.3, 0.012, 6.7

29.5. The Mass Spectrograph. An apparatus designed for the accurate determination of atomic masses is called a mass spectrograph. One

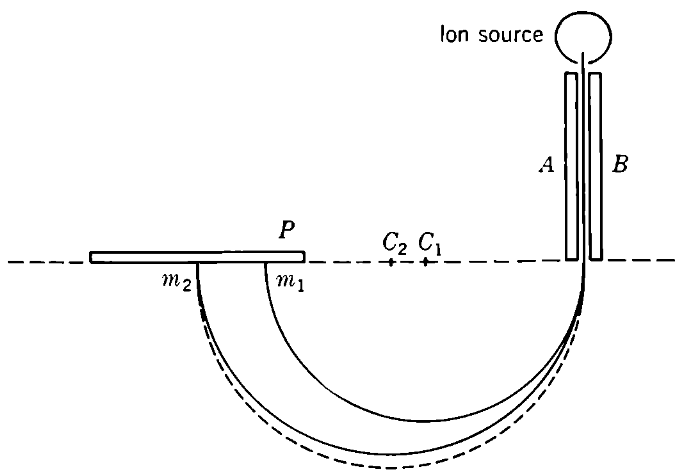


FIG. 29.6. A mass spectrograph with a velocity filter AB.

form of mass spectrograph is shown in Fig. 29.6. A uniform magnetic field, of large area and in a direction perpendicular to the paper, bends the moving ions in circular paths, and the ions impinge on a photographic plate at P, producing a blackening of the emulsion. Two closely spaced plates A and B form a velocity filter as well as a means to define the beam. An electric field between the plates counteracts the magnetic force for an ion with a definite value of velocity. Ions with other velocities strike the plates and are stopped. The semicircular path for an ion of mass m_1 has the center at C_1, and the path of a heavier ion of mass m_2 has the center at C_2. From the separation of the spots produced on the photographic plate and the distance of these spots from the exit of the velocity filter an accurate comparison of the masses m_1 and m_2 can be made. If, for example, ion 2 is a singly charged atom of hydrogen of mass number 2 (a deuterium ion) and ion 1 is a doubly charged atom of helium of mass number 4, the spots on the plate will be separated only slightly and the ratio of the masses can be accurately determined. Table 29.2 shows some of the mass values obtained with mass spectrographs. Correction must of course be made for the mass of the electrons removed to produce the ions. Mass determinations are made relative to the mass of the oxygen isotope of mass number 16.

Table 29.2. *The Masses of Some Isotopes as Measured by Mass Spectrographs*

(Values are relative to the mass of O^{16} = 16.0000)

Element	Atomic number	Mass	Element	Atomic number	Mass
Hydrogen.........	1	1.0081	Nitrogen.........	7	14.0075
	1	2.0147		7	15.0047
Helium...........	2	3.0170	Oxygen..........	8	16.0000
	2	4.0038		8	17.0045
				8	18.0049
Lithium..........	3	6.0169			
	3	7.0182	Fluorine..........	9	19.0045
Beryllium.........	4	9.0150	Neon............	10	19.9989
				10	21.0000
Boron............	5	10.0161		10	21.9986
	5	11.0129			
Carbon...........	6	12.0039			
	6	13.0076			

The arrangement just described has the property of focusing the ion beams. If an ion leaves the velocity filter at a small angle to the plates, it is bent in a path indicated by the dashed line in Fig. 29.6 and arrives at the same spot on the photographic plate as the ion that takes the solid-line path. For an arrangement to possess this property of focusing, the exit of the velocity filter, the center of curvature of the path, and the photographic plate must be on the same straight line.

Another mass spectrograph that also has the property of focusing the ions is shown in Fig. 29.7. Here the ion path is deflected only 60° by the magnetic field, but the defining slit B, the center of curvature C, and the detector D of the ions again lie on a straight line. The ions are given

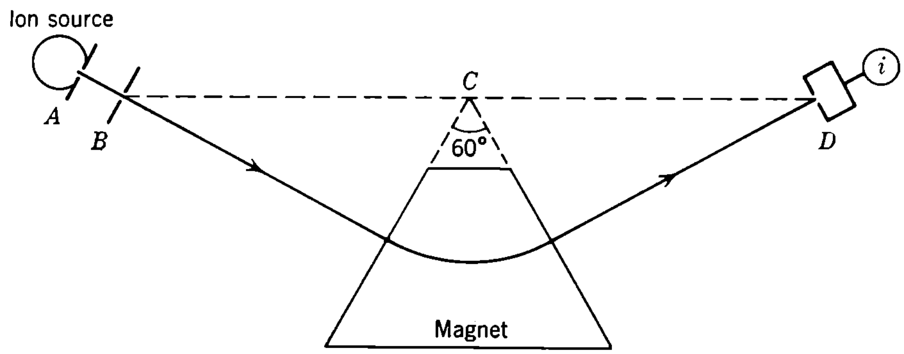


FIG. 29.7. A mass spectrograph with a 60° deflection.

a known velocity by the application of an accelerating potential V between metal slits A and B. If the charge is q, then

$$v = \sqrt{\frac{2qV}{m}}$$

and from Eq. (29.7)

$$\frac{q}{m} = \frac{2V}{r^2B^2} \tag{29.10}$$

The detector for ions is some sensitive current-measuring device. Hence this instrument is particularly suitable for the measurement of the relative numbers of each kind of ion present.

Mass spectrographs can also be used to separate small quantities of pure isotopes usually in amounts of the order of a microgram. Large equipment[1] of a similar nature has been built to separate large quantities of the isotope of uranium of mass number 235 which is used in "atomic bombs."

29.6. The Cyclotron. The cyclotron, an instrument that is in widespread use today for the investigation of the nuclei of atoms, utilizes the bending of the paths of charged particles in a magnetic field. Particles in a cyclotron are accelerated by the electric field between two hollow semicircular electrodes, or D's, indicated schematically in Fig. 29.8. These electrodes are contained in a vacuum chamber and placed between

[1] H. D. Smythe, "Atomic Energy for Military Purposes," Princeton University Press, 1945.

the poles of a powerful electromagnet. The magnetic field in Fig. 29.8 is perpendicular to the page and outward. A positive ion starting at the source S is accelerated until it reaches the hollow field-free space within one electrode, where it then moves in a circular path of radius given by Eq. (29.7). After traversing a semicircle the particle again reaches the space between the D's. If the electric field between the D's is an alternating one of the proper frequency, the field is now in the proper direction to accelerate the particle still further. With a larger value of v, the ion

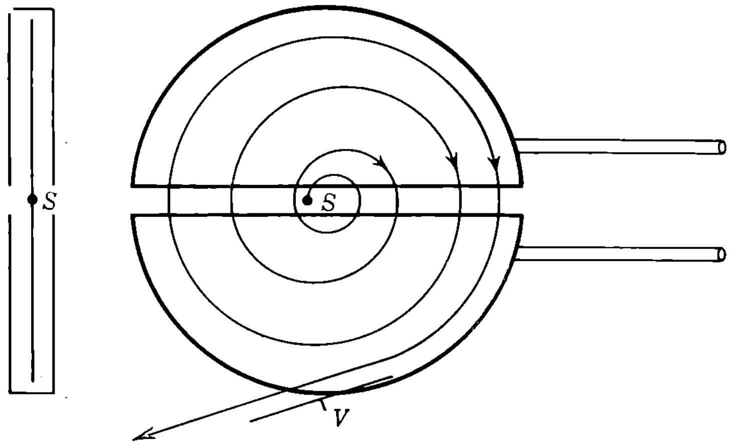


FIG. 29.8. The electrodes or D's of a cyclotron.

then describes a larger semicircle. The time t to travel a semicircle of radius r is

$$t = \frac{\pi r}{v} = \frac{\pi m}{qB} \tag{29.11}$$

if the value of r in Eq. (29.7) is used. Since this time is independent of v and r, a resonance condition exists and the electric field has the proper direction to accelerate the ion at every traversal if the field varies with a frequency f given by

$$f = \frac{1}{2t} = \frac{qB}{2\pi m} \tag{29.12}$$

The total energy that can be given to a particle is limited by the radius of the D's. The maximum energy W_{max} can be found by rearrangement of Eq. (29.7) as

$$W_{max} = \frac{1}{2} mv^2 = \frac{1}{2} \frac{q^2}{m} r_{max}^2 B^2 \tag{29.13}$$

When the particle reaches the maximum radius, it is bent out of the magnetic field by a high potential applied to the electrode V. Figure 29.9 is a photograph of the large cyclotron at the University of California.

A cyclotron can accelerate only relatively heavy particles such as protons, deuterons, or helium nuclei, since, to maintain the resonance condition, the mass of the particle must remain constant. When electrons are accelerated to speeds corresponding to millions of volts, their mass changes because of relativistic effects (see Chaps. 49 and 50). In consequence an electric field varying with a constant frequency does not remain in the proper time relation for acceleration to continue. If

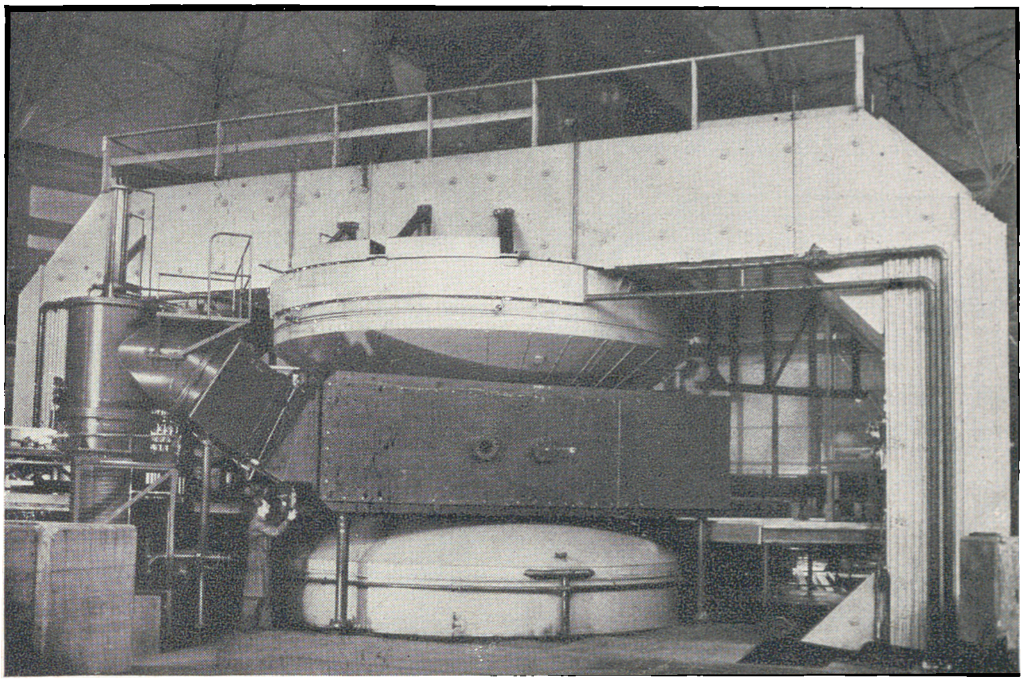


FIG. 29.9. A photograph of a cyclotron at the University of California. The diameter of the magnet pole pieces is 184 in. (*Courtesy E. O. Lawrence.*)

the frequency is slowly altered with time and the magnetic field changed, the resonance condition can be maintained and higher energies produced.

Worked Example. A cyclotron has a magnetic field of 10^4 gausses[1], or 1 weber/m^2, and the radius of the D's is 80 cm. What is the frequency of the alternating electric field that must be applied, and to what energy can deuterons be accelerated? For deuterons, $q = 1.6 \times 10^{-19}$ coulomb, and $m = 3.3 \times 10^{-27}$ kg. From Eq. (29.12), the frequency is

$$f = \frac{1.6 \times 10^{-19} \times 1}{2\pi \times 3.3 \times 10^{-27}} = 7.7 \times 10^6 \text{ cps} = 7.7 \text{ megacyles/sec}$$

[1]Although this plural may seem queer, the authors follow a recommendation of an International Commission on Electrical Units.

The maximum energy is, from Eq. (29.13),

$$W_{max} = \frac{1}{2}\frac{(1.6 \times 10^{-19})^2 \times (0.8)^2 \times 1^2}{3.3 \times 10^{-27}} = 2.5 \times 10^{-12} \text{ joule}$$

Since 1 ev = 1.6×10^{-19} joule, the energy in electron volts is found to be

$$W_{max} = 15.6 \times 10^6 \text{ ev}$$

29.7. The Magnetic Force on a Current. Since an electric current in a wire consists of moving electrons, these electrons, and therefore the wire, are also subject to a magnetic force. If there are n electrons per unit volume in a wire of area A, the total moving charge dq in a length dl of the wire is $neA\ dl$, whence $dq/dl = neA$. If v is the drift velocity of the electrons carrying a current i,

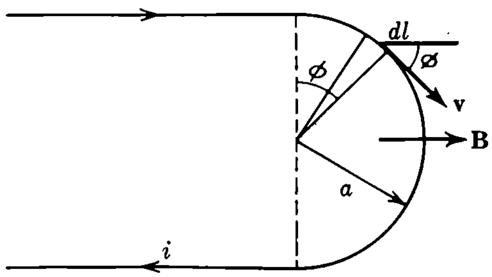


Fig. 29.10. A current-carrying wire in a field **B**.

$$i = neAv = \frac{dq}{dl} v \qquad (27.1)$$

Hence we may write $idl = vdq$. It should be remembered that for electrons both e and v are negative. Equation (29.2) written in differential form is $dF = dq\ v \sin \phi B$, ϕ being the angle between **v** and **B**. If now we replace vdq by idl, we find for the force on the length dl of the wire

$$dF = iB\ dl \sin \phi \qquad (29.14)$$

The total force on a wire must be found by summing Eq. (29.14) over the whole length of the wire. It must be remembered that $d\mathbf{F}$ is a vector quantity and therefore the sum must be a vector sum.

For example, consider the hairpin-shaped wire (Fig. 29.10) carrying a current i in a uniform field **B**. No force is produced on the straight sides since $\sin \phi = 0$. The force on the portion of the wire of length $dl = a\ d\phi$ is outward from the paper and equal to

$$dF = iBa\ d\phi \sin \phi$$

The total force on the wire is the integral of this, or

$$F = \int dF = \int iBa \sin \phi\ d\phi = iBa \int_0^\pi \sin \phi\ d\phi = 2iaB \qquad (29.15)$$

If a = 10 cm, i = 1 amp, and B = 1,000 gausses,

$$F = 2 \times 1 \times 0.1 \times 0.1 = 0.02 \text{ newton} = 2{,}000 \text{ dynes}$$

It should be noted that the force would have the same value if the wire ran across the end perpendicular to the sides as does the dashed line in the figure.

29.8. Torque of Current-carrying Coils in a Magnetic Field. Since a magnetic field produces a force on a wire in which a current flows, a torque is produced if a loop of wire with a current flowing in it is suspended in a magnetic field. Suppose we have a rectangular coil consisting of a single turn of breadth b and height a in a field B, as shown in Fig. 29.11, where the field is in the plane of the coil. There is no force on the top or bottom of the coil. The force on the element dl on the right side of the coil is directed outward from the paper and given by Eq. (29.14) with $\sin \phi = 1$. The torque about the axis xx' from the side of the coil is therefore $iaBb/2$. The force on the left side of the coil is directed into the paper, and the torque has the same magnitude. The total torque L is therefore

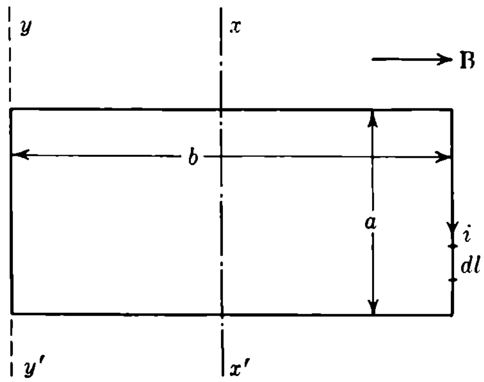


FIG. 29.11. A rectangular coil in a magnetic field.

$$L = iabB \tag{29.16}$$

and hence is proportional to the area ab of the coil. If the field B makes an angle θ with the plane of the coil, Eq. (29.16) is generalized to

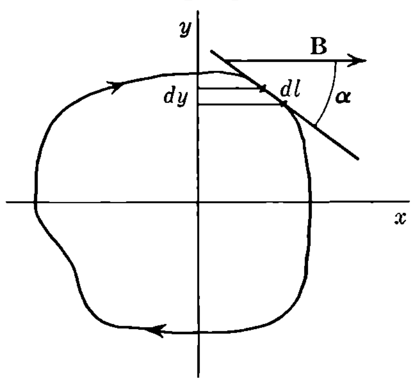


FIG. 29.12. A coil of arbitrary shape.

$$L = iabB \cos \theta \tag{29.17}$$

If **B** is perpendicular to the coil, then the torque is zero.

The value of the torque is independent of the axis about which the coil is supported. For example, if the axis yy' is chosen, only the righthand side of the coil contributes to the torque; but the distance to the axis yy' is b, twice as great as before, and Eq. (29.16) remains unchanged.

If the area of the coil is A, then the torque can be written as

$$\boxed{L = iAB} \tag{29.18}$$

It can be easily shown that Eq. (29.18) is true for a plane coil of any shape. Such a coil is shown in Fig. 29.12 with rectangular coordinate axes indicated. The torque about the y axis due to the current element dl is

$$dL = iB \sin \alpha \, x \, dl$$

where α is the angle between the field B and the tangent to the coil. The quantity $dl \sin \alpha$ is, however, equal to the element dy. The total torque, therefore, is

$$L = iB \int x \, dy = iBA$$

since the integral, with proper care taken about signs, is just the area A of the coil.

If the coil has n turns, each turn contributes to the torque and, for a field at an angle θ with the plane of the coil, the general expression can be written

$$L = niAB \cos \theta \qquad (29.19)$$

Since the normal flux Φ through the coil is $AB \sin \theta$, Eq. (29.19) can be written

$$L = ni \frac{d\Phi}{d\theta} \qquad (29.20)$$

A coil in a magnetic field thus tends to turn until $L = 0$, *i.e.* until the flux through it has a stationary value. Stable equilibrium requires Φ to be a maximum.

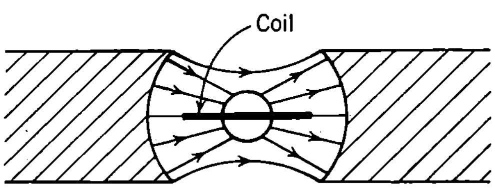


FIG. 29.13. The field **B** in an electric meter.

29.9. Ammeters and Voltmeters. The torque on a coil in a magnetic field forms the basis of most of the commonly used electrical instruments. A coil supported on jeweled bearings and furnished with a counter-weighted pointer is placed in the magnetic field of a permanent magnet. The ends of the magnet have a special shape, indicated in Fig. 29.13, so that the field **B** is approximately radial and is therefore independent of the orientation of the coil for a range of angles. If the torque due to current flowing through the coil is balanced by a restoring torque from a flat spiral spring, the angular deflection of the pointer is proportional to the current through the coil and the instrument is an *ammeter*. By small adjustments of the value of **B** the meter can be made to read a given current, say 1 ma, at full-scale deflection.

The range of the instrument can be extended and larger currents can be measured by the use of a resistor, called a *shunt*, in parallel with the meter. The current flowing through the ammeter and shunt divides, and a constant fraction flows through the meter coil. If the meter current and resistance are denoted by i_m and R_m and the corresponding quantities for the shunt by i_s and R_s, respectively, then, from Kirchhoff's laws (see Fig. 29.14),

$$\begin{aligned} i_m R_m &= i_s R_s \\ i &= i_m + i_s \end{aligned} \qquad (29.21)$$

The ratio of currents is therefore, on elimination of i_s,

$$\frac{i_m}{i} = \frac{R_s}{R_m + R_s} \qquad (29.22)$$

If a meter that reads 1 ma for a full-scale deflection has a resistance of 75 ohms, then the resistance of the shunt necessary to convert it to a meter reading 1 amp full scale is given by

$$\frac{0.001}{1} = \frac{R_s}{75 + R_s}$$

or $R_s = {}^{75}\!/\!_{999} \approx 0.075$ ohm

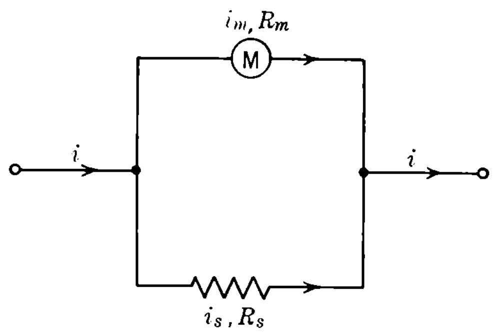


FIG. 29.14. An ammeter with a shunt resistor.

The shunt resistor is often incorporated into the meter case, and sometimes a switch is furnished to enable one of a series of shunts to be selected.

An electric meter can also be used to measure the potential difference between its terminals. The current through the meter is equal to the potential difference divided by the resistance; hence the angular deflection of the coil is proportional to the difference of potential. For a 1-ma meter with a resistance of 75 ohms, a potential difference of 75 mv across the terminals causes a full-scale deflection. Larger differences in potential can be measured by using a *multiplier* resistor in series with the meter, as indicated in Fig. 29.15. If the meter sensitivity is i amp for full-scale deflection, then the proper multiplier resistance R to make the meter read full scale for E volts across the terminals is given by

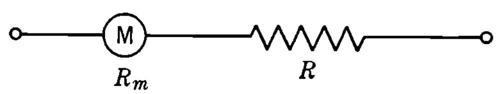


FIG. 29.15. A voltmeter.

$$E = i(R + R_m) \qquad (29.23)$$

The resistance R to convert the 1-ma meter to a 150-volt voltmeter is therefore

$$\begin{aligned} 150 &= 0.001(R + 75) \\ R &= (150{,}000 - 75) \text{ ohms} \end{aligned}$$

A voltmeter draws some current from the electric circuit in which it is used and often, therefore, introduces changes in the circuit behavior. To minimize these effects a meter as sensitive as possible should be used. The sensitivity is usually expressed, not as the current sensitivity of the meter movement, but as the reciprocal of this in units of ohms per volt. Thus a 1-ma meter would make a voltmeter with a sensitivity of 1,000 ohms per volt.

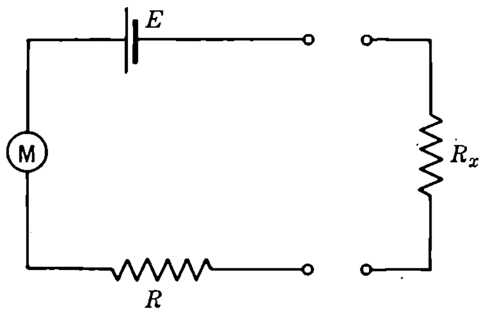


FIG. 29.16. The circuit of an ohmmeter.

A useful piece of test equipment can be made with an ammeter, a dry cell, and a resistor in series, as in Fig. 29.16. The series resistor R is so chosen that, when the terminals are short-circuited, the meter reads full scale. If this current is i and the emf of the dry cell E, then

$$E = iR$$

If an unknown resistor R_x is placed across the terminals, the current will be i_x and less than i, where

$$i_x = \frac{E}{R + R_x} = \frac{i}{1 + R_x/R} \tag{29.24}$$

The meter reading is therefore directly related to the value of R_x. The meter can be furnished with a special scale (Fig. 29.17) from which R_x

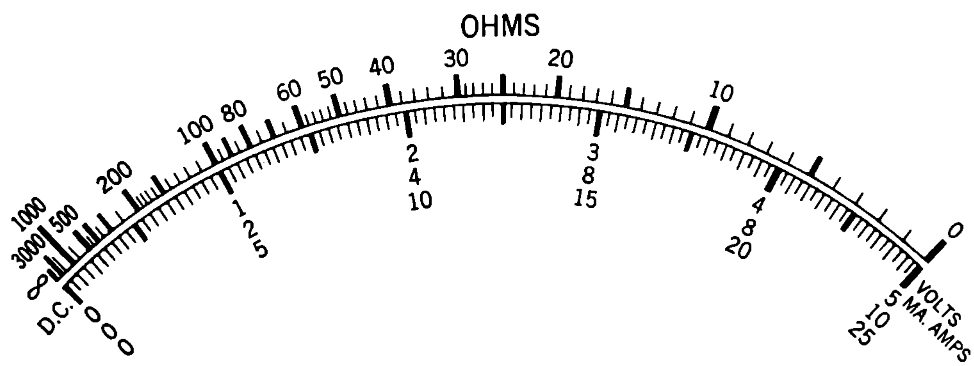


FIG. 29.17. An ohmmeter scale.

can be read directly. Such a device is called an *ohmmeter*. It is not a particularly accurate instrument but an extremely useful one for the location of faulty components in an electric circuit.

29.10. Galvanometers. A very sensitive current-detecting device can be made by suspending a coil in a magnetic field from a flat metallic strip, which acts as a torsion fiber to furnish the restoring torque. Current is conducted to the coil by this metallic suspension, and the circuit is completed by a loose spiral of thin wire or strip hanging from below the coil. Such a device is known as a *galvanometer*. The angular deflection of the coil can be indicated by the movement of a pointer, but in more sensitive instruments a mirror is rigidly attached to the moving system. A spot of light can then be reflected from the mirror to a scale, or the image of a fixed scale can be viewed in the mirror through a telescope. A galvanometer with a suspended coil is known as a *D'Arsonval* galvanometer. The sensitivities of D'Arsonval galvanometers usually lie in the range of 10^{-8} to 10^{-11} amp to produce a scale deflection of 1 mm on a scale 1 m from the mirror, a deflection of 0.001 radian.

Galvanometer sensitivities are also specified sometimes in *megohms*. The *megohm sensitivity* is the resistance in megohms that must be placed in series with one volt to produce a unit deflection of the galvanometer. Thus a current sensitivity of 10^{-8} amp corresponds to a sensitivity of 100 megohms.

Since galvanometers as well as pivoted-coil instruments are subject to a restoring torque that is proportional to the angular deflection, the motion of the coil can be described by the differential equation for damped SHM (Sec. 12.10). For ease of use an instrument is usually employed in such a way that it is critically damped or nearly so (Sec. 12.11). Thus the pointer or light spot comes to rest in the minimum time. The undamped period is therefore an important parameter in the selection of a galvanometer. Periods range from 1 to more than 10 sec. The more sensitive the galvanometer, the longer the period, since to increase the sensitivity the restoring torque must be made less; the other constants of the galvanometer can be changed relatively little.

If a sudden spurt of current is sent through a galvanometer, an angular momentum $I\omega$ is given to the coil. Since (see Secs. 10.5 and 10.6)

$$I\omega = \int L\,dt = K \int i\,dt = Kq, \tag{29.25}$$

where K is a constant of proportionality, the angular momentum is proportional to the charge q that passes. The coil deflects until the kinetic energy of rotation is transformed to potential energy of the twisted suspension. The maximum value of the deflection is therefore proportional to the charge. A galvanometer used in this manner is called a *ballistic* galvanometer, since the action is closely analogous to the action of a ballistic pendulum. A ballistic galvanometer can be calibrated easily by discharging a capacitor through it. Ballistic galvanometers are also useful in comparing capacitances by comparison of the deflections produced when the capacitors charged to the same potential are discharged through the galvanometer (see Prob. 15).

29.11. Direct-current Motors. One of the most important applications of the force on a current-carrying wire in a magnetic field is the electric motor. One arrangement is shown schematically in Fig. 29.18.

An *armature A* of soft iron is mounted in the magnetic field of a permanent magnet and has a coil of wire wound around it so that the current on the left side of the armature is flowing into the paper, as indicated by the crosses in the figure, and that on the right side is flowing out, as indicated by the dots. Each wire then has the current in the proper direction to produce a counterclockwise torque on the armature. As the armature rotates, a switching arrangement called a *commutator* (not shown in the figure) connects the armature windings so that the current directions shown do not change. The torque developed can be used to do mechanical work. If energy is extracted from the electrical system, however, a *counter emf* is developed in the armature. The magnitude of this counter emf times the current flowing in the armature is equal to the mechanical power transformed from electric power. Some power is lost, of course, in the resistance heating of the armature wires.

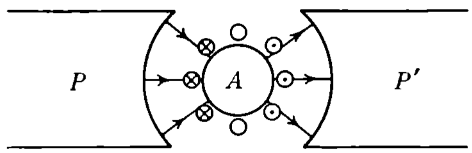


FIG. 29.18. A direct-current motor.

The magnetic field in a motor is often produced by current flowing in coils wound around the poles *PP′* of the motor. These *field coils* can be connected in parallel with the armature, as in the *shunt* motor, or in series with the armature, as in the *series* motor. In the *compound motor* there are two sets of field coils, one in parallel and the other in series with the armature. The motor indicated in the figure is called a *two-pole* motor. Other pairs of poles symmetrically placed can be added to form *four-pole* or *six-pole* machines.

PROBLEMS

1. What is the relation between 1 maxwell and 1 weber?

2. The electron moving in a hydrogen atom is placed in a field of 1,000 gausses perpendicular to the orbit. Compare the electrostatic attraction of the nucleus with the centripetal magnetic force; see Sec. 11.9.

3. Show that the pitch of the helical path of a particle moving with a velocity v at an angle ϕ with the field is $2\pi mv \cos \phi/qB$.

4. The field of the earth is roughly 0.6 gauss. What is the radius of curvature of an electron with 1,000 ev of energy?

5. A cyclotron is designed to have a field of 1.3 webers/m^2 over a circular area of radius 1.5 m. To what energies can protons be accelerated; deuterons; doubly charged helium ions?

6. What frequencies of the accelerating voltage are necessary for the three particles in Prob. 5?

7. A circular turn of wire of radius 10 cm is perpendicular to a field of 1,000 gausses. What is the tension across the cross section of the wire for a current of 1 amp?

8. A galvanometer coil is 1 by 3 cm and composed of 100 turns. In a field of 500 gausses parallel to the coil what is the torque produced for 1 μa of current?

9. A standard meter movement has a 10-ma full-scale deflection and a resistance of 62.5 ohms. What are the shunt resistors necessary to convert this movement to read full-scale deflection for 1 amp; 10 amp; 100 amp?

10. What is the sensitivity in ohms per volt of a meter giving full-scale deflection for 50 μa? How much resistance should be placed in series to make a 1,000-volt voltmeter?

11. A 1-ma meter of 75 ohms resistance is used to construct an ohmmeter with a dry cell of 1.5 volts emf. What series resistance should be used? Construct a table showing the resistance value of the external resistor that corresponds to each $\frac{1}{10}$ ma deflection.

12. A shunt motor has a field resistance of 200 ohms and an armature resistance of 10 ohms. If the motor while running draws a current of 5.5 amp when connected across a source of emf of 100 volts, what is the counter emf in the armature? How much mechanical power is produced?

13. A $\frac{1}{4}$-hp motor runs from a 115-volt power line. If the motor efficiency is 80 per cent, what current is drawn from the line?

14. Prove the statement in the last sentence in Sec. 29.7.

***15.** Show by using the principle of conservation of energy that, if ω is given by Eq. (29.25), the deflection of the galvanometer is proportional to q.

CHAPTER 30

THE MAGNETIC FIELD PRODUCED BY CURRENTS

30.1. The Field of a Current Element. It will be recalled that the electromagnetic interaction of two currents was divided into two problems. One current produces a magnetic field, and this field acts upon the second current. In the preceding chapter the second problem was discussed. We shall now treat the production of a field by a current.

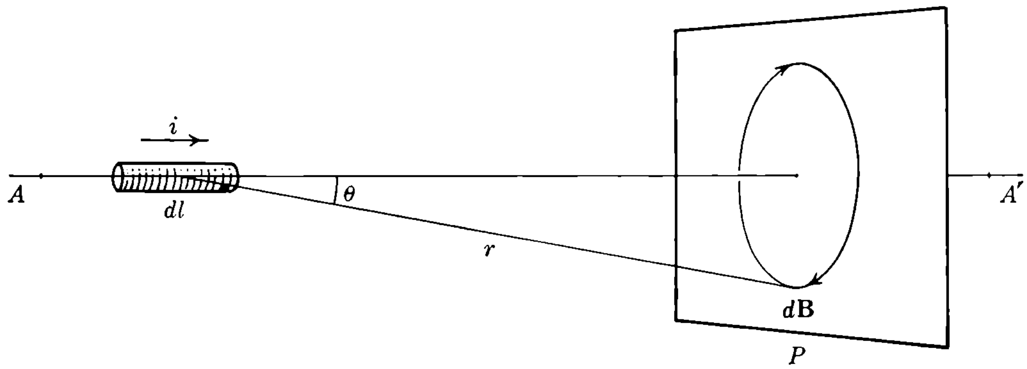


FIG. 30.1. The field of a current element.

The magnetic effect of a current as indicated by the deflection of a compass needle was first discovered by Oersted (1777–1851), a Danish physicist. It is said that the discovery was made during a lecture in which he was demonstrating that no such effect existed. Further experiments and the development of the theory were carried out by the Frenchmen Ampère, Biot (1774–1862), and Savart (1791–1841).

The magnetic effects of a current must be formulated in terms of current elements, since each portion of a complete circuit in which a current flows makes some contribution to the magnetic field at every point near the circuit. We imagine a current i flowing in an element of length dl. This current element produces an element of field $d\mathbf{B}$ that has a direction perpendicular to dl and a magnitude equal to

$$dB = k' \frac{i\, dl \sin\theta}{r^2} \tag{30.1}$$

where the quantities r and θ are shown in Fig. 30.1. The plane P is perpendicular to the axis AA' of the current element, and the lines of induction are closed circles on P. The direction of $d\mathbf{B}$ is the direction of rotation of a right-handed screw advancing in the direction of the current element. If the thumb of the right hand is held in the direction of the currents, then the fingers curl in the direction of the magnetic induction. Along the axis of the current element, $d\mathbf{B} = 0$ since $\theta = 0$; and $d\mathbf{B}$ is a maximum in the plane through the element where $\sin\theta = 1$. Equation (30.1) is sometimes known as *Ampère's law.*

The constant k' in Eq. (30.1) has a value dependent on the choice of units, similar to the constant k in Coulomb's law. The cgs system of units in which k' is unity is called the electromagnetic system of units. If i is measured in abamperes (abcoulombs per second) and dl and r in centimeters, then dB is expressed in gausses.

In the practical mks system of units, k' has exactly the value of 10^{-7}. For i in amperes, dl and r in meters, dB is expressed in webers per square meter. In order to avoid the occurrence of the factor 4π in other equations, the constant k' is usually written as

$$k' = \frac{\mu_0}{4\pi} \tag{30.2}$$

and hence $$\mu_0 = 4\pi \times 10^{-7} \text{ weber amp}^{-1}\text{ m}^{-1}$$

or $$\mu_0 = 1.257 \times 10^{-6} \text{ weber amp}^{-1}\text{ m}^{-1} \tag{30.3}$$

Worked Example. A wire 1 cm long carries a current of 1 amp. What is the field in a plane perpendicular to the wire at a point 1 m away from it?

In practical units,

$$dB = 10^{-7}\frac{1 \times 0.01}{1^2} = 10^{-9} \text{ weber/m}^2$$

In emu,

$$dB = \frac{0.1 \times 1}{(100)^2} = 10^{-5} \text{ gauss} = 10^{-9} \text{ weber/m}^2$$

since 1 amp = 0.1 abamp and 10^4 gausses = 1 weber/m.^{2}

30.2. The Field of a Straight Wire. To obtain the field produced by a long wire the contributions from each current element in the wire must be added by integration of Eq. (30.1). A long wire is shown in Fig. 30.2. The field dB at a point a distance a from the axis of the wire, due to the element dl, is

$$dB = k'\frac{i\,dl\sin\theta}{r^2}$$

The element dl and the distance r can be expressed in terms of a and the angle θ, thus

$$a = r \sin \theta$$

$$dl = \frac{r\,d\theta}{\sin \theta} = \frac{a\,d\theta}{\sin^2 \theta}$$

Hence

$$dB = k' \frac{i \sin \theta\, d\theta}{a}$$

The contribution of each element dl is in the same direction, and therefore

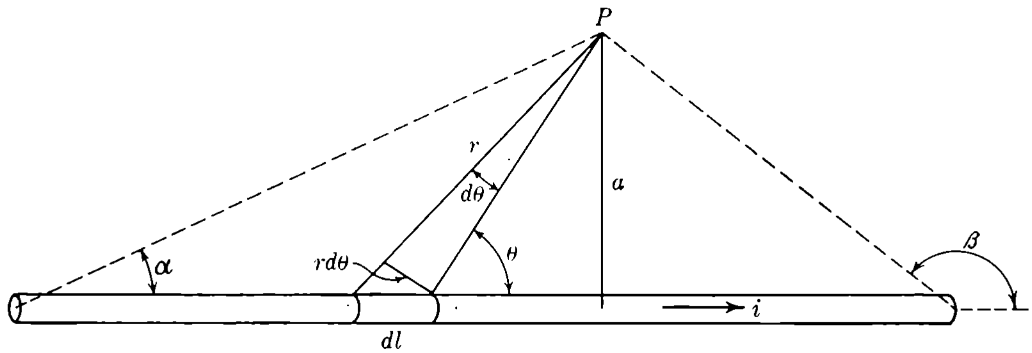


FIG. 30.2. The field near a long wire.

the total field is

$$B = \int dB = k' \frac{i}{a} \int_\alpha^\beta \sin \theta\, d\theta$$

or

$$B = k' \frac{i}{a} (\cos \alpha - \cos \beta) \tag{30.4}$$

where the angles α and β are shown in the figure. For a very long wire, $\alpha = 0$, $\beta = \pi$, and

$$B = k' \frac{2i}{a} \tag{30.5}$$

If emu are used, Eq. (30.5) becomes

$$B = \frac{2i}{a} \qquad \text{in gausses}$$

In practical units,

$$B = \frac{\mu_0}{2\pi} \frac{i}{a} \qquad \text{in webers/m}^2$$

30.3. The Force between Two Current-carrying Wires. We can now combine the result of the last section with Eq. (29.14) and obtain the electromagnetic force between two wires each carrying a current. Let us consider a length dl of the wire carrying a current i. The field at dl

due to another parallel wire a distance a away and carrying a current i' is

$$B = k' \frac{2i'}{a}$$

The field is perpendicular to the plane of the wires and hence $\sin \phi$ in Eq. (29.14) is unity. The force on dl is therefore

$$dF = k' \frac{2ii'}{a} dl \tag{30.6}$$

The direction of the force is such that the wires attract each other if the currents i and i' are in the same direction and repel each other if the currents are opposite.

Equation (30.6) is an extremely important one since by means of it a rigorous definition of the unit of current can be made. In the practical system of units the ampere is defined as that amount of current flowing in each of two parallel wires separated by 1 meter that causes a force of attraction of 2×10^{-7} newton per meter of length. A similar definition can be made for the emu of current. It is thus possible to make an absolute determination of the ampere, *i.e.*, a determination in terms of purely mechanical quantities: forces, and distances. In practice this is done by measuring the force between two coaxial coils in series. One coil is stationary, the other attached to one arm of a precision balance. Such a device is called a *current balance*. Once the value of the ampere is established, the value of the coulomb follows immediately as the amount of charge passed in a current of one ampere for one second. Units of potential and emf are then established in terms of the coulomb and purely mechanical quantities.

***30.4. Electric and Magnetic Forces between Charges.** Since a charge q moving with a velocity v constitutes a current element, the magnetic field produced by a moving charge follows from Eq. (30.1), with $i\,dl$ replaced by qv and dB by B [see Eq. (27.1)]

$$B = k' \frac{qv \sin \theta}{r^2} \tag{30.7}$$

We are now in a position to compare the electrostatic and electromagnetic forces between two charges. Suppose we have a charge q moving with a velocity v and a second charge q' moving with velocity v', and let us consider the simple case when v and v' have the same direction and the line joining the charges is perpendicular to their paths as in Fig. 30.3. The electrostatic force between the charges is given by Coulomb's law

$$F_{\text{elec}} = k\frac{qq'}{r^2} \tag{30.8}$$

The electromagnetic force F_{mag} is given by combining Eqs. (30.7) and (29.1) and is

$$F_{\text{mag}} = -k'\frac{qvq'v'}{r^2} \tag{30.9}$$

The negative sign occurs in Eq. (30.9) because for like charges F_{mag} is an attractive force and tends to decrease r. The total force between the charges is the sum of these two forces. To investigate the relative magnitudes, let us form the ratio

$$-\frac{F_{\text{elec}}}{F_{\text{mag}}} = \frac{k}{k'}\frac{1}{vv'} \tag{30.10}$$

Since the ratio of two forces must be a dimensionless quantity, we may conclude that k/k' must have the dimensions of the square of a velocity. This velocity must be a constant of nature, and experimental determinations show that it is equal to the velocity c of light in vacuum. Hence

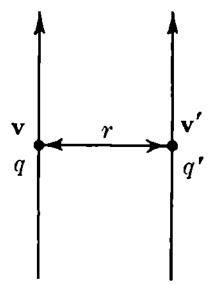


FIG. 30.3. Two moving charges.

$$\boxed{\frac{k}{k'} = c^2} \tag{30.11}$$

where $c = 2.998 \times 10^8$ m/sec

and

$$-\frac{F_{\text{elec}}}{F_{\text{mag}}} = \frac{c^2}{vv'} \tag{30.12}$$

The electrostatic forces thus far outweigh the magnetic forces between charges moving with ordinary velocities. As the velocities approach that of light, the total force between the charges approaches zero since the electromagnetic attraction approaches in magnitude the electrostatic repulsion.

In the practical mks system of units the constants k and k' are written in terms of two others, ϵ_0 and μ_0 by the relations

$$\left.\begin{aligned} k &= \frac{1}{4\pi\epsilon_0} \\ k' &= \frac{\mu_0}{4\pi} \end{aligned}\right\} \tag{30.13}$$

The ratio of the constants is

$$\frac{k}{k'} = \frac{1}{\epsilon_0\mu_0} = c^2 \tag{30.14}$$

The unit of current in the mks system is defined in terms of the attraction of two wires, and μ_0 has the value of $4\pi \times 10^{-7}$ exactly. The value of ϵ_0 can be derived most precisely from experiments on the velocity of light since

$$\epsilon_0 = \frac{1}{\mu_0 c^2} = \frac{10^7}{4\pi c^2} \tag{30.15}$$

In the cgs system of units, if electrostatic units of charge are used in Eq. (30.8), then $k = 1$. Furthermore, if electromagnetic units are used in Eq. (30.9), then $k' = 1$. It is evident, since Eq. (30.12) must remain true, that the units of charge in the two systems must have the ratio c, or

$$q \text{ (esu)} = \frac{1}{c} q \text{ (emu)} \tag{30.16}$$

It also follows from Eq. (30.16) that charge does not have the same *dimensions* in the electrostatic and electromagnetic system, but the ratio of dimensions must have the dimensions of a velocity.

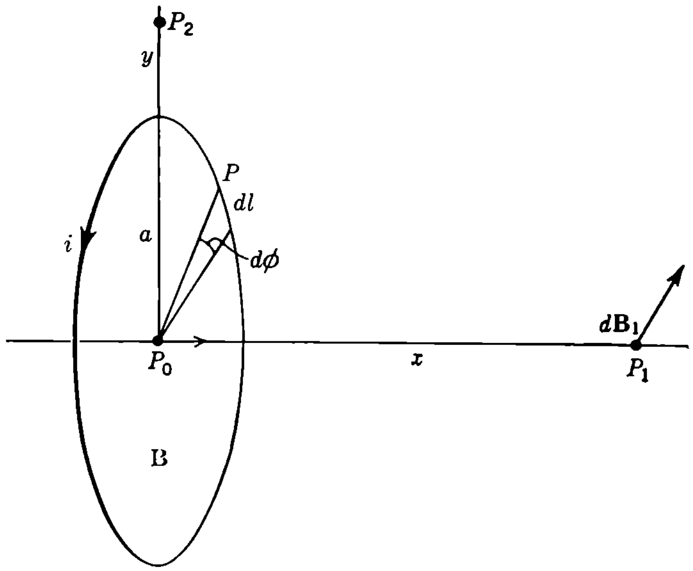


FIG. 30.4. The field of a circular loop.

30.5. The Magnetic Field of a Current Loop. Consider the point P_0 at the center of a circular loop of radius a carrying a current i, as in Fig. 30.4. The elements of field due to all current elements on the loop are in the same direction, and the total field B_0 is obtained directly by integration of Eq. (30.1),

$$B_0 = k' \int \frac{i}{a^2}\, dl$$

since $\sin\theta = 1$ and $r = a$. The element $dl = a\, d\phi$, and hence

$$B_0 = k' \frac{i}{a} \int_0^{2\pi} d\phi = k' \frac{2\pi i}{a} \tag{30.17}$$

If the coil has n turns, then the right-hand side of Eq. (30.17) is simply multiplied by n.

In emu, $k' = 1$,

$$B_0 = \frac{2\pi i}{a} \qquad \text{emu},$$

and for practical units, $k' = \mu_0/4\pi$,

$$B_0 = \mu_0 \frac{i}{2a} \qquad \text{mks units}$$

The field B_1 at the point P_1 on the axis but not at the center of the coil must also be found by integration. The field contribution from dl at the point P has the direction shown in the figure. It is obvious from the symmetry of the loop, however, that the resultant field lies along the axis of the loop and only the components of the contributions in the axial direction from the separate elements need be summed. The element of field $d\mathbf{B}_1$ lies in the plane passing through P, P_0, and P_1 as shown in Fig. 30.5. If the point P_1 is a distance x from the center of the loop, the contribution from one current element is

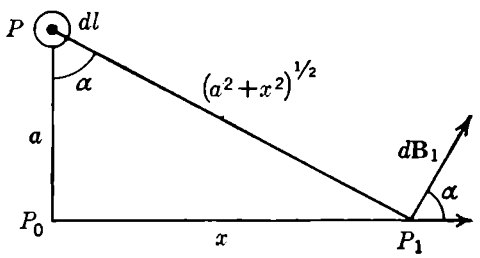


FIG. 30.5. The element of field from a portion of the circular loop of Fig. 30.4.

$$dB_1 \cos\alpha = k' \frac{i\,dl}{a^2 + x^2} \cos\alpha = k' \frac{i\,dl}{(a^2 + x^2)} \frac{a}{(a^2 + x^2)^{1/2}} = k' \frac{ai\,dl}{(a^2 + x^2)^{3/2}}$$

and the total field is

$$B_1 = k'i \frac{2\pi a^2}{(a^2 + x^2)^{3/2}} \tag{30.18}$$

since $\int dl = 2\pi a$ and x is constant.

At values of x large compared with the radius of the coil the field can be written

$$B_1 = k'i \frac{2\pi a^2}{x^3} \tag{30.19}$$

At a point P_2 in the plane of the loop, but a distance y from the loop center, the field dB_2 from each element must be split into components and summed as above. The result, however, is complicated and must be expressed as an infinite series. For values of y large compared with a, the result has a simple form,

$$B_2 = -k'i \frac{\pi a^2}{y^3} \tag{30.20}$$

The field $\mathbf{B}_2$ is parallel to the axis of the loop but in the opposite direction to the field at the loop center. The lines of induction about the loop are indicated in Fig. 30.6. It should be noted that all lines of B form closed curves. The axis AA' is an axis of symmetry of the figure.

Equations (30.19) and (30.20) are very similar to Eqs. (25.7) and (25.8), which describe the electric field at large distances from an electric dipole of moment p. To make the similarity complete, let us introduce a *magnetic dipole* moment analogous to p, calling it M. It is necessary, however, to keep an open mind about the constant of proportionality, which in Eqs. (25.7) and (25.8) was k'. For the magnetic case, we shall call it k'' and hold its value undetermined for later adjustment. This constant, as will appear soon, is the third and last important unit constant, besides k and k' which have already been introduced, in the subject of electromagnetism.

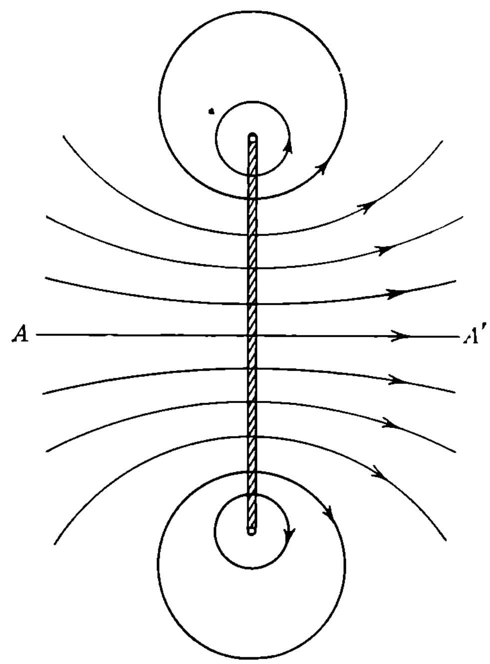


FIG. 30.6. The lines of induction about a circular loop.

Equations (30.19) and (30.20) then take the form

$$\left.\begin{aligned} B_1 &= \frac{2k''M}{x^3} \\ B_2 &= -\frac{k''M}{y^3} \end{aligned}\right\} \quad (30.21)$$

Comparison with Eqs. (30.19) and (30.20) yields

$$M = \frac{k'}{k''} i\pi a^2 = \frac{k'}{k''} iA \quad (30.22)$$

expressing the important fact that a circuit of area A carrying a current i is the magnetic equivalent of an electric dipole.

To anticipate, the constant $k'' = 1/4\pi$ in practical units; but for emu, $k'' = 1$.

30.6. The Magnetic Field of a Solenoid. A coil wound in the form of a solenoid (Greek *solen*, tube) can be considered for most purposes to be composed of a series of circular turns. The field on the axis of the solenoid is then the resultant of contributions each given by Eq. (30.18). If the solenoid has a length l and is composed of N turns, then the number of turns in the portion of the solenoid (see Fig. 30.7) between x and $x + dx$ is $(N/l)\,dx$. The field contributed by this portion is, from Eq. (30.18),

$$dB = k'i\frac{N}{l}\frac{2\pi a^2}{(a^2 + x^2)^{3/2}}\,dx$$

The total field is obtained by integration. It is convenient to change the independent variable from x to the angle θ. Since

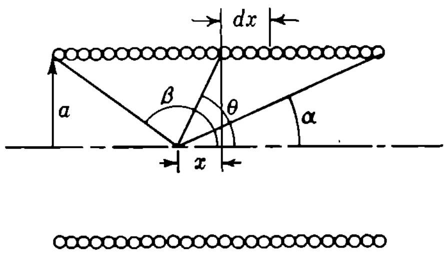


FIG. 30.7. The field of a solenoid.

$$x = a \cot \theta$$
$$dx = -a \csc^2 \theta \, d\theta$$

and the total field is

$$B = -k'i\frac{N}{l}2\pi \int_\beta^\alpha \sin\theta \, d\theta$$

$$B = k'i\frac{N}{l}2\pi(\cos\alpha - \cos\beta) \qquad (30.23)$$

If the solenoid is very long, the field near the center of the coil is

$$\boxed{B = k'\frac{4\pi iN}{l}} \qquad (30.24a)$$

since $\alpha = 0$, $\beta = \pi$. In practical units,

$$B = \mu_0\frac{iN}{l} \qquad \text{mks units} \qquad (30.24b)$$

In emu,

$$B = \frac{4\pi iN}{l} \qquad \text{emu} \qquad (30.24c)$$

At the end of a long solenoid, $\alpha = 0$, $\beta = \pi/2$, and

$$B = k'\frac{2\pi iN}{l} \qquad (30.25)$$

The field is therefore just half as great as at the center of the solenoid. The lines of induction about a solenoid are indicated in Fig. 30.8. Since the field at the end is only half that at the center, half the lines of force must cut the solenoid.

If a long solenoid is bent into a circle, a toroidal coil results. The field of a toroid is confined to the interior and must be uniform around the circumference. Since there are no end effects,

$$B = k'\frac{4\pi iN}{l} \qquad (30.26)$$

if the radii of the turns of the toroid are small compared with the radius of the toroid.

The fields of a solenoidal or toroidal coil have been calculated on the assumptions that the turns are truly circular and not helical and that

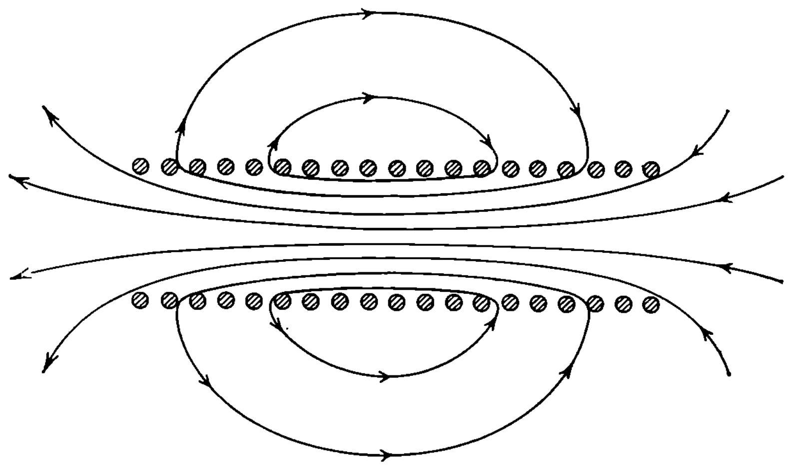


FIG. 30.8. The lines of induction about a solenoid.

every point on the axis can be considered to be at the center of a turn. These assumptions are tantamount to a replacement of the coil of wire by a *current sheet* that flows accurately perpendicularly to the axis of the solenoid as indicated in Fig. 30.9. The surface current density j or current per unit length of the solenoid is

$$j = \frac{iN}{l} \qquad (30.27)$$

and the field can be written

$$B = k' \times 4\pi j \qquad (30.28)$$

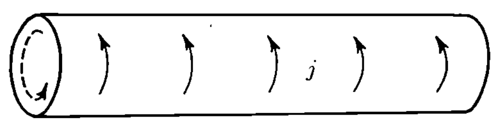


FIG. 30.9. A solenoid is equivalent to a current sheet of current density j.

***30.7. The Wattmeter.** A moving-coil instrument can be designed to measure directly the power consumed in a load on an electric circuit. A coil M pivoted on jeweled bearings and furnished with a restoring spring and a pointer as in a voltmeter is connected across the load R as in Fig. 30.10. The magnetic field in which the moving coil M rotates is provided by two fixed coils F, F', which are connected in series with the load R. The same current thus flows in these fixed coils and in the load, and the magnetic field produced is proportional to this current. The current through the moving coil is, however, proportional to the potential difference across the load. The torque on the moving coil, and therefore the coil deflection,

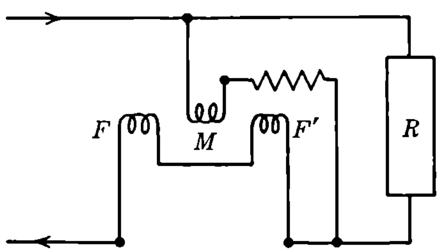


FIG. 30.10. Circuit diagram of a wattmeter.

is thus proportional to the product of the load current and the potential difference, and hence proportional to the power dissipated in the load. Such an instrument is called a *wattmeter*.

If the two sets of coils of a wattmeter are connected in series, the torque on the moving coil is proportional to the *square* of the current through the meter. If this current varies or even reverses its direction, the average torque on the coil has the same direction and is proportional to the average value of the square of the current, i^2. An instrument connected in this way is called a *dynamometer*.

The *watt-hour* meter is simply a precision motor whose speed is proportional to the power consumed. The number of revolutions is therefore a measure of the total energy dissipated in the load. The motor revolutions are counted mechanically, and watt-hour meters usually read directly in kilowatt-hours.

PROBLEMS

1. What is the magnetic field 10 cm away from a long wire carrying a current of 10 amp? Obtain the result in both practical units and emu.

2. Define the emu of current in terms of the force between two parallel current-carrying wires.

3. A current balance is constructed with coils 10 cm in diameter of 100 turns, having an average distance between coils of 2 cm. For a current of 1 amp through the coils, what mass is needed on the balance to compensate the force of attraction?

★4. Two electrons of 10 kv of energy are moving in parallel paths 1 mm apart. What are the electrostatic and magnetic forces between them? What is the total force?

★5. From Coulomb's law and Eq. (30.16) find the dimensions of charge in the electrostatic and electromagnetic systems of units.

6. What is the magnetic field at the nucleus of a hydrogen atom, due to the motion of the electron in the first circular orbit? See Sec. 11.9.

7. Two large circular coils, parallel and separated by a distance equal to a coil radius, are often used to produce a uniform magnetic field at a point on the axis halfway between them. Derive the value of this field.

8. Derive an expression for the torque on a coil carrying current, in terms of the magnetic moment of the coil.

9. A solenoid 1 m in length and 10 cm in diameter contains 1,000 turns. What is the field at the center of the solenoid for a current of 2 amp?

10. For the solenoid in Prob. 9, find the field along the axis at points 10, 20, 30, 40, and 50 cm from the center.

★11. A circular coil of 5,000 turns is wound in a flat pancake of inner radius 10 cm and outer radius 30 cm. What is the field at the center of the coil for a current of 0.1 amp through it?

12. A square coil 25 cm on the side and having 10 turns carries a current of 0.5 amp. What are the magnitude and direction of the field at the center of the coil?

CHAPTER 31
MAGNETIC FIELDS AND POLES

31.1. The Field near a Long Magnet. In the neighborhood of a natural magnet of lodestone (magnetite) or near a permanent magnet of iron or steel, there is found to be a force exerted on a moving charge or on a current element. A field of induction **B** therefore exists there. If the magnet is a very long and thin bar, the field near one end is radial and varies inversely with the square of the distance from the end. Near the center of the magnet the field is small. It thus appears that the origin of the magnetic effects is at the ends of the bar or at the *poles* of the magnet. At one end the field is radially outward, and this end is called the *north pole* of the magnet. At the other end the field is radially inward, and this end is called the *south pole*. The terms "north" and "south" derive, of course, from the fact that a magnet free to turn about a vertical axis perpendicular to its length aligns itself approximately north and south, with the north pole in the northward direction.

The field **B** near one pole can be expressed by the equation

$$B = k'' \frac{m}{r^2} \tag{31.1}$$

where r is the distance to the pole, m is a measure of the strength of the magnet, and k'' is a constant, introduced in Sec. 30.5, whose value depends upon the choice of units. The quantity m is called the *pole strength* of the magnet. Equation (31.1) serves to *define* pole strength and the units in which it is measured. If B is expressed in emu, r in cm, and k'' is equal to unity, the equation defines the emu of pole strength m. In practical mks units, r is measured in meters, B in weber/m^2, $k'' = 1/4\pi$, and the mks unit of m is defined. The occurrence of 4π here is the result of the fact that the mks system is a rationalized system; if the factor were omitted here, it would occur in some other, more commonly used equation. In neither the mks nor the electromagnetic system of units does the unit of pole strength have a separate name. Since it has the same dimensions as magnetic flux, the units weber or maxwell may be used, respectively. *Caution should be exercised, however. The total flux from a unit pole in the electromagnetic system is 4π maxwells; the flux from a unit pole in mks units is 1 weber.*

The similarity of Eq. (31.1) to the equation for the electric field near a charge is immediately obvious. The pole strength m therefore plays the role of a magnetic charge that is concentrated at the end of the magnet. In fact Eq. (31.1) was once believed to demonstrate the existence of magnetic charge, but we know now that this is not the case. Magnetic charges do not exist, but the magnetic field of a permanent magnet is the result of moving charges within the magnet, just as the magnetic field near a current-carrying wire is the result of the current directly. Magnetic charges are often a convenient artifice, however. The field of a permanent magnet can, for example, be described easily as the sum of the fields due to two equal and opposite poles at the ends of the magnets. The field due to a number of poles or to a continuous distribution of poles can be calculated in exactly the same way as the electric field due to a distribution of charges is calculated. North poles are conventionally taken as positive poles with positive values of m; south poles are considered negative.

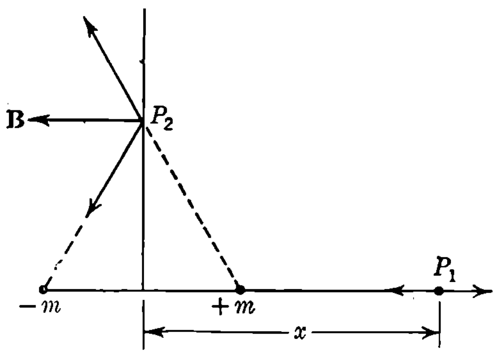


FIG. 31.1. The field of a magnetic dipole.

31.2. The Magnetic Dipole. As an example of the calculation of the magnetic induction field from the pole strength, let us consider a magnetic dipole consisting of two poles of strength $+m$ and $-m$ separated by a distance l, as in Fig. 31.1. The field B_1 at the point P_1 is the sum of two oppositely directed components, and the net field is directed away from the pair of poles and of magnitude

$$\begin{aligned} B_1 &= \frac{k''m}{[x - (l/2)]^2} - \frac{k''\,m}{[x + (l/2)]^2} \\ &= \frac{2xlmk''}{[x^2 - (l^2/4)]^2} \end{aligned} \tag{31.2}$$

At the point P_2 the field B_2 is the sum of two vectors, as indicated in the figure, and is equal to

$$B_2 = -\frac{k''ml}{[y^2 + (l^2/4)]^{3/2}} \tag{31.3}$$

These relations are identical with the expressions for the electric field near an electric dipole [Eqs. (25.7 and 25.8)]. The product ml is called the *magnetic moment* M of the dipole. At large distances the fields become

$$\left.\begin{aligned} B_1 &= \frac{2k''M}{x^3} \\ B_2 &= -\frac{k''M}{y^3} \end{aligned}\right\} \tag{31.4}$$

Equations (31.4) are also identical with Eqs. (30.22), which describe the field B about a current loop. The field, then, of a short magnet of moment $M = ml$ is indistinguishable from the field of a small coil of moment $M = k'iA/k''$.

31.3. Intensity of Magnetization. Experiment shows that a magnetic charge or pole never occurs as an isolated magnetic charge; no isolated poles exist. If a magnet has a north pole of strength $+m$ at one end, the south pole at the other end has a strength $-m$. The *total* magnetic charge on a body is always zero. The magnetic charges in a magnet are, however, polarized, just as the charges in a dielectric are polarized by an applied electric field. The intensity of this polarization, called the *intensity of magnetization, is measured by the magnetic moment per unit volume.* The intensity of magnetization **I** is a vector quantity, just as is the analogous polarization **P** of a dielectric; the direction of **I** is taken to be from the south to the north end of the dipole moment.

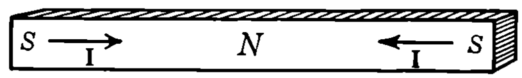


FIG. 31.2. Consequent poles.

The relation between the intensity of magnetization **I** and the pole strength m is easily found. If a magnet of length l and cross-sectional area A is *uniformly* magnetized, then

$$I = \frac{ml}{Al} = \frac{m}{A} \tag{31.5}$$

The units of **I** are consequently the same as the units of **B** [see Eq. (31.1)] and this is the case in both the electromagnetic and mks systems of units.

The picture of magnetization as the polarization of magnetic charges offers an easy explanation of many phenomena. For example, if a magnet is broken in two parts, each part has poles at the ends equal in strength to the poles of the original magnet and hence the magnetization **I** is not altered in the process. Moreover a magnet in the form of a ring can exist, which has no poles at all. If such a ring magnet were broken, however, the pieces would have poles of the strength m given by Eq. (31.5). Magnets are not always uniformly magnetized. Thus, if **I** were in one direction over part of the length of a bar magnet and in the opposite direction in the rest of the length, there would be a south pole, say, at each end of the bar and a north pole at the center, as in Fig. 31.2. Such poles are called *consequent poles.*

31.4. Equivalent Surface Currents in a Magnet. As we have repeatedly emphasized, all magnetic effects are the result of moving electric charges and are not caused by magnetic charges or poles. The intensity of magnetization of a magnet is therefore the result of electric currents within the magnet and indeed within the atoms of the material of the magnet. Although this point of view was first suggested by Ampère in 1820, complete confirmation has been obtained only within the last 20 years. In steel and in the other substances from which permanent magnets can be made, the electric circuit responsible for magnetic effects is the electron itself. Each electron behaves as though it were a spherical distribution of charge that is spinning rapidly about some axis. An electron is thus a small electric circuit and possesses a magnetic moment. In an unmagnetized body the axes of the electrons are oriented at random, and no net magnetic moment exists. In a magnet the electron axes have some preferred orientation, the spins are aligned in more or less the same direction, and magnetization results.

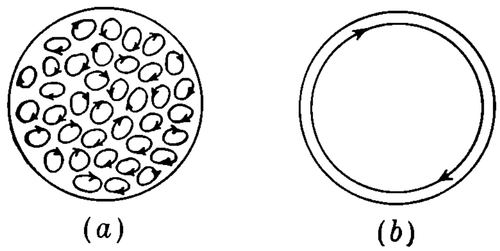


FIG. 31.3. Cross section of a magnet showing the electron currents and the equivalent surface current.

A cross section of a magnet would thus expose a multitude of tiny electric circuits with the current circulating in each in the same direction, as indicated in Fig. 31.3*a*. It should be noted that in the interior of the magnet each part of one of the elementary circuits is adjacent to a part of a neighboring circuit in which the current is flowing in the opposite direction. The two opposite currents then produce no net effect, and the whole assemblage of small circuits can be replaced by a current flowing around the surface of the magnet, as indicated in Fig. 31.3*b*. The relation between the strength of surface currents and the intensity of magnetization can be found easily.

Let di be the surface current flowing in a length dl around the magnet, as indicated in Fig. 31.4. The magnetic moment dM of dl is, from Eq. (30.21),

$$dM = \frac{k'}{k''} A\, di \qquad (31.6)$$

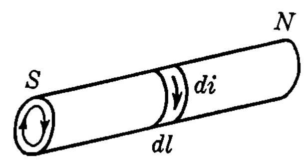


FIG. 31.4. The surface current on a magnet.

where A is the cross-sectional area of the magnet. The vector **I** has therefore the magnitude

$$I = \frac{dM}{dV} = \frac{dM}{A\,dl} = \frac{k'}{k''}\frac{di}{dl} \qquad (31.7)$$

The quantity di/dl is the surface current density j_s, and

$$I = \frac{k'}{k''} j_s \qquad (31.8)$$

The similarity of this equation to the solenoid equation (30.28) will be noted on recalling that $k'' = 1/4\pi$ in mks units.

The magnetic state of a material can therefore be expressed in any one of three ways, (1) by the intensity of magnetization **I**, (2) by the pole strength per unit area, or (3) by the surface current per unit length.

31.5. The Torque on a Magnet in a Field. It is instructive to calculate the torque exerted on a magnet in an external field **B**. Let us assume that the magnet has a length l and cross-sectional area A, that it is uniformly magnetized, and that the external field makes an angle ϕ with the axis of the magnet, as indicated in Fig. 31.5. Each element of length dl of the magnet is equivalent to a single turn of area A carrying a current $j_s dl$. The torque dL on this element is, by Eq. (29.19),

$$dL = -j_s dl\, AB \sin\phi$$

since $\sin\phi = \cos\theta$ of Eq. (29.19), and the negative sign means that the torque is in the direction to tend to reduce ϕ. This torque is a *couple* (Sec. 3.6), and hence the total torque L is obtained by integration over the length of the magnet, or

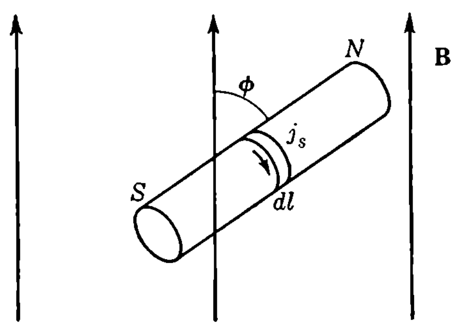


FIG. 31.5. The torque on a magnet.

$$L = -j_s lAB \sin\phi \tag{31.9}$$

The torque can be expressed in terms of I by use of Eq. (31.8), which gives

$$L = -\frac{k''}{k'} IlAB \sin\phi \tag{31.10}$$

The quantity IlA is, however, the total magnetic moment M of the magnet; hence

$$L = -\frac{k''}{k'} MB \sin\phi \tag{31.11}$$

Since the magnetic moment M is related to the pole strength at the ends of the magnet by

$$M = ml$$

the torque can also be written

$$L = -ml\frac{k''}{k'} B \sin\phi \tag{31.12}$$

The last form for the torque on the magnet is very significant since it enables us to evaluate the force F on a magnetic pole. The torque on the magnet caused by forces in the direction of **B** exerted at the ends must be

$$L = -Fl \sin \phi$$

and therefore the force on a pole of strength m in a field B is

$$F = m \frac{k''B}{k'} \tag{31.13}$$

***31.6. The Magnetic Intensity.** It is customary to define a second magnetic vector **H,** which is somewhat analogous to the electric displacement **D.** In the preceding section it was seen that the torque on a magnet in an external field is equal to the torque exerted by a force **F** acting at each end of the magnet, the magnitude of the force being given by Eq. (31.13). The magnetic intensity **H** is defined as the force on a unit magnetic pole; then for this case

$$\mathbf{H} = \frac{\mathbf{F}}{m} = \frac{k''}{k'} \mathbf{B} \tag{31.14}$$

The torque on the magnet can then be written in terms of H as

$$L = H \sin \phi \, ml = H \sin \phi \, M \tag{31.15}$$

It must be emphasized that **B** and **H** above are the values for the field applied to the magnet; they do not refer to the field of the magnet. The applied field **B** might be the result of currents flowing in wires or of the magnetization of another magnet.

In the neighborhood of a magnetic pole, *i.e.*, near the end of a long magnet, B is given by Eq. (31.1). The value of H is therefore, from Eq. (31.14),

$$H = \frac{(k'')^2}{k'} \frac{m}{r^2} \tag{31.16}$$

If a second pole of strength m' were present, it would experience a force in this field of

$$F = Hm'$$

The magnetic force exerted on one magnetic pole by the other is

$$F = \frac{(k'')^2}{k'} \frac{mm'}{r^2} \tag{31.17}$$

This equation is analogous to Coulomb's law for the force between electric charges. It would be entirely logical to begin the study of electromagnetism with Eq. (31.17) as the definition of pole strength, rather than Eq. (31.1). In emu, Eq. (31.17) becomes

$$F = \frac{mm'}{r^2} \qquad \text{emu} \tag{31.18a}$$

and in practical units

$$F = \frac{mm'}{4\pi\mu_0 r^2} \qquad \text{mks units} \tag{31.18b}$$

The field due to a single pole is

$$H = \frac{m}{r^2} \qquad \text{emu} \tag{31.19a}$$

$$H = \frac{m}{4\pi\mu_0 r^2} \qquad \text{mks units} \tag{31.19b}$$

If a distribution of poles is present, then the vector sum of the contribution from each pole must be taken in exactly the same way as the field **E** is computed from a distribution of electric charges. If electric currents are present, there will also be a contribution to **H** from these. From Eqs. (30.1) and (31.14),

$$dH = k'' \frac{i\, dl \sin\theta}{r^2} \tag{31.20}$$

If both poles and currents are present, a sum must be made to find the total value of **H.** It must be remembered that a *vector* sum is to be taken. Equations (31.20) and (31.16) are therefore the definitions of **H** in terms of the currents and poles present. It will be recalled that the field **B** is defined by its action on a moving charge [Eq. (29.2)]. The relations between **B** and **H** within magnetic materials will be discussed in Chap. 33.

Since H is force per unit pole, the practical units of **H** can be taken as newtons per weber. It is more customary to use the unit derived from the relation between H and i. From Eq. (31.20) it is evident that *H is measured in amperes per meter*. Sometimes the units of H are called "ampere turns" per meter, since the current i is often flowing in a coil, but in this text we shall use the simpler expression. The emu of H might be either abamperes per centimeter or dynes per maxwell. A special name has been given, however, to a unit field intensity; it is called the *oersted.* In free space in the electromagnetic system **B** and **H** have the same numerical value, but this is not true in the practical mks system. The relation between the units of **H** is

$$1 \text{ oersted} = \frac{10^3}{4\pi} \frac{\text{amp}}{\text{m}} \tag{31.21}$$

This relation can be derived from Eq. (31.20).

***31.7. The Magnetometer.** If a magnet of moment **M** is placed in a field of intensity **H,** the torque is given by Eq. (31.15). If the angle ϕ is small, L can be approximated by

$$L = -MH\phi$$

If the magnet is suspended so that it is free to turn, its motion is governed by the equation

$$I' \frac{d^2\phi}{dt^2} = -MH\phi \tag{31.22}$$

where I' is the moment of inertia of the magnet. The magnet oscillates with SHM about $\phi = 0$ (cf. Sec. 12.8). The frequency of oscillation is

$$f = \frac{1}{2\pi} \sqrt{\frac{MH}{I'}} \tag{31.23}$$

and is proportional to the square root of the magnetic field. The strength of one magnetic field can be found in terms of the strength of another by observation of the frequencies of oscillation of the same magnet in the two fields. If the two cases are denoted by the subscripts 1 and 2,

$$\frac{f_1}{f_2} = \sqrt{\frac{H_1}{H_2}} \tag{31.24}$$

An instrument that utilizes an oscillating magnet in this manner is called a *magnetometer*.

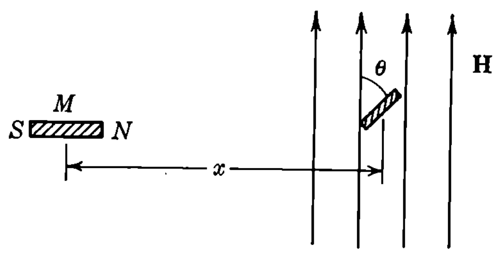


FIG. 31.6. Measurement of H by a magnetometer.

In order to find the absolute value of the field H it is necessary to know also the moment M of the magnet. With the aid of a second suspended magnet this can be determined. Suppose the second magnet is suspended in the field H and the magnet of moment M brought near it, as shown in Fig. 31.6. The suspended magnet will set itself in the direction of the resultant field at an angle θ with respect to the original field. If the field from the magnet of moment M is H_M, then

$$\cot \theta = \frac{H}{H_M}$$

From the first of Eqs. (31.4),

$$H_M = \frac{(k'')^2}{k'} \frac{2M}{x^3}$$

where x is the distance between the centers of the magnets. The observable angle θ is then given by

$$\cot \theta = \frac{k'}{(k'')^2} \frac{x^3 H}{2M} \tag{31.25}$$

in terms of the ratio H/M. Now the frequency f of oscillation in the field H alone is given by Eq. (31.23). When f is measured, both H and M are determined, for they can be computed by simultaneously solving Eqs. (31.23) and (31.25).

Magnetometer measurements are most suitable for weak fields. They are often employed in the measurement of the field of the earth and in geophysical prospecting.

31.8. Terrestrial Magnetism. Since the north pole of a compass needle points northward, the lines of magnetic field **B** or **H** of the earth run from south to north. A compass needle points, however, not to true north but rather at a small angle from the meridian. This angle is called the *declination*. If a long magnet is suspended about a horizontal axis, the equilibrium position, in the Northern Hemisphere, is with the north pole downward. The angle with the horizontal is called the angle of *dip*, or *inclination*. The declination and dip, as well as the intensity of the field, vary over the earth's surface. To a first approximation the field at the surface of the earth is that of a magnetic dipole near the center of the earth and inclined at an angle of about 17° with the earth's axis, as indicated by PP' in Fig. 31.7. The dipole moment is directed from north to south. The points on the earth's surface where the dip is 90° are called the *magnetic poles* of the earth. The north magnetic pole is in northern Canada at lat 70°N, long 96°W, and the south magnetic pole is at lat 72°S, long 157°E. The *magnetic equator* ee' is the line on the earth's surface where the dip is zero. It varies from lat 15°S in South America to 20°N in North Africa. Since the equivalent dipole is not exactly at the center of the earth, the intensity of the field along the magnetic equator is not constant but varies slightly from one side of the earth to the other. The intensity of the field at the magnetic equator is roughly 0.4 oersted and horizontal. At the magnetic poles it is vertical and twice as strong.

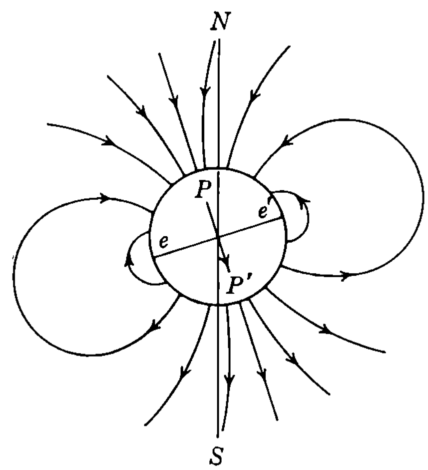


FIG. 31.7. The magnetic field of the earth.

One source of the field about the earth is the currents that flow in the upper atmosphere. At heights above about 80 km, free charges, electrons and positive ions, exist and form what is known as the *ionosphere*. In the ionosphere, convection currents result from the heating and cooling of the atmosphere by the sun, and such ionospheric currents produce a magnetic field. These convection currents vary with diurnal and seasonal periods, and consequently there is a diurnal and seasonal variation of the earth's field. At times, the convection currents are greatly disturbed, and *magnetic storms* occur. Magnetic storms are apparently connected with sunspot activity, and may be caused by a change in the number of charged particles in the ionosphere.

Atmospheric currents produce, however, only a small part of the total magnetic field. Most of the field is produced by something inside the earth. The causes of the field are not known, and very little about them has been surmised. The earth's field constitutes one of the outstanding unsolved problems of science. The field from the interior also appears to vary with time, but with much longer periods. These changes are called "secular" changes. They are most easily described by the statement that the north magnetic pole is moving about the North Pole of the earth with a period of about 1,000 years. The variation of the field over the surface of the earth does not give much information about the field within the interior. If the earth were uniformly magnetized, the external field would be exactly the same as the field from a dipole at the center of the earth.

PROBLEMS

1. A magnet in the form of a ring has a small air gap. If the pole strengths on the faces of the gap are $+m$ and $-m$ and the cross-sectional area of the magnet is A, what is the field B within the gap?

2. Two thin magnets, each 10 cm long, are parallel and separated by 10 mm. What force does one magnet exert upon the other, if the pole strengths at the magnet ends are 10 emu?

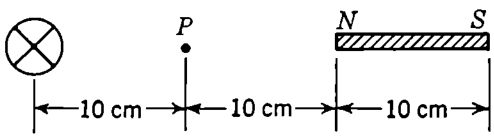


FIG. 31.8. Diagram for problem 3.

★3. A long wire carrying a current of 10 amp is perpendicular to a magnet with a pole strength of 10^{-5} mks unit. What is the field H at the point P in Fig. 31.8?

★4. A steel magnet, density 7.8 gm/cm², 10 cm long, and of cross-sectional area 1 mm², has poles of strength 5 emu. If the magnet is suspended in the earth's field of 0.3 oersted, what is the period of oscillation?

5. Assume that the earth's field is 0.4 oersted at the magnetic equator. What is the magnetic moment of the dipole at the center of the earth that would produce this field?

★6. A solenoid 50 cm long and having 100 turns carries a current of 2 amp. What are the values of H at the center of the solenoid in emu and in mks units?

7. What is the intensity of magnetization of the magnet of Prob. 4? What is the equivalent surface current density?

★8. What is the potential energy as a function of θ of a magnet of moment M set at an angle θ with a field H? [HINT: Integrate Eq. (31.15).]

★9. Find the dimensions of B and H in mks units. Show that for free space $B = \mu_0 H$.

★10. The period of oscillation of a suspended magnet in the earth's field of 0.3 oersted is 0.5 sec. When suspended inside a long solenoid the period is 0.1 sec. What is the strength of the field in the solenoid?

★11. A magnet suspended in the earth's field sets itself at an angle of 20° from magnetic north when a second magnet is placed 50 cm from it as shown in Fig. 31.6. If the earth's field is 0.3 oersted, what is the moment of the second magnet?

★12. A magnet of 1 cm² cross-sectional area and 5 cm long has an intensity of magnetization of 1,000 gausses. What is the surface current density around the magnet? Express the result in both emu and mks units.

CHAPTER 32

INDUCED ELECTROMOTIVE FORCE

32.1. Electric Fields Produced by Motion. The force on a charge moving in a magnetic field has one consequence of extreme importance. Suppose a vertical conductor is moving to the right, as in Fig. 32.1, with a velocity **v**, and that there is a magnetic field **B** directed inward to the paper. The electrons and positive ions in the conductor experience a force, given by Eq. (29.1), that tends to make the positive ions move upward and the electrons move downward. The positive ions are held in place by the interatomic forces, but the electrons are free to move. A separation of charges therefore results as indicated in the figure. The separation continues until an electric field E is produced that is just sufficient to counteract the magnetic force on the charges.

$$E = \frac{F}{q} = vB \qquad (32.1)$$

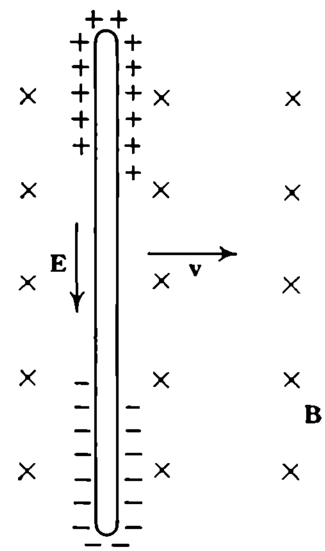


FIG. 32.1. A conductor moving in a magnetic field.

If the field **B** is uniform and **v** is constant, then the electric field produced is constant over the length of the conductor. An emf $d\mathcal{E}$ directed upward is thus produced in each element of length dl, which is (*cf.* Sec. 27.3)

$$\boxed{d\mathcal{E} = vB\,dl} \qquad (32.2)$$

and the total emf $\mathcal{E}$ is

$$\mathcal{E} = \int d\mathcal{E} = vBl \qquad (32.3)$$

In a more general case, let dl lie along the x axis and choose **B** to lie in the xy plane (see Fig. 32.2). If ϕ is the angle between the z axis and **v**, and θ that between **B** and the x axis, Eq. (32.2) is generalized to

$$d\mathcal{E} = (v \cos \phi)\ (B \sin \theta)dl \qquad (32.4)$$

The emf $\mathcal{E}$ is called an *induced* emf. It is produced by virtue of mechanical motion and hence forms the basis of all electric generators. The process of induction was discovered about a century ago by Faraday and by Henry[1] independently. Energy is expended in the separation of the charges in the conductor, and this energy is supplied from the mechanical power that moves the conductor. Since it is important to understand the situation clearly, let us consider what happens when the ends of the moving conductor are connected by a wire. The connecting wire can be stationary and the conducting rod imagined as sliding along the wire, as in Fig. 32.3. The induced emf causes an induced current i to flow in the circuit as indicated. The current flows in the direction of the emf upward along the moving conductor. There is a force, however, on a current

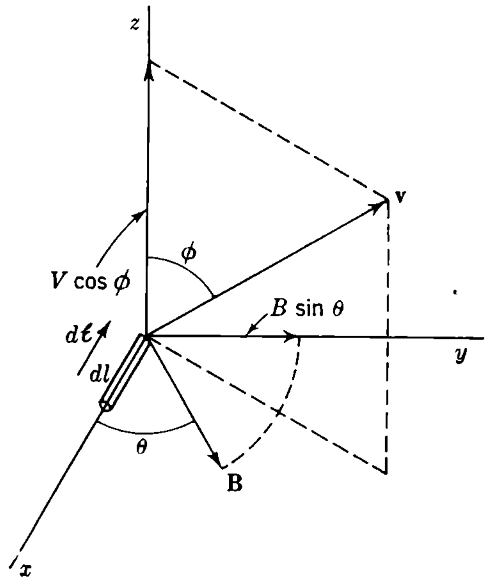


FIG. 32.2. Induced electromotive force.

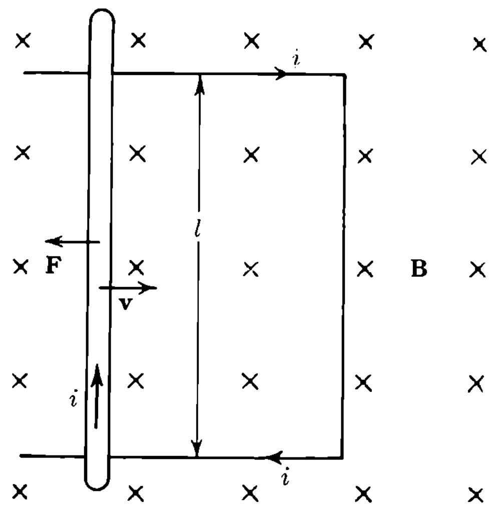


FIG. 32.3. An induced current.

flowing in a magnetic field that is given by Eq. (29.14). The force **F** is directed to the left in the figure and tends to oppose the motion. Its magnitude is

$$F = iBl \tag{32.5}$$

As the rod is moved, mechanical power is expended in the amount

$$P = Fv = iBlv \tag{32.6}$$

The mechanical power is transformed to electric power, which is

$$P = \mathcal{E}i \tag{32.7}$$

If Eqs. (32.6) and (32.7) are combined, we obtain again Eq. (32.3).

32.2. Faraday's Law of Induction. The emf produced in the moving conductor can be regarded from still another point of view. As the conductor moves, the total flux through the circuit changes. The change

[1] Joseph Henry (1797–1878), an American physicist.

in area dA for a motion $v\,dt = dx$ is

$$dA = l\,dx = lv\,dt$$

and hence the change in flux $d\Phi$ that results is

$$d\Phi = BdA = -Blv\,dt$$

or

$$\boxed{\frac{d\Phi}{dt} = -Blv = -\mathcal{E}} \qquad (32.8)$$

This result is known as *Faraday's law of induction: the induced emf around a circuit is equal to the negative rate of change of magnetic flux through the circuit.* The negative sign in Eq. (32.8) arises from the convention for the positive direction of $\mathcal{E}$. The emf is taken as positive if the direction of $\mathcal{E}$ around the circuit (clockwise in this case) is the direction of rotation of a right-handed screw advancing in the direction of the field **B** (into the paper here). A negative rate of change in Φ, for the present case, thus produces a positive $\mathcal{E}$. If the stationary wire were to the left of the moving conductor, as in Fig. 32.4, the rate of change of Φ would be positive; but the current i now would flow counterclockwise, and $\mathcal{E}$ would be negative.

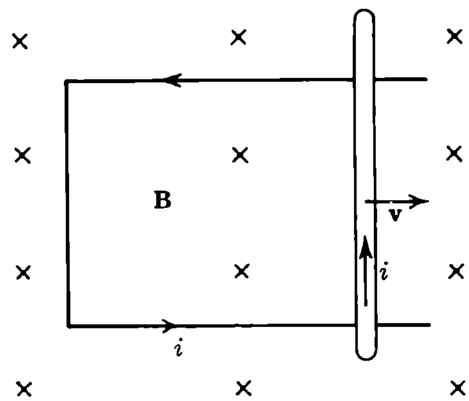


FIG. 32.4. A negative induced emf.

The direction of the induced emf can often be obtained most easily from *Lenz's law: The direction of an induced emf is such as to oppose through its magnetic effects the change producing the emf.* This law may be variously interpreted. For example, in the circuit of Fig. 32.4 the force on the moving rod is opposite to the velocity. In terms of the flux through the circuit the law may be applied as follows: The motion of the rod results in an increase in the flux; the emf must therefore produce a current in such a direction as to compensate this change. The field of the current must be outward from the paper, and hence the current must flow in a counterclockwise sense. Lenz's law is, of course, an expression of the principle of conservation of energy, not of the perversity of nature. When electric energy appears, mechanical energy disappears.

Faraday's law of induction is valid whether or not the change in flux is the result of the motion of the circuit. Thus a change in flux through a circuit caused by the approach of a magnet or by the change in the current of a neighboring circuit produces an emf given by Eq. (32.8). It should also be emphasized that the circuit containing the emf need not be a metallic circuit. Faraday's law is applicable to any closed curve.

Equation (32.8) is valid in both practical mks units and electromagnetic cgs units. If practical units are used, $\mathcal{E}$ is in volts if $d\Phi/dt$ is in webers

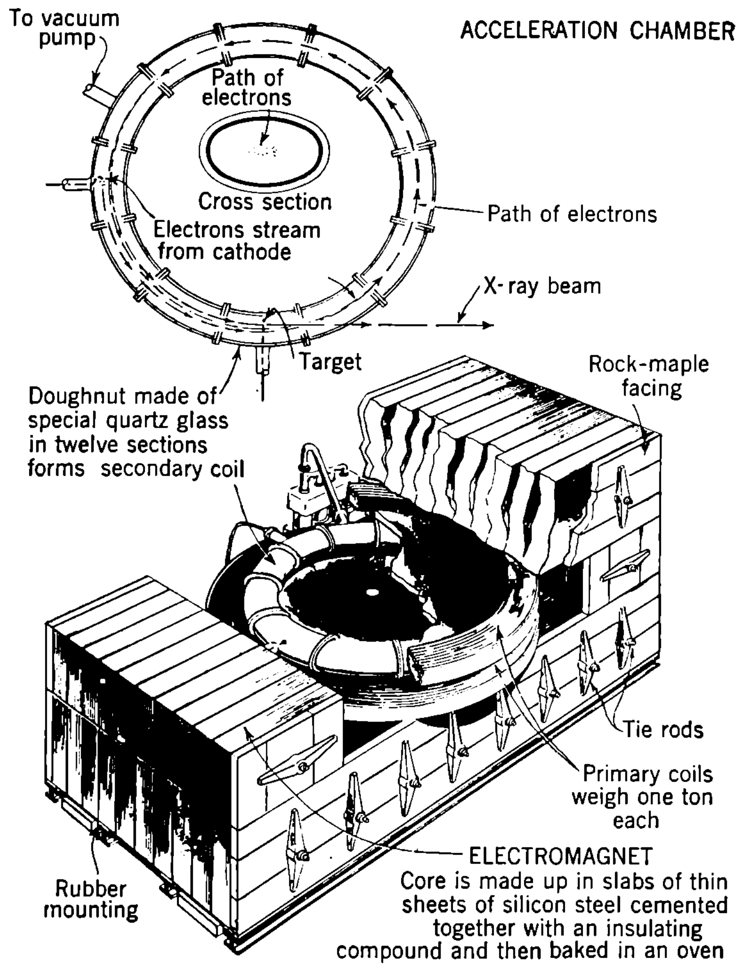


FIG. 32.5. Construction and operation of the 100-Mev betatron. (*Courtesy of Popular Science and The General Electric Company.*)

per second. In emu, $d\Phi/dt$ has the units of gauss cm^2 per second, and $\mathcal{E}$ is expressed in electromagnetic volts, or *abvolts*. Since the ratio of esu to emu of emf is $c = 3 \times 10^{10}$ cm/sec and 1 electrostatic volt = 300 volts, approximately, we have

$$1 \text{ esu} = 300 \text{ volts} = 3 \times 10^{11} \text{ emu}$$

and

$$1 \text{ emu of emf} = 1 \text{ abvolt} = 10^{-8} \text{ volt} \qquad (32.9)$$

32.3. The Betatron. The betatron, or induction electron accelerator, is a device that utilizes the emf produced by a change of flux to accelerate electrons to high energies. It was recently invented by D. W. Kerst, and several large betatrons are now in operation. The changing flux is produced by a large electromagnet excited by an alternating current. The field of the magnet serves at the same time to bend the electrons in a circular path and confine them to the region of changing flux. The arrangement is indicated schematically in Fig. 32.5. The electrons move within a doughnut-shaped vacuum chamber placed between the poles of the large electromagnet. The coils carry the alternating current. The time of revolution of an electron in its circular path is much smaller than the period of alternation of the field. As the field grows from a small value to its maximum, the change of flux within the path of the electron produces an emf whose magnitude is given by Eq. (32.8). The energy received by the electrons in one revolution is small, but a very large number of revolutions are possible, and electrons of high energy result. The electrons can be deflected from their path at any time by a surge of current through the magnet coils, and the high-energy electrons can be directed to a target.

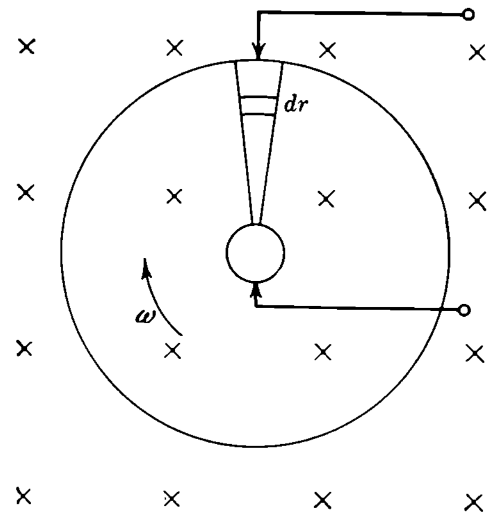


Fig. 32.6. The disk dynamo.

A betatron recently built by the General Electric Company produces electrons of energies up to 100 million electron volts. The vacuum chamber is constructed of sections of glass tubing and is 74 in. in outside diameter and 58 in. in inside diameter. The frequency of alternation for the magnet current is 60 cycles/sec. Electrons are accelerated by an electric field to 50,000 ev and shot into the tube tangential to a circular path 66 in. in diameter. At each revolution an electron gains about 400 volts and makes about 250,000 revolutions while the field is changing. The total energy gained is therefore $400 \times 250{,}000 = 10^8$ ev. The relativistic change of mass with velocity does not affect the operation of the betatron, since no resonance is involved, as in the cyclotron.

***32.4. The Disk Dynamo.** One of the first electric generators was invented by Faraday and is called the *disk dynamo.* The dynamo consists simply of a metal disk that can be rotated in a magnetic field about an axle perpendicular to the disk. Two sliding contacts, or brushes, one on the rim of the disk, the other on the axle, are the terminals of the source of emf. The dynamo is shown schematically in Fig. 32.6. The small section of the disk lying between the radii r

and $r + dr$ is moving in the field B with a velocity $v = \omega r$, where ω is the angular velocity of the disk. The emf generated in the section is

$$d\mathcal{E} = B\omega\, rdr$$

The total emf between the two brushes is therefore

$$\mathcal{E} = \int_a^b B\omega r dr \tag{32.10}$$

where a and b are the radii of the axle and disk, respectively. Evaluation of the integral results in

$$\mathcal{E} = \frac{B\omega}{2}(b^2 - a^2) \tag{32.11}$$

Dynamos of this type are not practical machines since the emf that can be generated is too small.

Worked Example. A disk dynamo rotates with a speed of 3,600 rpm in a field of 4,000 gausses. The radius of the disk is 10 cm, that of the axle 1 cm. What emf is generated? The field in practical units is 0.4 weber/m^2. The angular velocity

$$\omega = \frac{3{,}600}{60} 2\pi = 120\pi \text{ radians/sec}$$

From Eq. (32.11),

$$\mathcal{E} = \frac{0.4 \times 120\pi}{2}(0.01 - 0.0001) \approx \frac{0.4 \times 1.2\pi}{2} = 0.75 \text{ volt}$$

In emu,

$$\mathcal{E} \approx \frac{4{,}000 \times 120\pi}{2} \times 99 = 7.5 \times 10^7 \text{ abvolts} = 0.75 \text{ volt}$$

32.5. Dynamos. More practical electric generators can be made that utilize the emf induced in a coil rotating in a magnetic field. In the simplest case a coil of N turns, each of area A, rotates in a uniform field **B** about an axis in the plane of the coil, as indicated in Fig. 32.7. If θ is the angle of rotation, the flux Φ through the coil is

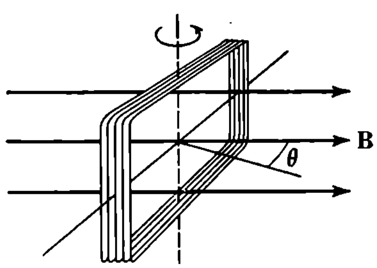


FIG. 32.7. A coil rotating in a field.

$$\Phi = BA \cos\theta \tag{32.12}$$

The emf developed in the coil is therefore

$$\mathcal{E} = -N\frac{d\Phi}{dt} = -NBA\frac{d}{dt}(\cos\theta)$$

If the angular velocity of the coil is $\omega = d\theta/dt$, a constant, then

$$\mathcal{E} = NBA\omega \sin\theta$$

If we assume that $\theta = 0$ at $t = 0$, then $\theta = \omega t$ and the emf is

$$\mathcal{E} = NBA\omega \sin \omega t \tag{32.13}$$

The emf is therefore zero at $t = 0$ and varies as a sine curve, reversing direction each half cycle, like the displacement in SHM. The emf is hence called an *alternating* emf. The maximum value is

$$\mathcal{E}_{max} = NBA\omega$$

and

$$\mathcal{E} = \mathcal{E}_{max} \sin \omega t \tag{32.14}$$

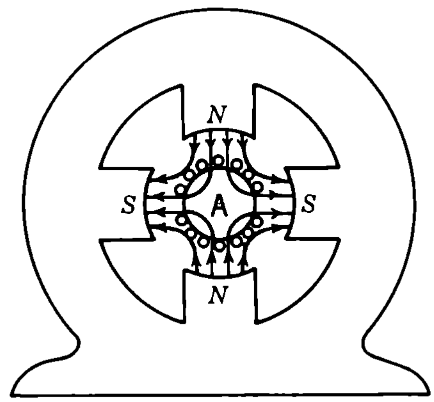


FIG. 32.8. A four-pole generator.

In order to utilize this emf it is necessary to make contact with the coil. This is accomplished by metal rings, called *slip rings*, attached to the shaft. Upon these rings are sliding contacts, usually of carbon, called brushes. The ends of the coil are connected to the slip rings, so that the emf is developed between the brushes. To obtain a large value of **B** the coil is wound upon a soft-iron *armature*. The frame of the generator is usually also wound with *field coils*, through which a current is passed to excite the generator. In very small generators the field is supplied by a permanent magnet. Most large generators have a more complicated field structure, with several pairs of poles. A four-pole machine is shown schematically in Fig. 32.8. The emf induced in the coil wound on the armature A is zero at four positions in each rotation of the shaft, instead of two as in the two-pole machine. The frequency f of the alternating emf in a machine with n poles rotating R turns per sec is

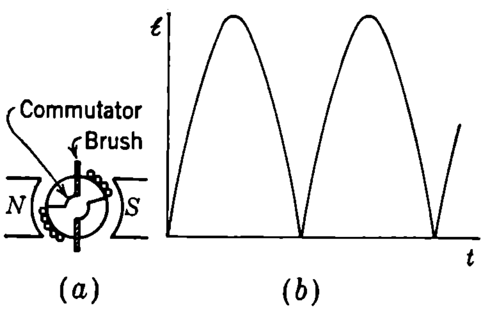


FIG. 32.9. (*a*) D-c generator with a two-segment commutator; (*b*) the emf produced.

$$f = \frac{\omega}{2\pi} = \frac{n}{2} R \tag{32.15}$$

Thus a four-pole machine at a speed of 1,800 rpm produces an alternating emf of the standard frequency, 60 cps.

Direct-current generators are constructed with several coils, and the slip rings are split into segments that provide switching from one coil to another. The slip ring is then called a *commutator*. For each coil there is a pair of segments on the commutator. For a single coil the arrangement as well as the variation of the emf with time is shown in Fig. 32.9.

If two coils are used whose planes are at right angles to each other and the emfs induced by the coils are connected in series, the resultant emf is as shown in Fig. 32.10. The "ripple" in the emf is much reduced. By a further increase of the number of coils and of the number of commutator segments a rather steady emf is obtained.

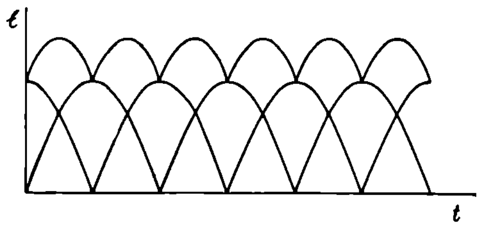


FIG. 32.10. The emf from two coils in series. Upper curve is sum of lower two.

The d-c motor described in Sec. 29.10 is nothing more than a dynamo. The conversion of energy is from mechanical to electric energy in the dynamo. The motor performs the conversion in the opposite direction.

***32.6. The Search Coil and Electromagnetic Damping.** By the use of the induced emf it is possible to measure a magnetic field. A small coil of N turns, each of area A, is connected to a ballistic galvanometer. If the coil is removed suddenly from a magnetic field B, there is an induced emf in the circuit and a current i flows through the galvanometer. If R is the resistance of the circuit of the coil and galvanometer, then

$$i = \frac{\mathcal{E}}{R} = -\frac{1}{R}\frac{d\Phi}{dt} = -\frac{NA}{R}\frac{dB}{dt} \tag{32.16}$$

The total charge q that flows through the galvanometer is equal to the time integral of the current, or

$$q = \int i\,dt = -\frac{NA}{R}\int_0^B \frac{dB}{dt}\,dt = \frac{NAB}{R} \tag{32.17}$$

if B decreases to zero. Since the galvanometer deflects an amount proportional to q, the field B can be determined by such a measurement. The coil can, of course, be simply turned over instead of removed from the field. In this case twice the charge flows through the circuit. If the field B can be changed by some other means, the coil can be left stationary.

In the calibration of a ballistic galvanometer for use with a search coil it is necessary to take into account another effect. As the coil of the galvanometer turns through its magnetic field, an emf is induced in it. By Lenz's law the emf will be induced in such a direction as to retard the rotation of the coil. Thus *electromagnetic damping* of the motion is produced. This damping is proportional to the angular velocity of the coil, as in damping of other sorts, and inversely proportional to the resistance of the circuit. Since the deflection of the galvanometer per unit charge depends on the damping, a ballistic galvanometer must be calibrated in a circuit of the same resistance as that in which it is used.

Electromagnetic damping is an important factor in the performance of nearly all electrical instruments. Pivoted-coil meters often have the coil wound on a light metal frame. The emf induced in this single closed turn provides the damping. Galvanometers are usually used with shunt resistors across the terminals, which can be adjusted for the optimum condition of damping. In most work an amount of damping slightly less than that necessary for critical damping is desirable. The current sensitivity of the galvanometer is, of course, altered by the shunt.

32.7. Mutual Inductance. The process of induction provides a means whereby two electric circuits can be coupled together. If the current i_1 in one circuit changes, the magnetic field in the vicinity of the second circuit changes proportionally and an emf $\mathcal{E}_2$ is induced in the second circuit. The *mutual inductance* M_{21} is defined by

$$M_{21} = \frac{\mathcal{E}_2}{di_1/dt} \tag{32.18}$$

and is a measure of the strength of interaction between the circuits.

Suppose, for example, that circuit 1 is a long solenoid of length l and N_1 turns through which a current i_1 is flowing and that circuit 2 is a shorter solenoid of N_2 turns wound coaxially with the first and near its center. The flux through circuit 2 due to the current i_1 is, by Eq. (30.24*a*),

$$\Phi_{21} = k' \frac{4\pi i_1 N_1}{l} A$$

where A is the area of solenoid 1 and k' the factor to take care of units. It will be recalled that, in the practical system, $k' = \mu_0/4\pi$ and, in the electromagnetic system, $k' = 1$. If i_1 changes, the emf in circuit 2 is

$$\mathcal{E}_2 = N_2 \frac{d\Phi_{21}}{dt} = k' \frac{4\pi N_1 N_2 A}{l} \frac{di_1}{dt}$$

From the definition the mutual inductance is therefore

$$M_{21} = k' \frac{4\pi N_1 N_2 A}{l} \tag{32.19}$$

In practical units,

$$M_{21} = \mu_0 \frac{N_1 N_2 A}{l} \tag{32.20}$$

and in emu,

$$M_{21} = \frac{4\pi N_1 N_2 A}{l} \tag{32.21}$$

The practical unit of inductance is called the *henry*. From Eq. (32.18) *a mutual inductance of one henry exists between two circuits when an emf of one volt is produced in one circuit by a change of current in the other circuit of one ampere per second.* The henry has the same dimensions, therefore, as one ohm times one second. The customary unit of μ_0 is expressed in terms of the henry; it is evident from Eq. (32.20), since N_1 and N_2 are pure numbers, that μ_0 has the units henrys per meter. We see also from Eq. (32.21) that in the electromagnetic system the unit of inductance is the centimeter.

If the circuit 2 is a closed circuit with a current i_2, any change in i_2 produces an emf $\mathcal{E}_1$ in circuit 1. The emf $\mathcal{E}_1$ is similarly proportional to the rate of change of i_2; therefore a mutual inductance can also be defined by the equation

$$M_{12} = \frac{\mathcal{E}_1}{di_2/dt} \qquad (32.22)$$

It can be shown, although not simply, that in *all cases*

$$\boxed{M_{12} = M_{21}} \qquad (32.23)$$

In what follows we shall therefore drop the subscripts and write the mutual inductance simply as M.

Since the flux Φ_{21} through circuit 2 produced by a current i_1 in circuit 1 is proportional to i_1, M can also be expressed as

$$M = \frac{N_2(d\Phi_{21}/dt)}{di_1/dt} = \frac{N_2 d\Phi_{21}}{di_1} = \frac{N_2\Phi_{21}}{i_1} \qquad (32.24)$$

or alternatively, because of Eq. (32.23),

$$\boxed{M = \frac{N_1\Phi_{12}}{i_2}} \qquad (32.25)$$

The mutual inductance between two circuits can be measured easily by a ballistic galvanometer. If the galvanometer is connected in circuit 2, the charge through it is

$$q_2 = \int i_2 dt = \int \frac{\mathcal{E}_2 dt}{R_2} = \frac{M}{R_2}\int \frac{di_1}{dt}\, dt = \frac{M}{R_2}\,\Delta i_1 \qquad (32.26)$$

where $\Delta\, i_1$ is the total change of the current in circuit 1.

32.8. Self-inductance. Just as a changing current in one circuit produces an emf in another circuit, an emf is also produced in the circuit in which the current changes. The magnitude of the emf depends on the shape of the circuit and the value of di/dt. The *self-inductance* L is defined by a relation similar to Eq. (32.18),

$$L = -\frac{\mathcal{E}}{di/dt} \qquad (32.27)$$

The negative sign in this equation arises from the fact that $\mathcal{E}$ is a *counter* emf. By Lenz's law the direction of $\mathcal{E}$ is such as to oppose the change in current. Thus, if the current increases, di/dt is positive, and $\mathcal{E}$ is negative. The units of self-inductance L are henrys, the same as for mutual inductance. By analogy with Eqs. (32.24) and (32.25), L can also be written as

$$L = \frac{N\Phi}{i} \qquad (32.28)$$

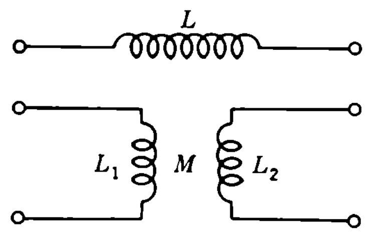


FIG. 32.11. Circuit symbols for inductances.

The self-inductance of a long solenoid or a toroid can be calculated from Eq. (30.24*a*),

$$\Phi = k' \frac{4\pi i N}{l} A$$

and therefore

$$L = k' \frac{4\pi N^2 A}{l} \qquad (32.29)$$

In emu, $k' = 1$; in practical units, $k' = \mu_0/4\pi$, and

$$L = \mu_0 \frac{N^2 A}{l} \qquad (32.30)$$

A coil possessing appreciable self-inductance is called an *inductor*, or an *inductance*. It can be represented in circuit diagrams as shown in Fig. 32.11. The symbol for two coils with mutual inductance is also indicated.

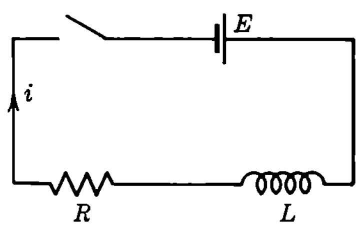


FIG. 32.12. A circuit with inductance.

***32.9. The Growth of Current in an Inductance.** It is of interest to study in more detail the retarding action of the counter emf. Let us consider a simple circuit consisting of an inductor of inductance L, a resistor of resistance R, and a battery of emf E in series with a switch, as in Fig. 32.12. The value of R is meant to include not only the resistance of the resistor but also the internal resistance of the cell and the resistance of the coil with inductance. Let us suppose that the switch is closed at $t = 0$ and no current is flowing before the switch is closed. At any time the total emf must be equal to the iR drop across the equivalent resistor, or

$$E + \mathcal{E} = iR \tag{32.31}$$

From Eq. (32.27) the value of $\mathcal{E}$ can be substituted, and with a rearrangement of terms, we have the differential equation for the current,

$$L\frac{di}{dt} + Ri = E \tag{32.32}$$

The variables can be separated in this equation, and

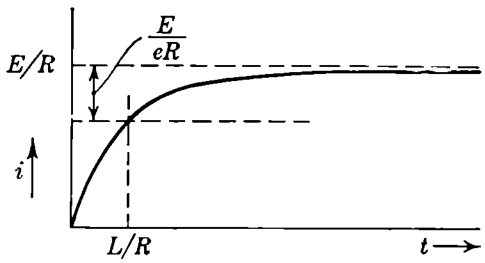


FIG. 32.13. The growth of current in an inductance.

$$\frac{di}{E - Ri} = \frac{dt}{L}$$

Multiplication of both sides of this equation by $-R$ and integration result in

$$\ln (E - Ri) = -Rt/L + \ln A$$

where $\ln A$ is the constant of integration. In exponential form the relation is

$$E - Ri = Ae^{-(R/L)t}$$

The constant A is determined by the initial condition, $i = 0$, $t = 0$; hence

$$A = E$$

The value of i can therefore be written as

$$i = \frac{E}{R}[1 - e^{-(R/L)t}] \tag{32.33}$$

which is the solution of the differential equation. We can see that i increases from zero and approaches a steady value E/R for large values of t. A graph of i vs. t is shown in Fig. 32.13. At the time $t = L/R$ the current is within $1/e$, or 37 per cent, of its final value E/R. This time is called the *time constant* of the circuit. The variation of the current through an inductance is therefore very much like the variation of the potential across a capacitor charged through a resistor.

The power relations in this circuit are also interesting. The battery is supplying power at the rate Ei, and this, by Eq. (32.32), is

$$P = Ei = Li\frac{di}{dt} + Ri^2 \tag{32.34}$$

The last term on the right represents the power dissipated in the resistor, and the first term on the right must therefore be power supplied to the inductance. The inductance is storing energy, and this stored energy W is growing at the rate given by

$$\frac{dW}{dt} = Li\frac{di}{dt} \tag{32.35}$$

The total energy can be found by integrating this; and since $W = 0$ when $i = 0$,

$$W = \tfrac{1}{2}Li^2 \tag{32.36}$$

The energy W can be thought of as stored in the magnetic field surrounding the inductance. It reappears and is given back to the electric circuit when the current decreases. If the switch in Fig. 32.12 is opened after a current i is established, the energy in the magnetic field partly reappears in the spark made when the circuit is broken.

★32.10. Inductances in Series. If a circuit contains two coils of inductances L_1 and L_2, respectively, the resultant inductance of the coils connected in series is simply the sum of the individual inductances, provided that the mutual inductance M of the two coils is negligible. If M cannot be neglected, then the resultant inductance can be found from a consideration of the emfs induced. Let us call the induced emf, in coil 1, $\mathcal{E}_1$ and the currents in the coils i_1 and i_2. Then

$$\mathcal{E}_1 = L_1\frac{di_1}{dt} \pm M\frac{di_2}{dt} \tag{32.37a}$$

where the $\pm$ sign must be used since the windings of the two coils may be in the same or in opposite directions. Similarly, for coil 2,

$$\mathcal{E}_2 = \pm M\frac{di_1}{dt} + L_2\frac{di_2}{dt} \tag{32.37b}$$

The total emf is the sum $\mathcal{E}_1 + \mathcal{E}_2$ if the coils are in series; and the currents must be equal, $i_1 = i_2 = i$. Hence

$$\mathcal{E}_1 + \mathcal{E}_2 = (L_1 + L_2 \pm 2M)\frac{di}{dt}$$

and the effective inductance L is

$$L = L_1 + L_2 \pm 2M \tag{32.38}$$

★32.11 The Fundamental Equations of Electromagnetism. We are now in a position to summarize the laws of electricity and magnetism. The study of electromagnetism is the study of the forces between charges, both at rest and in motion. Charge is an indestructible quantity, and conservation of charge is a fundamental principle. The forces between charges are described in terms of two fields, the electric field **E** and the magnetic field **B.** The force due to the electric field on a charge q' is

$$F_1 = q'E \tag{25.1}$$

and the force due to the magnetic field, which depends on the velocity v of the charge, is

$$F_2 = q'vB\sin\phi \tag{29.2}$$

The total force on the charge q' is the vector sum of F_1 and F_2.

The fields **E** and **B** are produced by other charges at rest or in motion. The electric field is determined by Coulomb's law [Eq. (25.3)] or by the equivalent expression of Coulomb's law, Gauss's law,

$$\int_{\text{surface}} E \cos \theta \, dA = 4\pi \, kq \tag{25.9}$$

A similar law must hold for the magnetic field. Both **E** and **B** are forces per unit charge that are inversely proportional to the square of the distance to the electric charge or to the magnetic charge (pole). Since no isolated magnetic poles (charges) exist, the lines of induction **B** must form closed curves (Sec. 30.5). Within any surface the magnetic charge must therefore be zero. Hence we can write

$$\int_{\text{surface}} B \cos \theta \, dA = 0 \tag{32.39}$$

which corresponds directly to Eq. (25.9). Equation (32.39) is not sufficient, however, to determine the values of B, but we must utilize Ampère's law

$$dB = k' \frac{i \, dl \sin \theta}{r^2} \tag{30.1}$$

If the magnetic field changes with time, an electric field must be present also. The magnitude of the electric field has been expressed in terms of the emf $\mathcal{E}$ in a closed circuit by Faraday's law,

$$\mathcal{E} = -\frac{d\Phi}{dt} \tag{32.8}$$

In a completely analogous way an electric field that changes with time produces a magnetic field. Equation (30.1) is correct only for steady currents that do not change with time. If the currents, and hence the electric field, vary with time, an additional term must be added to Eq. (30.1); and we can write it as

$$dB = k' \frac{i \, dl \sin \theta}{r^2} + \text{contributions from } \frac{dE}{dt} \tag{32.40}$$

The existence of a magnetic field produced by a changing electric field was first recognized by J. C. Maxwell (1831–1879), an English physicist.

The electric and magnetic fields, or the *electromagnetic* field, is thus determined by the four relations, Eqs. (25.9), (32.39), (32.8), and (32.40). If the fields are known, the forces on the charge q' can be calculated from Eqs. (25.1) and (29.2). The four equations that determine the fields are known as Maxwell's equations. Maxwell was able to prove from these expressions that electromagnetic energy can exist in free space without the presence of charges or currents and that the fields are oscillating and wavelike in nature (see Sec. 39.2). This electromagnetic radiation is identical with radio waves, light waves, and gamma rays (Chap. 50), the only difference being the frequency of oscillation of the fields.

PROBLEMS

1. A wire 1 meter long is moving in a magnetic field of 1 weber/m², as shown in Fig. 32.1, with a velocity of 1 cm/sec. What is the emf produced?

2. One end of the wire in Prob. 1 is fixed and the other end moves in a circle, in a plane perpendicular to the magnetic field, with a velocity of 1 cm/sec. Calculate the emf.

3. It is desired to make a betatron produce electrons with a maximum energy of 10 Mev and a velocity of c. What magnetic field and what radius of path could be used?

4. Apply Lenz's law to find which terminal is positive in the disk dynamo of Fig. 32.6.

5. A two-pole generator has a field of 8,000 gausses and runs at 3,600 rpm. How many turns must be on the armature of area 20 cm² to produce a maximum emf of 100 volts?

6. A rectangular coil is rotating in a magnetic field about one side as an axis. Find the emf induced in the coil from Eq. (32.3), and show that this agrees with Eq. (32.13).

★7. A search coil of 2 cm² area and 100 turns is removed from a field of 1,000 gausses. How much charge flows in a circuit of 10 ohms resistance connected to the coil?

8. What is the relation between 1 henry and 1 emu of inductance?

9. What is the self-inductance of a solenoid 1 m long, wound with 1,000 turns, each 2 cm in diameter?

★10. Suppose that the solenoid of Prob. 9 is regarded as two coils that are connected in series, find the inductance by means of Eq. (32.38).

★11. A large electromagnet has a time constant of 5 sec and resistance of 10 ohms; what is the inductance?

★12. An RL circuit has a time constant of 0.01 sec. How long does it take the current to attain 90 per cent of its final value; 99.9 per cent?

★13. An inductance of 20 millihenrys carries a current of 1 amp. How much energy is stored in the inductance?

14. A copper rod is placed within a coil. Explain why the inductance of the coil is reduced.

★15. After a steady current i_0 has been established in the circuit of Fig. 32.12, the emf of the cell vanishes so that the circuit becomes simply an inductance and a resistance in series. Show that the current decays according to $i = i_0 e^{-(R/L)t}$.

CHAPTER 33

THE MAGNETIC PROPERTIES OF MATTER

***33.1. The Magnetic Vectors.** We have already introduced three vector quantities to describe a magnetic field—the magnetic induction **B**, the intensity of magnetization **I**, and the magnetic intensity **H**. The magnetic state of matter is described by the vector **I**, which represents the magnetic moment per unit volume of the material. It will be recalled that **I** can also be expressed in terms of magnetic poles or in terms of a surface current density (Sec. 31.3). In free space, **I** is zero, and **B** and **H** are related by the simple equation

$$\mathbf{B} = \frac{k'}{k''}\mathbf{H} \tag{33.1}$$

where $k' = \mu_0/4\pi$ and $k'' = 1/4\pi$ in practical units; in the electromagnetic system, $k' = k'' = 1$. Within matter, Eq. (33.1) is no longer true; the surface currents flowing on the matter (or we can equally well say the magnetic poles) also contribute to the field **B**, and their contribution must be taken into account. For this purpose let us consider the simple case of a toroidal coil that is wound upon a ring of the material under discussion. The case is a simple one because the material has no ends where free poles can exist. If an additional small coil is wound over a portion of the toroid and connected to a galvanometer, then the galvanometer deflections are proportional to the changes in **B** within the coil and **B** can be measured. Such an arrangement is known as a Rowland ring, after the American physicist A. H. Rowland (1848–1901), who used it extensively.

Within the ring the field **B** is everywhere uniform and consists of one part arising from the current in the coil [Eq. (30.28)] and another part arising from the surface currents in the material of the core. Hence the total field is

$$B = k' \times 4\pi j + k' \times 4\pi j_s \tag{33.2}$$

We have seen in Sec. 31.6 that **H** in free space was the force on a unit pole. In a magnetic medium **H** is defined as the field that arises from actual currents and poles only, with exclusion of the surface currents. The field **B** includes the surface currents.

The magnetic intensity **H** within the ring arises only from the current in the coil since there are no poles present to contribute. The quantity **H** has the same value in the coil as if the core were not present, and its value is simply obtained from Eqs. (30.28) and (33.1) as

$$H = k'' \times 4\pi j \tag{33.3}$$

Therefore the field **B** can be written for this particular case as

$$\mathbf{B} = \frac{k'}{k''}\mathbf{H} + k'' \times 4\pi\mathbf{I} \tag{33.4}$$

where Eq. (31.8) has been employed. Although we have considered only a toroidal coil, *Eq.* (33.4) *represents a general relation and is always true.*

In the electromagnetic system of units, **B** is given by

$$\boxed{\mathbf{B} = \mathbf{H} + 4\pi\mathbf{I} \qquad \text{emu}} \tag{33.5}$$

and in the practical system of units

$$\boxed{\mathbf{B} = \mu_0\mathbf{H} + \mathbf{I} \qquad \text{mks}} \tag{33.6}$$

***33.2. Susceptibility and Permeability.** It is customary to think of the vector **H** as the "cause" of the intensity of magnetization **I** of a substance, although all that can be observed experimentally are the relations between **I** and **H** and between **I** and **B**. For many substances, **I** is indeed proportional to **H** within the accuracy of the observations. The constant of proportionality is called the *magnetic susceptibility* χ of the material, and therefore

$$\boxed{\chi = I/H} \tag{33.7}$$

In terms of χ, Eq. (33.4) can be written as

$$\mathbf{B} = \left(\frac{k'}{k''} + k'' \times 4\pi\chi\right) H$$

and the combination $[(k'/k'') + k'' \times 4\pi\chi]$ is called the *permeability* μ of the medium; hence

$$\boxed{B = \mu H} \tag{33.8}$$

and

$$\boxed{\mu = \left(\frac{k'}{k''} + k'' \times 4\pi\chi\right)} \tag{33.9}$$

In practical units, $k' = \mu_0/4\pi$, $k'' = 1/4\pi$, and Eq. (33.9) becomes

$$\mu = \mu_0 + \chi$$

The susceptibility χ and the permeability μ are measured in the same units as μ_0, that is, in henrys per meter. Since $\chi = 0$ for free space, the quantity μ_0 is sometimes called the "permeability of free space." A third quantity is also used, *i.e.*, the ratio of μ to μ_0. This is called the *relative permeability* μ_r, and

$$\mu_r = \frac{\mu}{\mu_0} \tag{33.10}$$

The magnetic properties of a material are therefore specified by any one of the three quantities χ, μ, or μ_r. These quantities are related by the equation

$$\mu = \mu_0 - \chi = \mu_r \mu_0 \tag{33.11}$$

In the electromagnetic system of units, $k' = k'' = 1$, and Eq. (33.9) becomes

$$\mu = 1 + 4\pi\chi$$

Since μ is dimensionless and since the permeability of free space is unity, the permeability is equal to μ_r,

$$\mu_r = 1 + 4\pi\chi = \frac{B}{H} \qquad \text{emu} \tag{33.12}$$

Since the value of the relative permeability μ_r is independent of the system of units, we can use it to find the relation between the susceptibilities in the mks and emu systems; thus

$$\mu_r = 1 + 4\pi\chi\ (\text{emu}) = 1 + \frac{\chi\ (\text{mks})}{\mu_0}$$

or

$$\chi\ (\text{mks}) = 4\pi\mu_0\chi\ (\text{emu}) \tag{33.13}$$

These relationships are very similar to those involving the electric susceptibility and the permittivity of a dielectric medium (cf. Sec. 26.2).

***33.3. Diamagnetism.** The magnetic properties of all substances are the result of electronic motions within the atoms of the material. A magnetic moment may arise from a variety of phenomena, depending on the material. One phenomenon that is universally present arises from the induced emfs in the atomic circuits. Suppose an electron is revolving in a circular orbit as indicated in Fig. 33.1. If the electron is traveling clockwise as shown, the conventional direction of the current in the elementary circuit is counterclockwise. Now suppose an external field **B**, extending into the paper, is then applied perpendicular to the orbit. The field **B** indicated does not include the field, outward from the paper, produced by the motion of the electron. During the growth of the field **B** from zero to its final value an emf is induced in the elementary circuit. By Lenz's law this emf is in the counterclockwise direction, and the electron is speeded up in its motion. The larger speed of rotation results in a larger field produced by the electronic motion, and thus the total field is reduced. If the electron traverses its orbit in the counterclockwise direction, then the application of the field reduces the electronic speed and again a change of field is produced to oppose **B**. If the substance contains electron orbits oriented at random, the total magnetic moment vanishes in the absence of an external field. The induced

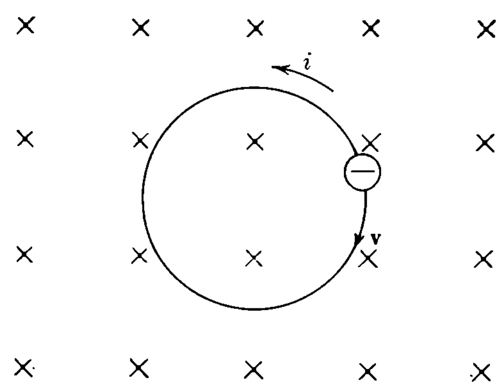


FIG. 33.1. An electron in a field **B**.

emfs in the orbits all tend, however, to produce a change in the moments of the orbits such as to *oppose* the applied field. The net magnetic moment produced is *negative*, that is, in the opposite direction to B.

Since a negative magnetic moment is produced by induction, **I** is directed oppositely to **B**, and hence also to **H**, and the susceptibility χ is negative. When no other mechanisms are operative, the substance is called *diamagnetic*. If χ is negative, then μ_r is less than 1 and μ is less than μ_0. Diamagnetic effects are usually quite small. Sample values are shown in Table 33.1. The magnitude of χ depends primarily on the areas of the electron orbits. If the area is large, a large emf is induced and the susceptibility is large. Since many of the orbits are deep within the atom, and therefore unaffected by heat motions of the molecules, there is no large change of diamagnetic susceptibility with temperature.

When a bar of diamagnetic material is placed in a magnetic field, a torque is exerted that tends to set the bar at right angles to the field since, it can be shown, in this position the energy is a minimum. From this phenomenon the name is derived, the prefix *dia* meaning "across."

***33.4. Paramagnetism.** In some materials the diamagnetic effects are exceeded by others, which produce a positive magnetic moment, thus making the resultant susceptibility positive. Such behavior is called *paramagnetic*. Each of the electron orbits in a substance has a magnetic moment and consequently experiences a torque when a field is applied just like the coil of a galvanometer. If the orbits are free to turn, then there is a tendency for the moments to align themselves with the field. This alignment is retarded by the thermal agitation of the atoms, and except at very low temperatures the net magnetic moment produced is very small. A *paramagnetic* rod tends to set itself *parallel* to the lines of magnetic field. Some values of paramagnetic susceptibilities are given in Table 33.1.

Table 33.1. *Values of the Magnetic Susceptibility*

Substance	Temp	χ (emu)	χ (mks) henry/m	$\mu_r - 1$
Bismuth	18	-13×10^{-6}	-21×10^{-11}	-17×10^{-5}
Carbon (diamond)	20	$-$ 1.8	$-$ 2.8	$-$ 2.1
Copper	18	$-$ 0.8	$-$ 1.2	$-$ 1.0
Silver	18	$-$ 2.1	$-$ 3.4	$-$ 2.6
Aluminum	18	1.6	2.7	2.2
Iron ammonium alum	-258	1,030	1,620	1,290
Iron ammonium alum	17	53	84	67
Cupric chloride	18	29	46	37
Ferric chloride	20	240	380	300
Oxygen, liquid	-219	349	550	438

The magnitude of the susceptibility depends principally on the freedom of the orbits to turn. Within a single atom the electron orbits are usually arranged in oppositely revolving pairs, and hence the magnetic moments cancel. The resultant magnetic moment of an atom is produced by only one or two outer electrons not revolving oppositely. The net magnetic moment of the atom then changes its direction as the atom as a whole rotates with energy of thermal agitation. The susceptibility increases with decreasing temperature since the energy of agitation is decreased. Most substances show a variation of the form

$$\chi = \frac{A}{T + T_0} \tag{33.14}$$

where A and T_0 are empirical constants and T is the absolute temperature. For most substances, χ is independent of the field H. There are a few materials in which, at low temperatures and high fields, χ decreases as H increases.

In an atom the electrons not only move in orbits about the nucleus but also spin like small tops. Each electron thus has a magnetic moment. Just as the electron orbits are arranged in pairs with no net magnetic moment, so the axes of electron spin are parallel and in opposite directions by pairs. Unbalanced spin magnetic moments do contribute, however, to the paramagnetic susceptibility in the same way as do the orbital moments. Atomic nuclei also possess magnetic moments, but these are too small to have an appreciable effect on the susceptibility.

***33.5. Ferromagnetism.** Iron, nickel, cobalt, and a few alloys whose constituents are only weakly magnetic show values of the relative permeability μ_r that are very much larger than unity. These substances are called *ferromagnetic*. In addition the susceptibility and permeability are not constant but vary over large ranges as the field H is varied. If a sample of soft iron, for example, is initially unmagnetized and then subjected to a steadily increasing field intensity H, the values of B and I vary as shown in Fig. 33.2. The curve of B vs. H is called the *magnetization curve* of the substance. The permeability, the ratio of B to H, is the slope of the line drawn from the origin to a point on the magnetization curve. As H increases, the permeability thus increases to a maximum value at the field corresponding to the point where the line to the origin is tangent to the magnetization curve (line A in the figure) and then decreases again. From the figure, the maximum permeability occurs for H = 110 amp/m and B = 0.6 weber/m². The maximum value is therefore μ_{max} = 0.6/110 = 0.0054 henry/m. The maximum relative permeability is the ratio of μ_{max} to μ_0, or

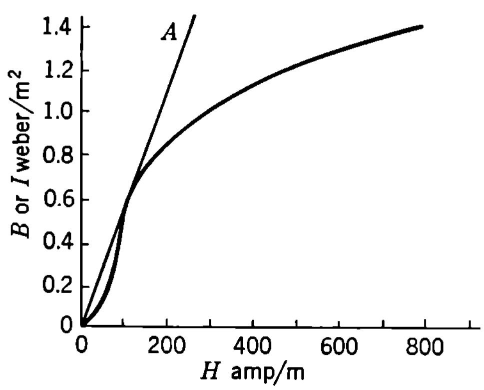


FIG. 33.2. Magnetization curve of soft iron (mks units).

$$\mu_{r\max} = \frac{\mu_{max}}{\mu_0} = \frac{5.4 \times 10^{-3}}{1.26 \times 10^{-6}} = 4{,}280$$

Over the range of values of H shown in Fig. 33.2 the value of I is not appreciably different from that of B. The susceptibility χ (mks) is therefore equal to the permeability μ. For larger magnetic intensities, I approaches a constant value: the flux density B continues to increase slowly, however, as H increases. In this range of H the iron is said to be *saturated.*

In ferromagnetic materials the elementary circuits responsible for magnetization are the spinning electrons in the material. When H is sufficiently large that all the axes of spin of the electrons are pointed in the same direction, saturation occurs. Ferromagnetic materials differ from paramagnetic ones chiefly in the ease with which the alignment of electrons occurs. In ferromagnetics the heat motions in the substance also prevent perfect alignment of the electron spins, and a higher value of I at saturation is obtained if the material is cooled to a low temperature. At temperatures above room temperature, smaller and smaller saturation values are obtained until, at a rather sharply defined temperature, the substance ceases to be ferromagnetic and becomes paramagnetic. This transition temperature is called the *Curie temperature*. Figure 33.3 shows how the saturation value of I changes with T in iron. The Curie temperature lies at about 760°C. The transition at the Curie temperature is similar to phase changes (changes in crystal structure, for example) that take place in other materials. A latent heat exists, and the specific heat shows anomalous variations (see Sec. 18.2). These phenomena have been called collectively *recalescence* (Latin *calere*, to be hot). The presence of these anomalies is easily made evident by observation of an iron wire heated through the Curie temperature.

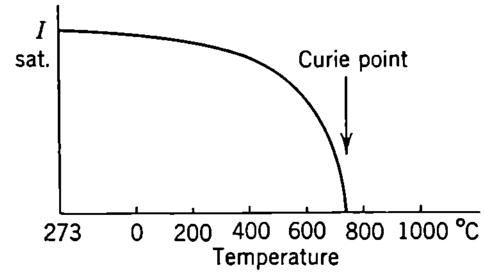


FIG. 33.3. The saturation magnetization in iron vs. temperature.

★33.6. Hysteresis. If an unmagnetized sample of iron is slightly magnetized by currents flowing near it, the induction B is given as a function of H by the magnetization curve of Fig. 33.2. If the magnetic intensity is now decreased, the induction does not decrease along the magnetization curve but follows a new curve, as indicated in Fig. 33.4. The induction depends, therefore, not only upon the value of H but also upon the previous state of magnetization. This phenomenon is known as *hysteresis* (see Sec. 4.6). Even when the field H is reduced to zero, the induction has a large value indicated in Fig. 33.4 as B_r. This value of the induction is called the *remanence*, or *retentivity*, of the material. The material is now a permanent magnet. Such a magnet is not really permanent, since the application of a field H in the *reverse* direction, shown as a negative value of H in the figure, causes the induction to decrease and come to a zero value when $H = -H_c$. This value of the field H is called the *coercive force*. A good material

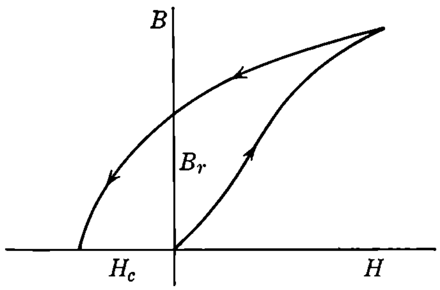


FIG. 33.4. Remanence B_r and coercive force H_c.

for a permanent magnet is one in which B_r and H_c have large values. In mild steel, for example, $B_r = 0.95$ weber/m^2 and $H_c = 3{,}600$ amp/m; in a special alloy of iron, nickel, cobalt, and aluminum, Alnico 5, designed to be used for permanent magnets, $B_r = 1.25$ weber/m^2 and $H_c = 44{,}000$ amp/m.

If the value of H oscillates between two extreme values, as when an alternating current is applied to a magnetizing coil, B traces out a *hysteresis loop* such as that shown in Fig. 33.5. It is frequently difficult to obtain a ferromagnetic material in which B is truly zero. Two methods are possible. A series of hysteresis loops can be traced out with successively smaller extreme values of H. The successive loops wind closer and closer about the origin, and B can be made quite small. The second method is made evident in Fig. 33.5. If H is brought to zero when the state of the magnetic material corresponds to a certain point A, on the reverse side of the loop, then B travels along the dashed curve to the origin.

It is apparent that permeability and susceptibility lose much of their significance when hysteresis is encountered. Thus, at the point a in Fig. 33.5, B/H has an infinitely large value; and over the segments of the hysteresis loop from a to b and c to d, B/H is negative. Hysteresis is not the only phenomenon that limits the usefulness of the concept of permeability. In large crystals of iron, or in sheets of iron alloys that have been rolled, it is possible that B and H do not even have the same direction. Even in these complicated cases, however, Eq. (33.4) remains true, and the vector **B** is obtained by taking the vector sum of (k'/k'') **H** and $(k'' \times 4\pi)$ **I**.

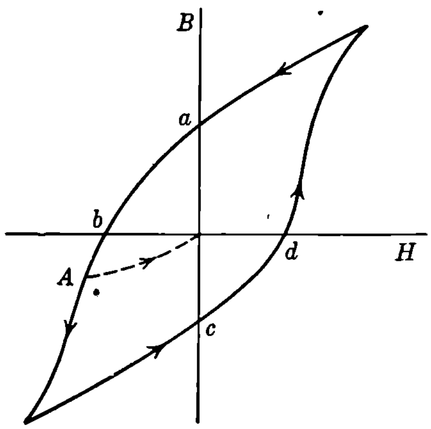


FIG. 33.5. A hysteresis loop.

An important consequence of hysteresis effects is that heat is produced in a ferromagnetic material as the result of a sort of friction in the magnetization process. This heat energy is produced at the expense of the electric energy used to magnetize the specimen. The amount of heat developed can be calculated. Suppose we have a toroid of mean circumference l and cross-sectional area A, wound with N turns of wire. If the current in the winding is increased from i to $i + di$ in the time dt, there is a change in flux $d\Phi = A\,dB$ through the winding and an induced emf arises equal to

$$\mathcal{E} = N\frac{d\Phi}{dt}$$

The power P necessary to magnetize the specimen is $P = \mathcal{E}i$, and the energy dW supplied is $P\,dt$. Combining these relations, we have

$$dW = \mathcal{E}i\,dt = Ni\,d\Phi = NiA\,dB$$

The magnetic intensity H is given by

$$H = k''\frac{Ni}{l}$$

and hence dW can be written in terms of H as

$$dW = AlH\ dB/k'' = VH\ dB/k''$$

where V is the volume of the toroid. If a hysteresis loop is traced out, $\int H\ dB$ taken from the maximum to the minimum value of B and back again does not equal zero but is equal to the area of the hysteresis loop. The total energy W that must be supplied is therefore

$$W = (V/k'') \int_{\text{loop}} H\ dB \tag{33.15}$$

Steinmetz has given a useful empirical relation from which the energy loss can be found. His equation is

$$W = RVB_{\text{max}}^{1.6} \tag{33.16}$$

where R is an empirical constant.

***33.7. Ferromagnetic Domains.** As already stated, the magnetic properties of ferromagnetic materials are the result of the magnetic moments of the spinning electrons within the material. Not only are these electrons acted upon by applied external fields, but also each electron is in the magnetic field of all the other electrons in the ferromagnetic body. The internal field at any one electron is usually very much larger than the applied external field. Moreover the direction of the internal field is always such as to make the moments of the electrons align with one another, and we should expect that the ferromagnetic material would always be magnetized to saturation. This is actually the case! A piece of iron when unmagnetized as a whole is made up of a collection of *domains*, as they are called, each domain consisting of 10^{12} to 10^{15} atoms. Each domain is magnetized to saturation; but the directions of magnetization of the separate domains are all different, and the resultant magnetization of the iron is zero.

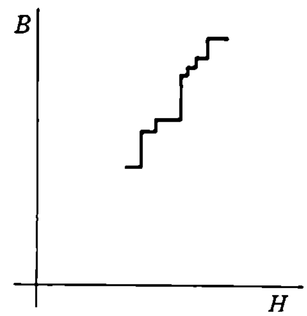


FIG. 33.6. Portion of magnetization curve showing the Barkhausen effect.

When an external field is applied, several processes take place. At low fields the directions of magnetization of those domains that are almost lined up with the field are slightly changed to more nearly exact alignment. In addition those domains that are favorably oriented grow in size at the expense of unfavorably oriented ones. The region of the magnetization curve where these effects predominate is the portion that is concave upward at the foot of the curve (below $H = 80$ amp/m in Fig. 33.2). The sharply rising portion of the magnetization curve where the permeability is high (between $H = 80$ amp/m and 200 amp/m) is characterized by sudden complete reversals of the direction of magnetization of the domains. If a small portion of the magnetization curve in this region is examined with care, it is found to consist of a series of discontinuous steps, as shown in Fig. 33.6. The steplike increases in B represent the reversal of single domains. This effect is known after its discoverer as the *Barkhausen effect*. It can be demonstrated easily by connecting headphones to the output terminals of an amplifier whose input terminals are connected to a search coil surrounding the magnetic specimen. The sudden changes in B produce clicks in the headphones.

In the high-field region (above $H = 200$ amp/m in Fig. 33.2) the increase in magnetization is accomplished by the rotation of the direction of magnetization of each domain as a whole until the domains are all aligned when saturation is reached.

The magnetization of a ferromagnetic material is accompanied by changes in the size of the specimen. This phenomenon is known as *magnetostriction*. Although the phenomenon is not well understood, it appears to be intimately connected with all the ferromagnetic phenomena. Magnetostriction is utilized in oscillators which generate ultrasonic waves.

***33.8. Iron-cored Inductances.** In Sec. 32.8 the inductance of a long solenoid with no core was calculated. We wish now to find the inductance of a solenoid wound on a magnetic core. By Eq. (32.28) the inductance can be found from the expression

$$L = \frac{N\Phi}{i} \tag{32.28}$$

The flux Φ in a core of permeability μ is

$$\Phi = BA = \mu HA$$

and since

$$H = 4\pi k'' \frac{iN}{l}$$

$$L = 4\pi k''\mu \frac{N^2 A}{l} \tag{33.17}$$

This result is the same as that for an air-core inductance with k' replaced by $k''\mu$. If we write

$$L_0 = k' \frac{4\pi N^2 A}{l}$$

then

$$L = \frac{k''\mu}{k'} L_0 = \mu_r L_0 \tag{33.18}$$

The relative permeability μ_r of a medium is thus equal to the ratio of the inductance of a solenoid wound on a core of the medium to the inductance of the solenoid with no core. It will be recalled that in a similar fashion the specific inductive capacity is equal to the ratio of capacitance of a capacitor containing a dielectric to that of the capacitor without the dielectric.

We can now derive an expression for the energy density within a magnetized material. The stored energy in an inductance, derived in Sec. 32.9, is

$$W = \tfrac{1}{2}Li^2 \tag{32.36}$$

By substitution for L from Eq. (33.17) we have

$$W = \frac{1}{2} k''\mu \times 4\pi \frac{N^2 A}{l} i^2 = \frac{1}{2} \mu \frac{1}{4\pi k''} \left(\frac{4\pi k'' Ni}{l}\right)^2 Al$$

The quantity Al is the volume of the magnetic material, and

$$H = 4\pi k'' \frac{Ni}{l}$$

Hence, since $B = \mu H$, the energy density is

$$\frac{W}{Al} = \frac{1}{2}\frac{\mu H^2}{4\pi k''} = \frac{1}{2}\frac{BH}{4\pi k''} = \frac{1}{2}\frac{B^2}{\mu \times 4\pi k''} \tag{33.19}$$

If a ferromagnetic material is used as the core of an inductance, the value of μ to be used in Eq. (33.17) is uncertain. In fact the inductance of a solenoid with an iron core is not a constant but depends on the current i through the solenoid. Nevertheless inductors with iron cores are used in many applications. For example, to control an alternating current through a resistive load, such as a bank of lamps, a variable inductor is used in series with the lamps. The counter emf induced in the inductor results in a reduction of current. A second application of iron-cored inductors is their use as elements in circuits to reduce a small alternating current superimposed on a larger direct current. Inductors for such purposes must therefore be used with the proper direct current to obtain the specified inductance.

***33.9. The Magnetization of Short Specimens.** In most of the preceding sections we have considered only cases where the magnetic material is in the form

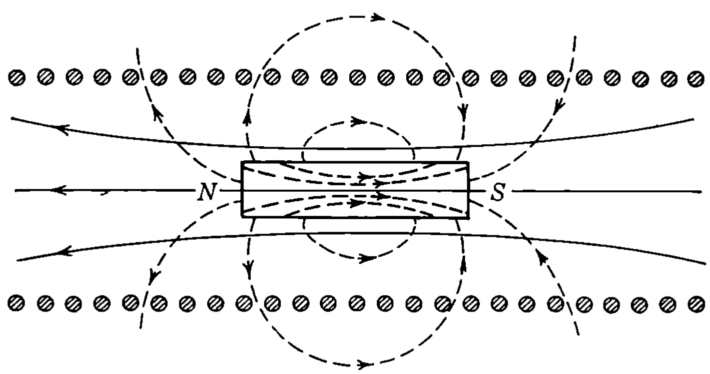


FIG. 33.7. The components of H due to the solenoid (solid lines) and due to the poles (dotted lines).

of a long rod surrounded by a uniform magnetizing winding or in the form of a toroid. The contribution of the poles at the ends of the specimen to the value of **H** is zero in a toroid, since there are no free poles, and in a long rod the ends are so far away that the effect they cause is negligible. If a short bar of a ferromagnetic material is placed in a uniform field, for example, within a long solenoid, the situation is more complicated. The vector **H** is made up of one portion arising from the current flowing in the solenoid and another portion arising from the free poles at the ends of the specimen. These two portions of **H** are indicated in Fig. 33.7. The solid lines represent the uniform field arising from the current in the solenoid. The dotted lines represent the field from the poles at the ends of the magnetized specimen. The field of the poles is not at all uniform but varies considerably over the length of the bar. If the pole strength is appreciable, the resultant of these two portions is not uniform over the specimen. Consequently the specimen is not uniformly magnetized. It should be noted that within the specimen the field of the poles is opposite to the field of the solenoid. The poles thus produce a demagnetizing effect.

In one case the demagnetizing effect of the poles can be accurately calculated. If the specimen is an ellipsoid, or a sphere in a special case, the magnetic charges distribute themselves over the surface in such a way that the field **H** within the specimen is uniform, although less than the field of the solenoid alone.

PROBLEMS

1. What are the values of the constants A and T_0 of Eq. (33.14) for the iron ammonium alum listed in Table 33.1?

2. Using the data of Fig. 33.2, plot the variation of μ with H for soft iron.

3. What are the values of the remanence and coercive force of Alnico 5, in emu?

4. Calculate the total stored magnetic energy and the energy density in an iron-cored coil of length 0.4 m, cross-sectional area 4 cm², and wound with 100 turns, the permeability of the iron being 1,000.

5. What is the inductance of a toroidal coil of 500 turns wound on a core of cross-sectional area 10 cm² and mean circumference 50 cm, if the effective relative permeability is 200?

6. An iron toroid 15 cm in diameter and 5 cm² in cross-sectional area is wound with 1,000 turns. A current of 0.7 amp produces an induction of 1 weber/m². What is the permeability of the iron?

7. Plot the magnetization curve and a curve showing the variation of μ with H from the following data for soft iron:

B, gausses	H, oersteds
10,600	5
12,800	10
14,000	20
15,500	40
16,700	70
17,000	100

8. A domain in iron contains 10^{14} atoms. What is its volume? If the domain has a cubical shape, what is the length of a side?

9. A steel permanent magnet 1 cm² in cross section and 5 cm long is uniformly magnetized to a value of B of 10,000 gausses. What are the values of I, μ, j_s, and the total magnetic moment of the magnet?

10. What is the stored energy in a magnet of volume 10 cm³ in a field of 100 amp/m? Use the magnetization curve in Fig. 33.2 and the results of Sec. 33.6.

11. A solenoid 1 m long is wound with 1,000 turns and carries a current of 1 amp. What is the value of H at the center? A thin steel rod of cross-sectional area 0.2 cm² and 20 cm long is placed at the center of the solenoid and becomes uniformly magnetized to a value of $B = 10^4$ gausses. What is now the value of H at the center of the steel rod?

CHAPTER 34
ALTERNATING CURRENTS

34.1. The Effective Value of an Alternating Current. The emf produced in an a-c generator as described in Sec. 33.5 causes a current i to flow in a resistor of resistance R. This current is also alternating and can be represented by the equation

$$i(t) = i_{\max} \cos (2\pi ft - \phi) \tag{34.1}$$

where $i_{\max}$ is the maximum value of the current and $-\phi$ is the phase of the current when $t = 0$. The power P dissipated in the resistor is

$$P = Ri^2 = i_{\max}{}^2 R \cos^2 (2\pi ft - \phi)$$

The power thus varies with time from $i_{\max}{}^2 R$ to 0. The steady current I that produces the same *average* power loss in R is called the *effective current.* Its value is given by

$$\bar{P} = RI^2 = Ri_{\max}{}^2 \overline{\cos^2(2\pi ft - \phi)} \tag{34.2}$$

where the bar denotes the average value. It will be recalled that the average value of any function $f(x)$ over the interval between x_1 and x_2 is defined by

$$\bar{f} = \frac{\int_{x_1}^{x_2} f(x)dx}{x_2 - x_1}$$

This formula can be applied to find the average value of $\cos^2 (2\pi ft - \phi)$. By means of the theorem from trigonometry that

$$\cos^2 \alpha = \tfrac{1}{2} - \tfrac{1}{2} \cos 2\alpha$$

we can write

$$\cos^2 (2\pi ft - \phi) = \tfrac{1}{2} - \tfrac{1}{2} \cos (4\pi ft - 2\phi)$$

The average value of the power can be calculated, taking the time interval to be 1 cycle or from $t = 0$ to $t = 1/f$, as

$$\bar{P} = Ri_{\max}{}^2 \left[\frac{\int^{1/f} \tfrac{1}{2}\, dt}{1/f} - \frac{\int_0^{1/f} \tfrac{1}{2} \cos (4\pi ft - 2\phi)}{1/f}\, dt \right]$$

The first integral has the value $\frac{1}{2}$ and the second is $\frac{1}{2}$ times the mean value of a cosine function over 2 cycles, or zero. Hence

$$\bar{P} = \tfrac{1}{2} Ri_{\max}{}^2$$

The effective current is thus equal to the rms value of the current, and

$$I = i_{\max} \sqrt{\tfrac{1}{2}} = 0.707 i_{\max} \tag{34.3}$$

The variation of the current and power with time is shown in Fig. 34.1. It is customary in dealing with alternating currents to specify the effective or rms value of the current rather than the maximum value. The same convention is adopted for an alternating emf or potential difference. Thus a lighting circuit in a house is designated as a 110-volt circuit.

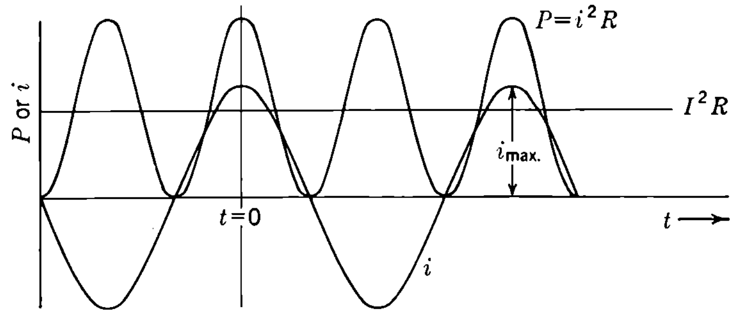


FIG. 34.1. The effective value of an alternating current.

The potential difference between the terminals of a lamp socket varies between $\sqrt{2}\ 110 = 155$ volts in one direction to 155 volts in the opposite direction.

34.2. The Circuit with Resistance, Inductance, and Capacitance. If an alternating emf produces a potential difference V, where

$$V = V_m \cos 2\pi ft \tag{34.4}$$

across a resistor of resistance R, an inductance L, and a capacitance C connected in series, as shown in Fig. 34.2, the potential difference of the source must equal the sum of the potential differences across the components of the circuit. Hence we have

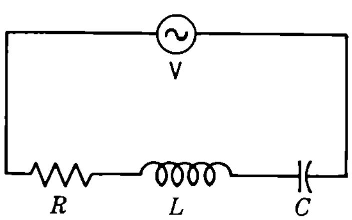


FIG. 34.2. A series *RLC* circuit.

$$V_m \cos 2\pi ft = Ri + L\frac{di}{dt} + \frac{q}{C} \tag{34.5}$$

where i is the current in the circuit and q is the charge on the capacitor. Since the rate of change of charge is equal to the current, $dq/dt = i$, and Eq. (34.5) can be written as

$$L\frac{d^2q}{dt^2} + R\frac{dq}{dt} + \frac{q}{C} = V_m \cos 2\pi ft \tag{34.6}$$

This differential equation must be solved to find the charge q. We have already met this equation in another guise in Sec 12.12, where the equation was applied to forced harmonic oscillations of a mechanical system. It will be recalled that a complete solution consists of two parts, the steady-state solution and the transient that dies out rapidly. As before, we shall concern ourselves with the steady-state solution only.

By comparison with Eqs. (12.24) to (12.26), using Table 34.1, we find

$$q = \frac{V_m}{2\pi f Z} \sin (2\pi f t - \phi)$$

$$i = \frac{V_m}{Z} \cos (2\pi f t - \phi) \tag{34.7}$$

where the constants Z and ϕ are given by

$$Z = \sqrt{R^2 + \left(2\pi f L - \frac{1}{2\pi f C}\right)^2} \tag{34.8}$$

$$\tan \phi = \frac{2\pi f L - 1/2\pi f C}{R} \tag{34.9}$$

The quantity Z is called the *impedance* of the circuit, just as in the mechanical case. In fact the complete correspondence between the electrical and mechanical cases is shown in Table 34.1.

Table 34.1. *Corresponding Quantities in Forced Electrical and Mechanical Oscillations*

Electrical System	*Mechanical System*
Charge q	Displacement x
Current i	Velocity v
Inductance L	Mass m
Resistance R	Damping factor R
$1/C$	Force constant k
Potential amplitude V_m	Force amplitude F_0
Radian frequency $2\pi f$	Radian frequency ω'
Phase constant ϕ	Phase constant α

As in the mechanical case, resonance occurs when Z is a minimum with respect to f, that is when $2\pi f L = 1/2\pi f C$, and the frequency of resonance f_0 is given by

$$f_0 = \frac{1}{2\pi \sqrt{LC}} \tag{34.10}$$

At resonance, $Z = R$, and $\phi = 0$.

34.3. Electric Impedance and Reactance. The separate terms that make up the electric impedance are given special names and symbols.

The quantity

$$X_L = 2\pi fL \tag{34.11}$$

is called the *inductive reactance;*

$$X_C = -\frac{1}{2\pi fC} \tag{34.12}$$

is called the *capacitive reactance.* The total reactance X of the circuit is the sum of the inductive and capacitive parts,

$$X = X_L + X_C$$

The reactance of a circuit can thus be either positive or negative. The impedance of the circuit can be written

$$Z = \sqrt{R^2 + X^2} = \sqrt{R^2 + (X_L + X_C)^2} \tag{34.13}$$

A circuit for which X is not zero is called a reactive circuit. The reactances, like the resistance, are measured in ohms if L and C are measured

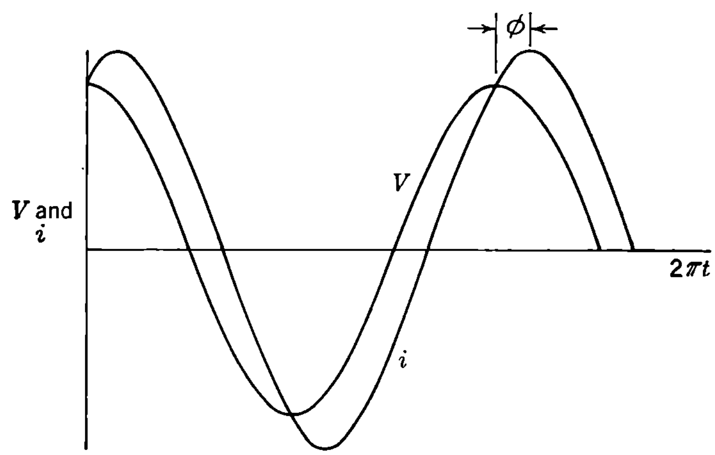


Fig. 34.3. Current and potential in an inductive circuit.

in henrys and farads, respectively. At the common power-line frequency of 60 cycles per second (cps), $2\pi f = 377 \text{ sec}^{-1}$, and 1 henry has the reactance X_L of 377 ohms. A capacitance of 1 μf has a reactance X_C that is negative and equal to $-10^6/377 = -2{,}650$ ohms.

Although the value of the impedance does not depend upon the sign of the reactance, since the reactance is squared, the phase difference between the current and the applied potential does depend upon the sign. From Eq. (34.9),

$$\tan \phi = \frac{X_L + X_C}{R} = \frac{X}{R} \tag{34.14}$$

In an inductive circuit where X is positive, $\tan \phi$ is positive, and ϕ lies between 0 and 90°. The maximum value of the current occurs at a *later* time than does the maximum value of the potential. The current, therefore, *lags* behind the applied potential. The phase relations are shown in Fig. 34.3. In a capacitive circuit, on the other hand, X is negative, and the current *leads* the voltage as in Fig. 34.4.

The relations of Eqs. (34.13) and (34.14) are easily summarized in the vector impedance diagram shown in Fig. 34.5. The diagram is self-

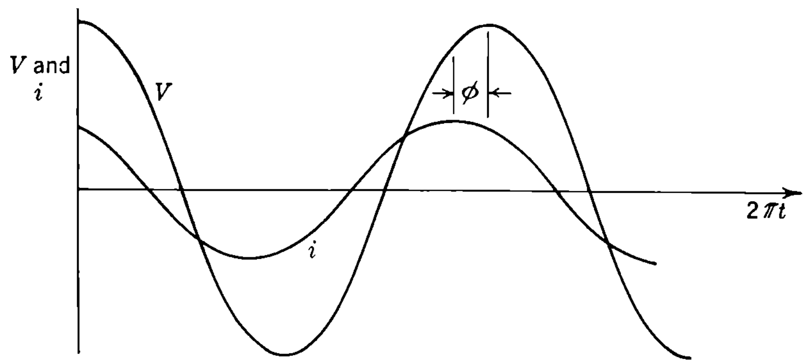


FIG. 34.4. Current and potential in a capacitive circuit.

explanatory. The angle ϕ is sometimes called the *phase of the impedance.* The phase of a resistive impedance is zero, and the phase of a purely reactive impedance is $\pm 90°$.

34.4. Potential Differences across Parts of an A-c Circuit. The results of Sec. 34.2 are immediately applicable to part of an a-c circuit as well as to the circuit as a whole. Equation (34.5) was obtained by equating the sum of the potential differences across each component to the potential difference of the generator. For the circuit as a whole, if only maximum values, not instantaneous ones, are considered we have

$$V_m = i_m Z \qquad (34.15)$$

This relation is valid also, then, for a part of the circuit. Across the resistor we have

$$V_R = i_m R$$

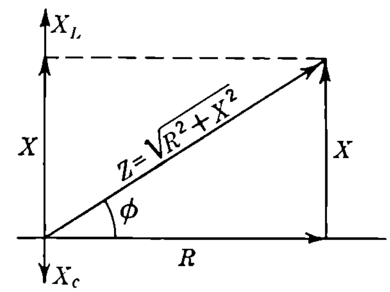


FIG. 34.5. Vector impedance diagram.

and across the inductance and capacitance we have

$$V_L = i_m X_L = i_m 2\pi f L \qquad (34.16)$$

$$V_C = i_m |X_C| = i_m \frac{1}{2\pi f C}$$

respectively. We must use the absolute value of X_C in these equations, since the impedance is the square root of $X_C{}^2$ and is always positive.

It must be emphasized that the potential difference across these elements in series is not the sum of V_R, V_L, and V_C. The total potential difference is the sum of the instantaneous values and not the sum of the maximum values, since phase differences exist. The phase of the current is, of course, the same throughout a series circuit, but the phases of the potential differences across the components vary from one to another. In general, then, Eq. (34.15) is valid for part of the circuit as well as for the whole. The equation can also be written in terms of the rms values. If these are denoted by V and i without subscripts, then

$$V = iZ$$

Worked Example. Let us consider an inductance of 1 henry in series with 1 μf of capacitance and a resistance of 1,000 ohms. At a frequency of 60 cps the total reactance is 377 − 2,650 = −2,270 ohms. The impedance is

$$Z = \sqrt{(1{,}000)^2 + (2{,}270)^2} = 2{,}480 \text{ ohms}$$

If the generator produces a potential difference of 110 volts rms, the current i is

$$i = \frac{110}{2{,}480} = 0.0443 \text{ amp}$$

The potential differences across R, L, and C are

$$V_R = 0.0443 \times 1{,}000 = 44.3 \text{ volts}$$
$$V_L = 0.0443 \times 377 = 16.7 \text{ volts}$$
$$V_C = 0.0443 \times 2{,}650 = 117.6 \text{ volts}$$

The potential difference V_C across the capacitor thus exceeds the line voltage.

Suppose now the frequency of the generator is varied until resonance takes place. By Eq. (34.10), this occurs for

$$f_0 = 159 \text{ cps}$$

The impedance of the whole circuit is

$$Z = R = 1{,}000 \text{ ohms}$$

and hence

$$i = \frac{110}{1{,}000} = 0.110 \text{ amp}$$

The potential difference across the resistor is 110 volts. The inductive reactance at resonance is

$$2\pi f_0 L = \sqrt{\frac{L}{C}} = 1{,}000 \text{ ohms}$$

and the value of the capacitive reactance is therefore $-1{,}000$ ohms.

Hence $$V_L = V_C = V_R = 110 \text{ volts}$$

***34.5. Rotating-vector Diagrams.** Since the voltage-current relations in an a-c circuit are so complicated, some graphical representation of them is very desirable. It is possible, of course, to plot the current and the potential differences as functions of time, as was done for two simple cases in Figs. 34.3 and 34.4. This is a time-consuming procedure, however, and a simpler method is advantageous. The vector impedance diagram (Fig. 34.5) can fortunately be modified for this purpose. It will be recalled that a convenient representation of SHM is obtained from a uniform circular motion. If we have a point revolving about the origin with a constant angular velocity $\omega = 2\pi f$, then the projections of this point on the x and y axes move with SHM. Suppose we consider the potential difference expressed by Eq. (34.4). In Fig. 34.6 the projection of the vector V_m on the x axis is $V_m \cos 2\pi ft$, and this projected length varies with time in the same manner as does the potential difference V. If it is desired to represent on the same diagram a current or another potential difference, then another vector may be added whose length is proportional to i_m, for example. Since the phase difference between the voltage and current remains constant, the second vector has a constant angle with respect to the first and the combination rotates as a whole about the origin. It is customary to choose for the representation that particular instant of time when one vector lies along the x axis and then to imagine the rotation of the whole picture. To represent the applied potential difference of Eq. (34.4) and the current in the circuit given by Eq. (34.7), let us choose the instant $t = 0$. The vector diagram is then shown in Fig. 34.7. The phase angle ϕ has been taken positive, and hence Fig. 34.7 represents an inductive circuit with a lagging current. Figure 34.7 thus corresponds to Fig. 34.3.

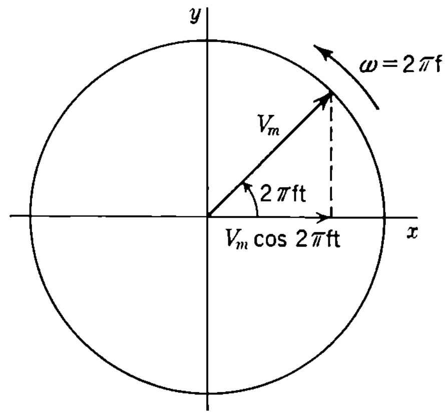


FIG. 34.6. Representation of an alternating potential difference.

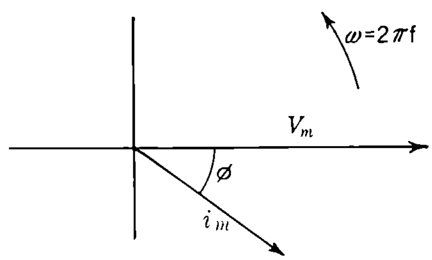


FIG. 34.7. Vector diagram of the current and voltage in an inductive circuit.

To represent the potentials across the components of the series circuit, we utilize Eq. (34.16). If the vector impedance diagram of Fig. 34.5 has all the quantities multiplied by the current i, then it becomes a diagram of potential differences. The diagram is redrawn in Fig. 34.8 with the proper change in notation. Since the current i has the same phase as the potential V_R across the resistor, the vector $\mathbf{i}_m$ has been added along the x axis. It should be noted that Fig. 34.7 shows the same situation as Fig. 34.8 at a *previous time*. Figure 34.8 is drawn for the time when the phase of the current is zero, or when

$$2\pi ft = \phi$$

Figure 34.7 was drawn for the time $t = 0$.

Another example of a vector diagram can be constructed for the resonance case discussed in Sec. 34.4. For the frequency of resonance, $V_L = V_C$, and $\phi = 0$. The vector diagram is shown in Fig. 34.9.

It is evident that vector diagrams can be constructed using either the maximum values of the quantities or the rms values. The difference is only a difference in scale. If a current and a potential difference are represented on the same diagram, a separate scale must, of course, be used for each.

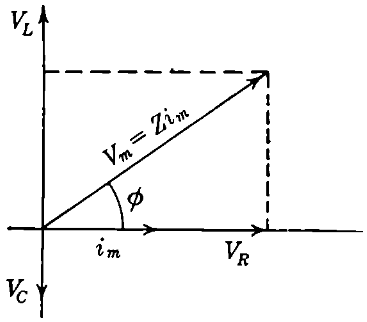


FIG. 34.8. Vector diagram of potential differences corresponding to Fig. 34.5.

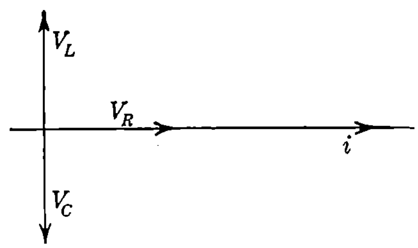


FIG. 34.9. Vector diagram at resonance.

34.6. Power Relations. The power delivered by an a-c generator to a circuit varies with time just as the current and potential difference vary. The power P is given by

$$P = Vi \qquad (34.17)$$

where V and i are instantaneous values. If V and i are given by

$$V = V_m \cos 2\pi ft$$
$$i = i_m \cos (2\pi ft - \phi) \qquad (34.18)$$

then the instantaneous power is

$$P = V_m i_m \cos 2\pi ft \cos (2\pi ft - \phi) \qquad (34.19)$$

In Fig. 34.10 the variations with time of V, i, and P are shown, with the assumption that the circuit is inductive and hence the angle ϕ is positive. The power P has a sinusoidal variation with time but varies about the average value $\bar{P}$ indicated by the dashed line.

The average power $\bar{P}$ dissipated in the circuit can be easily found by a trigonometric transformation of Eq. (34.19). Expansion of $\cos (2\pi ft - \phi)$

and the use of the relations

$$\cos^2 2\pi ft = \tfrac{1}{2} + \tfrac{1}{2}\cos 4\pi ft$$
$$\sin 2\pi ft \cos 2\pi ft = \tfrac{1}{2}\sin 4\pi ft$$

results in

$$P = \tfrac{1}{2}V_m i_m[\cos\phi + \cos(4\pi ft - \phi)] \tag{34.20}$$

The average power P is, therefore,

$$\bar{P} = \tfrac{1}{2}V_m i_m \cos\phi \tag{34.21}$$

The average power can also be expressed in terms of the rms voltage and current. Since

$$V_m = \sqrt{2}\,V_r$$
$$i_m = \sqrt{2}\,i_r$$

where V_r and i_r are the rms values,

$$\bar{P} = V_r i_r \cos\phi \tag{34.22}$$

The quantity $\cos\phi$ is called the *power factor* of the circuit. Since ϕ lies between $\pm 90°$, $\cos\phi$ lies between 0 and 1. The power factor is

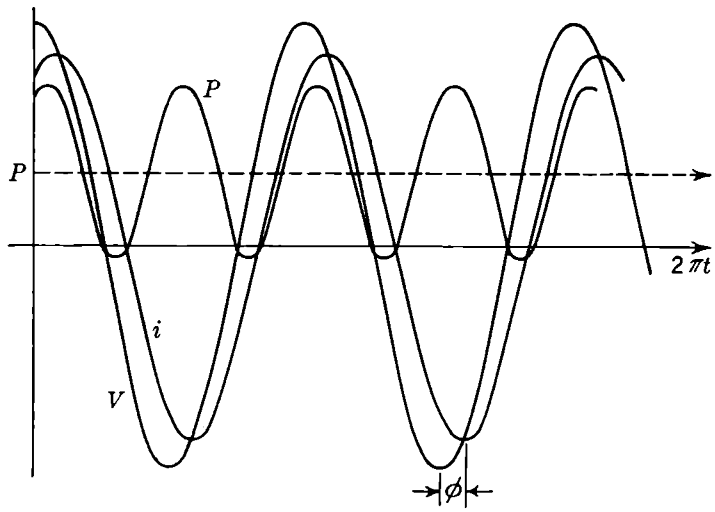


FIG. 34.10. The quantities P, V, and i in an inductive circuit.

often, therefore, expressed as a percentage. In a purely resistive circuit, $\phi = 0$, $\cos\phi = 1$, and the power absorbed by the circuit is the product of the effective voltage and current. An incandescent lamp is almost purely resistive at 60 cps. If the circuit contains reactance, the power factor is reduced. From Eqs. (34.13) and (34.14), we can express the power factor as

$$\cos\phi = \frac{R}{\sqrt{R^2 + X^2}} = \frac{R}{Z} \tag{34.23}$$

It is evident from Fig. 34.10 that part of the time the power P is negative. This means that, during this time, energy is being transferred from the circuit to the generator. The larger the reactive component of the circuit, the greater the time interval over which P is negative. If there is no resistive component, $\cos \phi = 0$ and the curve of P vs. t is symmetrical about the time axis. In a reactive circuit, then, energy is transferred from the generator to the circuit and back again. This transfer of energy performs no useful function, and power circuits are usually arranged to have power factors as large as possible. Since most motors are inductive, capacitors are often added to reduce the reactance in the circuit and to increase the power factor.

The wattmeter described in Sec. 30.7 measures the average power transferred since the deflection is proportional to the average of the product vi. The power factor can be determined by measuring V_r and i_r with a voltmeter and an ammeter and $\bar{P}$ with a wattmeter. The power factor is then determined from Eq. (34.22).

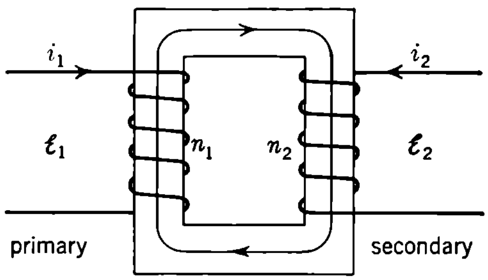


FIG. 34.11. A transformer.

34.7. Transformers. One of the principal reasons for the wide use of alternating current is the ease with which the potential difference or the current can be changed by means of transformers. A transformer consists of two coils closely coupled together by winding them on the same iron core, as indicated in Fig. 34.11. The coil to which power is applied is called the *primary* of the transformer; power is taken from the *secondary*. If the secondary coil is unconnected or open-circuited, the primary acts as a large inductance and a small current 90° out of phase with the applied potential flows. This current is called the *magnetizing current*. In a good transformer the inductance is large and the magnetizing current small. It is sufficient, however, to establish a large flux in the transformer core that varies sinusoidally with time. The changing flux induces emfs in both the primary and secondary coils. If the flux through the two coils is the same, then the counter emf in the primary and the emf in the secondary are simply proportional to the number of turns and

$$\frac{\mathcal{E}_2}{\mathcal{E}_1} = \frac{n_2}{n_1} \tag{34.24}$$

The quantity n_2/n_1 is called the *turn ratio* of the transformer. If the turn ratio is larger than 1, the transformer is a *step-up* transformer and $\mathcal{E}_2/\mathcal{E}_1 > 1$; if the reverse is true, the transformer is called a *step-down* transformer.

Suppose now that a load is connected to the secondary and a current i_2 flows. By Lenz's law the current is in the direction to decrease the flux through the secondary coil and hence also through the primary. With a smaller flux the counter emf of the primary is reduced, and the current in the primary increases until the flux is restored to its original value. If the magnetizing current and hysteresis losses in the core are neglected (a

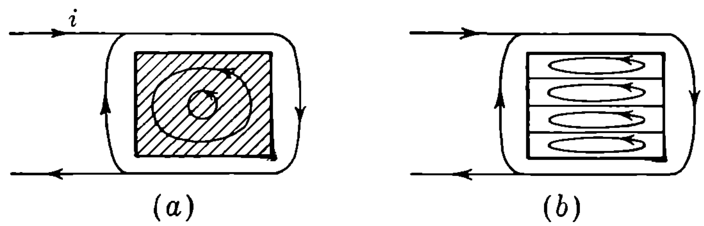


FIG. 34.12. The use of laminations to supress eddy currents.

justifiable procedure in well-designed equipment), by conservation of energy we must have

$$\mathcal{E}_1 i_1 = \mathcal{E}_2 i_2 \tag{34.25}$$

Consequently, if Eq. (34.24) is used,

$$\frac{i_1}{i_2} = \frac{n_2}{n_1} \tag{34.26}$$

The current is thus decreased in the same ratio as the emf is increased.

In practice, transformers usually have efficiencies above 90 per cent and, in large installations, even up to 99 per cent; consequently the simple relations just derived are good approximations. It is necessary, however, to avoid the presence of large induced emfs in the transformer core itself. If the core were made from a solid piece of iron, it would act as a single short-circuited secondary turn on the transformer. Large currents would flow in the iron, and the i^2R losses in the core would be excessive. Such currents are called *eddy currents*. Figure 34.12*a* shows these currents in a cross section through the primary coil. The eddy currents are minimized in transformers and other iron-cored coils by constructing the core of thin laminations, as indicated in Fig. 34.12*b*. The laminations are coated with some electrical insulator such as shellac, and the eddy currents are confined to circulate within a single lamination. Since, for a given flux density, the induced emf is proportional to the area, the magnitudes of the currents are very small in a lamination of small thickness.

Although the laminations in a transformer are tightly fastened together, the core is not perfectly rigid. The forces between the magnetized parts are usually strong enough to produce appreciable vibrations in the core, which result in the familiar transformer hum. The hum has a fundamental frequency of twice that of the alternating current.

One common application of transformers is to reduce the losses in a long transmission line. In a generator it is desirable to have fairly low potential differences between the parts of the circuit since clearances are usually small. The low emf and large current of a generator are changed by a transformer to a high emf and low current for transmission on the line. The losses on the line are proportional to the resistance of the line and to the square of the current, and hence a low current is desirable. At the end of the line, another transformer is used to lower the emf again.

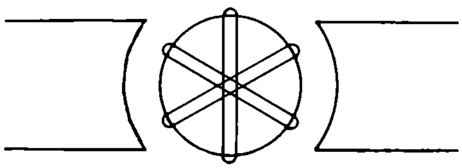


FIG. 34.13. A three-phase generator.

***34.8. Three-phase Currents.** If three coils are wound on the armature of an a-c generator so that the plane of one coil is separated by 60° from the plane of the next, as indicated in Fig. 34.13, then by means of a pair of slip rings for each coil three alternating emfs are produced that differ in phase. The polarity of the coils may be so chosen that a vector diagram of the three emfs is as shown in Fig. 34.14. It is possible, however, to interconnect the armature coils so that only three slip rings and three wires are necessary. The resultant of the vectors $\mathbf{V}_2$ and $\mathbf{V}_3$ in Fig. 34.14 is equal and opposite to the vector $\mathbf{V}_1$ and is indicated by the dashed vector. Therefore, if the coils are connected as shown in Fig. 34.15, the potential differences between any pair of the wires 1, 2, and 3 are all equal. The circuit on the left is known as a *delta* connection, that on the right as a *Y connection*. Loads can be connected to any pair of wires and power drawn from the generator. It is, of course, desirable to have equal currents through the three wires. When this condition is achieved, the loads are said to be *balanced*.

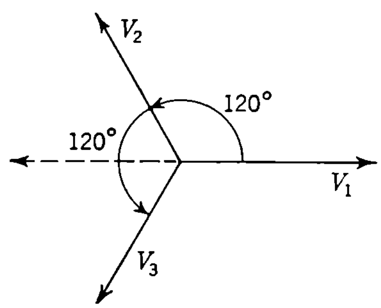


FIG. 34.14. Vector diagram of three-phase emf's.

The most important application of three-phase power is the *induction motor* The three emfs V_1, V_2, and V_3 differ in phase by 120°. Each emf is applied to a stator coil of an induction motor to produce a field that differs in phase by 120° and also differs by 120° in azimuth around the shaft from the fields of the two other coils. The coils might be arranged as in Fig. 34.16. Actually a six-pole arrangement is ordinarily used. The three field coils are connected in either the Y or delta connection, and the directions of winding are such that, when the pole (1) is a north pole, poles (2) and (3) are south poles.

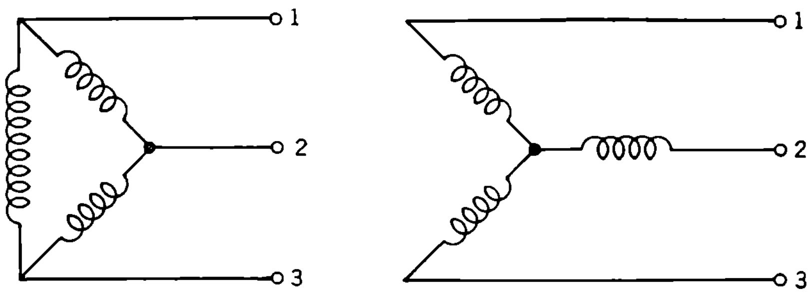


FIG. 34.15. The delta and Y connection.

The resultant field can be found most easily analytically. Let us choose axes as shown in Fig. 34.16, and compute the x and y components of the field. The fields of the separate poles are

$$\left.\begin{aligned} B_1 &= B_0 \cos 2\pi ft \\ B_2 &= B_0 \cos\left(2\pi ft - \frac{2\pi}{3}\right) = B_0\left(-\frac{1}{2}\cos 2\pi ft + \frac{\sqrt{3}}{2}\sin 2\pi ft\right) \\ B_3 &= B_0 \cos\left(2\pi ft - \frac{4\pi}{3}\right) = B_0\left(-\frac{1}{2}\cos 2\pi ft - \frac{\sqrt{3}}{2}\sin 2\pi ft\right) \end{aligned}\right\} \qquad (34.27)$$

The x component of the resultant field is

$$B_x = \frac{\sqrt{3}}{2} B_2 - \frac{\sqrt{3}}{2} B_3 \qquad (34.28)$$

and the y component is

$$B_y = -B_1 + \tfrac{1}{2}B_2 + \tfrac{1}{2}B_3 \qquad (34.29)$$

If the values of B_1, B_2, and B_3 are inserted from Eqs. (34.27), we find

$$\left.\begin{aligned} B_x &= \tfrac{3}{2}B_0 \sin 2\pi ft \\ B_y &= -\tfrac{3}{2}B_0 \cos 2\pi ft \end{aligned}\right\} \qquad (34.30)$$

The angle θ that the resultant B makes with the x axis is given by

$$\tan\theta = \frac{B_y}{B_x} = -\cot 2\pi ft$$

and hence $$\theta = \frac{\pi}{2} + 2\pi ft \qquad (34.31)$$

The field therefore *rotates* with the same frequency as the frequency of the alternating current.

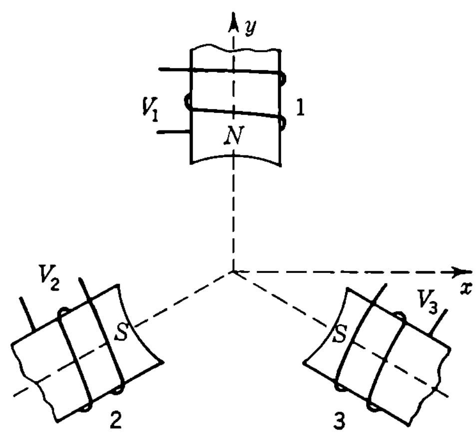


FIG. 34.16. The production of a rotating field.

The armature of an induction motor is constructed of a laminated iron core with heavy copper bars inserted parallel to the shaft and connected together by heavy copper rings at the ends, as shown schematically in Fig. 34.17. From the shape of the armature such a motor is called a "squirrel-cage" motor. When the armature is subjected to a rotating field, currents are induced on the rotor and exert torques on the motor shaft. If there is no load on the motor, the shaft rotates with the alternating frequency. When a load is present, there is some *slip* between the shaft and the field.

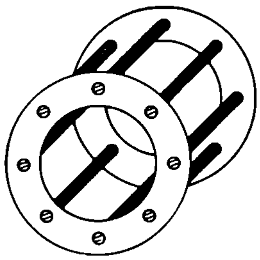


FIG. 34.17. The armature of an induction motor.

Two-phase circuits and motors are also in common use. The field coils are placed 90° apart on the stator, and the phase difference of the emf is also 90°.

34.9. Oscillations. If the series circuit of Fig. 34.2 contains no generator, but initially the capacitor C is charged, it will discharge through the circuit. The variation of the charge q with time is given by Eq. (34.6), with the driving emf set equal to zero, or

$$L\frac{d^2q}{dt^2} + R\frac{dq}{dt} + \frac{q}{C} = 0 \tag{34.32}$$

This equation will be recognized as the equation for damped oscillations discussed in Sec. 12.10. The solution is

$$q = q_0e^{-\alpha t}\cos 2\pi ft \tag{34.33}$$

where q_0 is the charge on the capacitor at $t = 0$ and α and f are

$$\alpha = \frac{R}{2L} \tag{34.34}$$

and

$$f = \frac{1}{2\pi}\sqrt{\frac{1}{LC} - \frac{R^2}{4L^2}} \tag{34.35}$$

If the resistance is not too large and $1/LC > R^2/4L^2$, the charge varies sinusoidally with time but the oscillations are damped and die out, as shown in Fig. 34.18. If $1/LC = R^2/4L^2$, the variation of q is critically damped. If $1/LC < R^2/4L^2$, the oscillations are overdamped and the decay is less rapid than for the case of critical damping. It should be noted that the frequency of free oscillation given by Eq. (34.35) is somewhat less than the frequency of resonance of the circuit when driven, as given by Eq. (34.10).

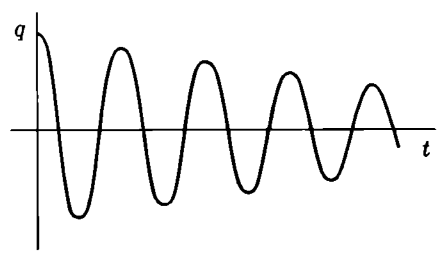


FIG. 34.18. The charge q vs. t in damped oscillations.

It is instructive to consider these oscillations from an energy standpoint. Initially the electric energy in the circuit is all stored in the capacitor C; and if the charge is q_0 initially, this energy is

$$W_E = \frac{1}{2}\frac{q_0^2}{C} \tag{34.36}$$

When current flows in the circuit, energy is transferred from the capacitor to the inductance. The current i, by differentiation of Eq. (34.33), is

$$i = -q_0 e^{-\alpha t}(2\pi f \sin 2\pi f t + \alpha \cos 2\pi f t) \tag{34.37}$$

The current first reaches a large and negative value at the time t_0 given, approximately for small α, by

$$\boxed{2\pi f t_0 = \frac{\pi}{2}} \tag{34.38}$$

At this time, by Eq. (34.33), $q = 0$. If α is not too large, this value of i is approximately

$$i = -q_0 e^{-\alpha t_0} 2\pi f \tag{34.39}$$

The magnetic energy W_H stored in the inductance, at this time, is

$$W_H = \tfrac{1}{2}Li^2 = \tfrac{1}{2}Lq_0{}^2(2\pi f)^2 e^{-2\alpha t_0} \tag{34.40}$$

Since α is assumed to be small,

$$(2\pi f)^2 \approx \frac{1}{LC}$$

and

$$W_H = \frac{1}{2}\frac{q_0{}^2}{C} e^{-2\alpha t_0} \tag{34.41}$$

which is, except for the exponential factor, equal to the electric energy W_E that was stored in the capacitor. The difference in energy is the energy dissipated in the resistance R. After the current reaches its maximum value, it decreases again and energy is returned to the capacitor. Oscillations of an electric circuit can thus be regarded as a fluctuation of energy from the electric field between the plates of the capacitor to the magnetic field surrounding the inductance. Each time the transfer takes place, some energy is transformed into heat in the resistance.

The rate of decay of oscillations is commonly described in terms of the Q, or Q factor, of the circuit according to the general definition

$$Q = 2\pi f \frac{\text{energy stored in the circuit}}{\text{energy lost per sec}} \tag{34.42}$$

The value of Q can now be found. The energy stored in the circuit initially is given by Eq. (34.36). The energy lost in the interval t_0 is the difference between Eqs. (34.36) and (34.41); hence

$$\text{Energy lost per sec} = \frac{W_E - W_H}{t_0}$$

$$= \frac{1}{2}\frac{q_0{}^2}{Ct_0}(1 - e^{-2\alpha t_0})$$

For small values of α, this can be approximated[1] as

$$\text{Energy lost per sec} = \frac{1}{2}\frac{q_0^2}{C}2\alpha$$

The value of Q is therefore

$$Q = \frac{2\pi f}{2\alpha} \tag{34.43}$$

The student can easily show that alternative forms for Q are

$$Q = \frac{X_L}{R} = \frac{2\pi fL}{R} = \frac{\sqrt{L/C}}{R} = \frac{1}{2\pi fRC} \tag{34.44}$$

PROBLEMS

1. How much inductance must be placed in series with a capacitance of 100 $\mu\mu$f for the combination to be resonant at a frequency of 1 Mc/sec?

2. Draw the vector impedance diagram for a resistance of 100 ohms, an inductance of 1 mh, and a capacitance of 1,000 $\mu\mu$f in series, at a frequency of 60 cps. What are the magnitude and phase of the impedance?

3. Draw the rotating-vector diagrams corresponding to Figs. 34.3 and 34.4.

4. Draw the impedance diagram corresponding to the first example in Sec. 34.4.

5. What is the power factor in each of the two worked examples in Sec. 34.4?

6. An ammeter connected in series with a motor reads a current of 5 amp, and a voltmeter across the motor terminals reads 110 volts. A wattmeter measures the average power consumed to be 490 watts. What is the power factor?

7. Explain why a motor is an inductive load.

8. Transformers are often designed for 60 cps, with 4 turns per volt. To what flux does this correspond? How many turns would there be on the primary and secondary coils of a power transformer to step down from 2,200 to 110 volts?

9. A resistance of 100 ohms is connected to the secondary of a 2:1 transformer. What resistance connected across the primary voltage source would show the same current as the transformer? (Answer: 25 ohms.)

***10.** Prove analytically that the sum of two components of a three-phase emf is equal to the third component.

11. The frequency of resonance of 1 henry, 1 μf, and 1,000 ohms in series was shown in Sec. 34.4 to be 159 cps. What is the frequency of damped oscillations in this circuit? What is the Q of the circuit?

12. Calculate the average power developed in a resistor during 1¼ cycles.

13. The reactance of a motor running on 60-cps alternating current is 50 ohms. What is the effective inductance of the motor? How much capacitance should be added in series with the motor to give a power factor of unity?

14. A step-down transformer from 440 to 110 volts at 60 cps is connected to a resistor and an inductor in series. The resistor has a resistance of 60 ohms and the inductor has an inductance of 2 henrys and a resistance of 10 ohms. What is the magnitude and phase constant of the current in the secondary circuit; in the primary of the transformer?

[1]Note that $1 - e^{-2\alpha t_0} = 1 - (1 - 2\alpha t_0 + \cdots) \cong 2\alpha t_0$.

CHAPTER 35
ELECTRONICS

35.1. Thermionic Emission. It is not necessary to have a material conductor in order to have a flow of current. Edison found that current passes from a heated wire in vacuum to a positively charged electrode. No current flows to a negatively charged electrode. The current is caused by the motions of electrons that are emitted from the hot wire, or *filament*, and are attracted to the anode. This process is known as the *thermionic emission* of electrons. It is found that the number of electrons emitted per second by the filament is independent of the potential difference across the tube, provided that this potential difference is large enough, but depends sensitively on the temperature of the filament and on the material of which it is composed. Thermionic emission is analogous to the vaporization of a liquid. Within the metal the electrons have a distribution of velocities that is characteristic of the temperature. Those with the highest velocities can escape and form the thermionic current. The energy necessary to remove an electron from a metal is a quantity analogous to the heat of vaporization of a liquid. This energy is usually expressed by the difference in potential ϕ between a point inside the metal and a point outside. The energy to remove one electron is therefore $e\phi$. The quantity ϕ is called the *work function* of the material and is usually expressed in volts.

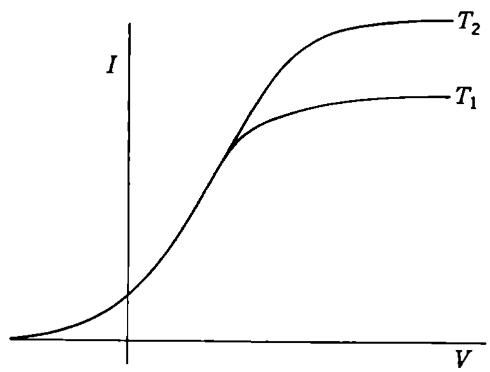


FIG. 35.1. Thermionic current for two temperatures of the filament.

If the variation of the thermionic current I is plotted as a function of the potential difference V across the tube, the curves shown in Fig. 35.1 result. The current is not zero when V is zero since some electrons possess sufficient energy to reach the anode, or *plate*, without the aid of an accelerating field. In fact, to stop the current completely a slightly negative potential is necessary. The shape of the curve for negative values of V can be used to find the distribution in velocity of the emitted electrons.

As the potential is increased, the current increases as shown, until it approaches a *saturation value* that is independent of V, and all the electrons emitted are collected by the anode. The saturation value does depend on the temperature, however, and increases with increasing temperature. A second curve is shown in the figure for a temperature $T_2 > T_1$. The magnitude of the saturation current is proportional to the area of the filament. O. W. Richardson and S. Dushman have shown, both theoretically and experimentally, that the saturation-current density J_s is given by

$$J_s = AT^2 \exp(-e\phi/kT) \tag{35.1}$$

where T is the absolute temperature and A is a constant that for pure metals has the value of 60 amp/cm². The quantity k is Boltzmann's constant. It will be recalled (Sec. 20.3) that k is the gas constant per molecule and that it has the value $k = 1.37 \times 10^{-23}$ joule/deg C. Since T and ϕ both occur in the exponent, J_s varies rapidly with both. The work function ϕ for pure tungsten has a value slightly larger than 5 volts. The presence of gas or other impurities on the emitting surface changes ϕ markedly. Since a large emission is obtained with a low work function, impurities that reduce ϕ are usually added. The presence of some thorium in the tungsten lowers ϕ to about 3 volts, and thus thoriated tungsten filaments are often used.

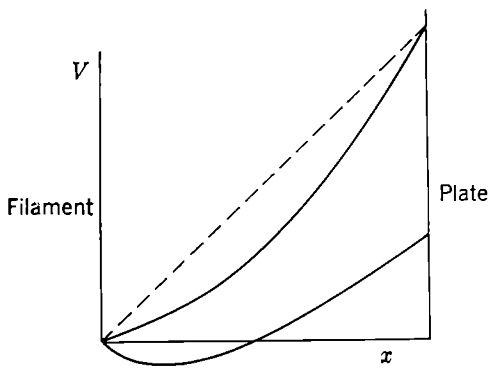


FIG. 35.2. The variation of potential with distance from the filament.

If a potential is applied to the anode that is not sufficient to produce saturation, some electrons are emitted that do not reach the plate but return to the filament. To understand the situation let us assume that the electrons are emitted from a filament which is in a plane and that the anode is a flat plate a small distance away. In the absence of any emission the electric field is uniform between the filament and plate, and the potential increases linearly across the tube, as indicated by the dashed line in Fig. 35.2. If some electrons are present, the potential is lowered below the dashed line, as indicated by the upper solid curve. If the potential difference is decreased below that necessary to produce saturation, the potential is indicated by the lower solid curve. This curve possesses a minimum, and the electric field (the slope of the curve) has opposite directions on the two sides of the minimum. On the right, electrons are drawn to the anode; on the left of the minimum the field accelerates electrons back toward the filament. At the minimum the field vanishes, and a cloud of electrons forms a *space charge* in this region.

35.2. Diodes and Rectification. Devices that employ thermionic emission are much in use today in control circuits and in communication equipment. The simplest of these devices consists of only a source of electrons and a collector in a vacuum tube. Since there are two electrodes, the tube is called a diode (Greek *hodos*, path). In most diodes, the filament is replaced by a cathode that is heated electrically by a separate heater circuit. The cathode surface is covered with a substance of low work function. A mixture of the oxides of barium and other alkaline earths on a nickel surface is ordinarily used. Often two separate diodes are enclosed within the same glass envelope.

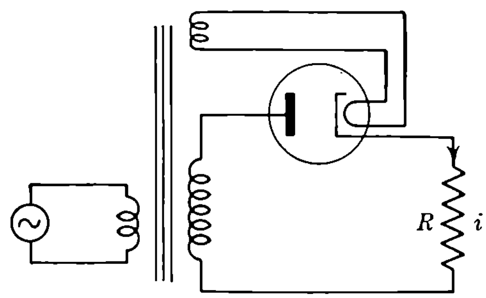


FIG. 35.3. A diode as a rectifier.

In diodes, and other thermionic tubes, the current is not linearly related to the potential applied to the tube—the relationship that exists is a complicated one. If the anode is made negative with respect to the cathode, no conduction occurs across the tube; and even with a positive potential difference across the tube the current I is not proportional to V, (see Fig. 35.1). Diodes are therefore *nonlinear* circuit elements, whereas resistors, capacitors, and inductances are linear. This nonlinearity is an extremely useful property but a difficult one to treat quantitatively.

Perhaps the simplest application of a nonlinear element is to produce a unidirectional current from an alternating one, or to *rectify* the current. If a diode is connected in series with a resistor to some source of alternating emf such as the secondary winding of a transformer, current flows through the diode only from the plate to the cathode, in the reverse direction to the electronic motion. Such a circuit is shown in Fig. 35.3, where the conventional symbol for a diode is used. The source of emf is the transformer. The current through the resistor R is shown in Fig. 35.4 as a function of time. The bottom halves of the sinusoidal waves are "cut off" by the nonlinear action of the diode, whereas the positive peaks are relatively unchanged. The resultant current is unidirectional although fluctuating. A diode used in this fashion is called a *half-wave*

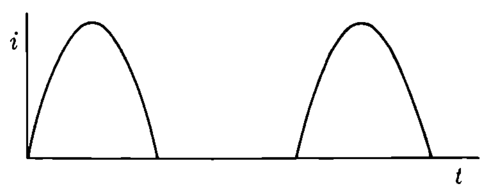


FIG. 35.4. The current i in the circuit of Fig. 35.3.

rectifier. In Fig. 35.3 the current for the cathode heater is shown supplied by a separate secondary winding on the transformer. The heater circuit may or may not be connected to the cathode.

Full-wave rectification can be obtained by employing two diodes in the circuit, as shown in Fig. 35.5. The current i through the resistor is shown in Fig. 35.6. The heater circuit is not shown in the diagram. The peak value of the current i is approximately $\mathcal{E}/R$ if the potential drop across the diodes can be neglected. The transformer winding must therefore produce a total emf of $2\mathcal{E}$. It should be noted that there is no connection between the wires that cross in Fig. 35.5. An electrical connection is indicated by a solid dot at the crossing.

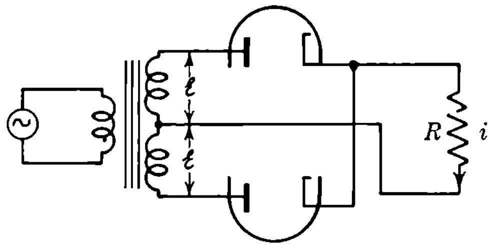


FIG. 35.5. Circuit for full-wave rectification.

If it is desired to produce not only a unidirectional current but also a steady one, a *filter* must be added to the circuit. The simplest filter is merely a capacitor connected in shunt with the load resistor R, as indicated in Fig. 35.7*a*. The capacitor C is charged by the current through the diodes and discharged through the resistor R. If the time constant RC is long compared with the period of oscillation of the alternating emf, a nearly steady current is obtained. If better filtering is desired or if larger currents are needed, iron-cored inductances in series and other capacitors in shunt are added

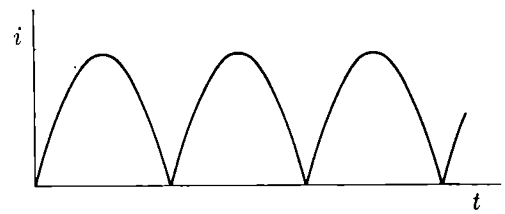


FIG. 35.6. Wave form of the current i in Fig. 35.5.

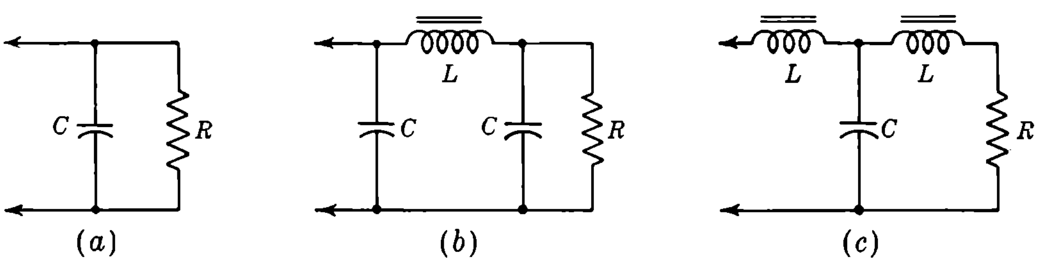


FIG. 35.7. Filters to produce a steady current from a full-wave rectifier.

to the filter, as indicated in Fig. 35.7*b* and *c*. The fluctuating current produces a counter emf in the inductances, which flattens the peaks in the current, whereas the capacitors serve to maintain the potential difference across R when the diodes are supplying only small currents.

35.3. Triodes. Lee deForest found that by introducing a third electrode between the plate and the cathode of a diode the current through the diode could be controlled. The third electrode consists of an open mesh of fine wires and is called a *grid*. The resulting thermionic tube is called a *triode*. Triodes are also constructed with a cylindrical cathode, having the heater inside, surrounded by a coaxial cylindrical plate. The grid is a helix of fine wire between the plate and cathode. The conventional circuit symbol for a triode is shown in Fig. 35.8. The plate current i_p of a triode is a function of two variables—the potential difference between the plate and cathode, e_p, and that between the grid and cathode, e_g. Figure 35.9 shows the current i_p as a function of e_p for a series of values of e_g taken with a commercial triode, type 6C5. Such curves are known as the *characteristic curves*, or *characteristics*, of the triode. Each curve is similar in shape to the lower portion of the diode characteristic shown in Fig. 35.1. Triodes are ordinarily used with the grid at a negative potential with respect to the cathode. Consequently no electrons reach the grid, and the grid current is very small. The potential of the grid can therefore be controlled by a very small current, and hence by a low power, to produce a relatively large change of current and power in the plate circuit.

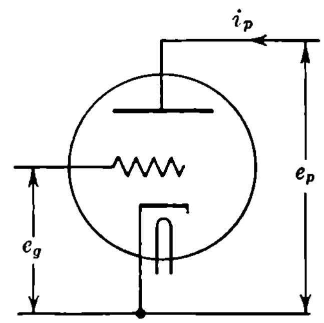


FIG. 35.8. Conventional symbol for a triode.

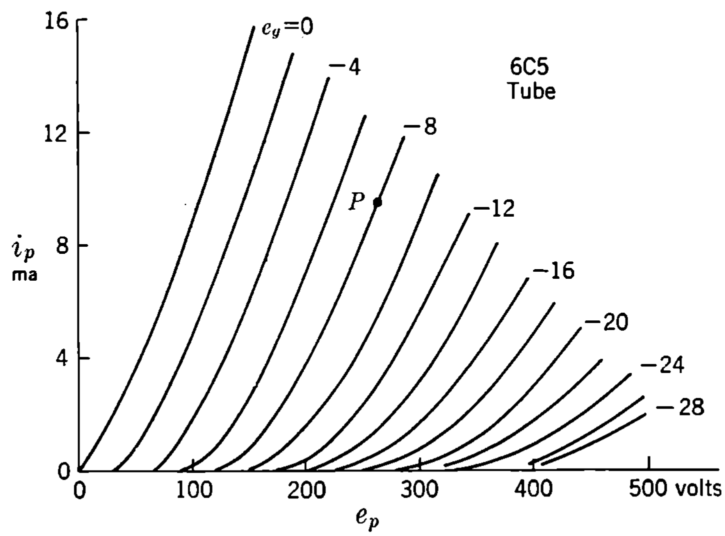


FIG. 35.9. Characteristic curves for a 6C5 triode.

The characteristics of Fig. 35.9 cannot be represented easily by analytic functions, but there are some useful parameters that represent the behavior of the tube. Triodes are usually used in circuits that are sensitive to *changes* in the values of e_g, e_p, and i_p rather than to the values of these variables themselves. Some operating conditions are chosen, such as those represented by the point P in Fig. 35.9, and the changes in the variables from the values at the operating point are utilized. The slope of the curve through P has the dimensions of a conductance, or the inverse of a resistance; hence we can write

$$\frac{\Delta i_p}{\Delta e_p} = \frac{1}{R_p} \qquad e_g = \text{const} \tag{35.2}$$

where R_p is called the plate resistance of the tube. In making the changes Δi_p and Δe_p the value of e_g is held constant, as indicated in the equation. In a similar fashion two other parameters can be defined as the ratios of changes in two variables when the third variable is held fixed. Thus

$$\frac{\Delta i_p}{\Delta e_g} = g_m \qquad e_p = \text{const} \tag{35.3}$$

$$\frac{\Delta e_p}{\Delta e_g} = \mu \qquad i_p = \text{const} \tag{35.4}$$

where g_m is called the *mutual conductance* and μ the *amplification factor*. The three parameters are not independent but are related by

$$\frac{\mu}{R_p} = g_m \tag{35.5}$$

It is found that, for a wide range of operating conditions, the values of the amplification factor μ and the mutual conductance g_m are very nearly constant. If only small changes in the variables are significant, the triode can be considered as a *linear* device. In actual practice the changes that are not small can be treated to a good approximation as linear. For the 6C5 tube with the characteristics shown in Fig. 35.9 the values of the parameters are $\mu = 20$, $R_p = 10{,}000$ ohms, $g_m = 2{,}000$ μmhos for the operating conditions of $e_g = -8$ volts, $e_p = 250$ volts, $i_p = 9$ ma.

***35.4. Triode Amplifiers.** A common application of a triode is as a voltage amplifier. A circuit diagram is shown in Fig. 35.10. The small alternating voltage to be amplified, Δe_g, is applied to the grid through the coupling capacitor C, which has a negligible impedance. The grid is maintained at the operating potential e_g applied through the resistor R_g, which is very large. The potential at the grid is thus the sum $e_g + \Delta e_g$. The plate potential of the triode is applied to the tube through the load resistor R_L from a source of emf equal to e_s. Since the steady plate current i_p flows through R_L, the value of e_p is less than e_s and is given by the equation

$$e_p = e_s - R_L i_p \tag{35.6}$$

When the grid potential changes in this circuit, a change in plate potential results, not only directly from the change in grid potential, but also from a change in the potential drop through the resistance R_L. Consequently, in terms of the tube parameters, we can write[1]

$$\Delta e_p = R_p \Delta i_p + \mu \Delta e_g \tag{35.7}$$

From Eq. (35.6) we also have

$$\Delta e_p = -R_L \Delta i_p \tag{35.8}$$

Elimination of Δi_p from Eqs. (35.7) and (35.8) results in

$$\Delta e_p = -\mu \Delta e_g \, R_L/(R_L + R_p) \tag{35.9}$$

The output voltage from the amplifier is taken through a second large coupling capacitor C'. The output voltage Δe_p is proportional to the input voltage Δe_g; hence let it be denoted by $G \, \Delta e_g$. The quantity G is called the *gain* of the amplifier, and

$$G = -\mu R_L/(R_p + R_L) \tag{35.10}$$

The negative sign in this expression means simply that, when the applied voltage on the grid is increased, the output voltage decreases. Several stages of amplification can be used together. If n stages are used, the over-all gain is G^n.

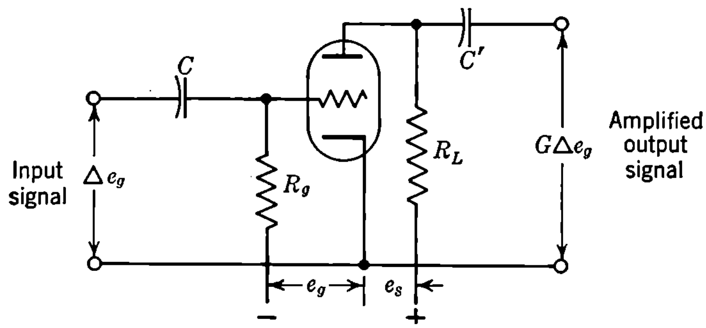


FIG. 35.10. A triode amplifier.

The relations just expressed analytically can also be obtained graphically from the characteristic curves. In Fig. 35.11 the line drawn through the value of the supply potential e_s with a slope of $-1/R_L$ intersects the characteristic curve of the chosen grid voltage e_g at the operating point of the triode, since the operating point must lie on the characteristic curve and must also satisfy the relation

[1]The student may recognize that Eqs. (35.2) to (35.4), which define the tube parameters, are expressed in terms of the *partial* derivatives of i_p and e_p. Equation (35.7) is the total differential of e_p, or

$$de_p = \frac{\partial e_p}{\partial i_p} di_p + \frac{\partial e_p}{\partial e_g} de_g$$

If the values of the partial derivatives are substituted from Eqs. (35.2) and (35.4), Eq. (35.7) results.

$$i_p = \frac{e_s}{R_L} - \frac{e_p}{R_L} \tag{35.11}$$

which is obtained from Eq. (35.6). If the grid potential is now changed to $e_g + \Delta e_g$ (on the figure this is shown as a negative change), the operating point shifts to P' and there is a change Δi_p in the plate current. The potential drop across R_L changes, and this change is indicated by $G\,\Delta e_g$.

It is evident from Eq. (35.11) that the gain increases with increasing R_L. For a given operating point, however, the plate-supply potential e_s must also be increased. It is often desirable to compromise by choosing $R_L = R_p$, and for this choice $G = \mu/2$.

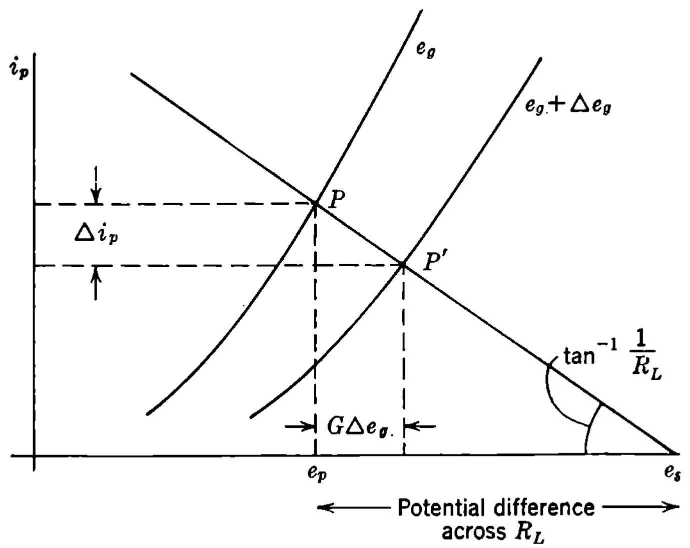


FIG. 35.11. Construction to locate the operating point of a triode.

★35.5. Pentodes. Thermionic tubes with more electrodes can be designed with characteristic curves of other desirable forms. In a pentode, two more grids are added between the control grid and the plate. The grid nearest the plate is called the *suppressor grid* and is maintained at cathode potential. It is used to slow down the electrons before they strike the plate, to prevent secondary-electron emission (cf. Chap. 36). Between the suppressor grid and the control grid is a *screen grid*. This grid is maintained at a positive potential and draws a small current. It acts as an electrostatic shield between the control grid and the plate and alters the distribution of space charge within the tube. In a pentode, very high amplification factors are obtained, and correspondingly large plate resistances. The characteristic curves of a 6SJ7 pentode are shown in Fig. 35.12.

The gain of a pentode used as a voltage amplifier is given by Eq. (35.10); but since μ and R_p are large, it is better to express the gain G in terms of the mutual conductance g_m. If Eq. (35.10) is multiplied and divided by R_p, it becomes

$$G = -\frac{\mu}{R_p}\frac{R_L R_p}{R_L + R_p} = -\frac{\mu}{R_p} 1/(1/R_L + 1/R_p)$$

If $R_p \gg R_L$, $1/R_p$ in the denominator can be neglected and we have

$$G = g_m R_L \tag{35.12}$$

For a 6SJ7 pentode, $g_m = 1{,}600$ μmhos. When the plate and screen-grid potentials are each 100 volts and the control grid is 3 volts negative with respect to the cathode, the plate resistance is about 0.7 megohm. Hence, with a load resistor of 100,000 ohms, a gain of nearly 160 can be attained.

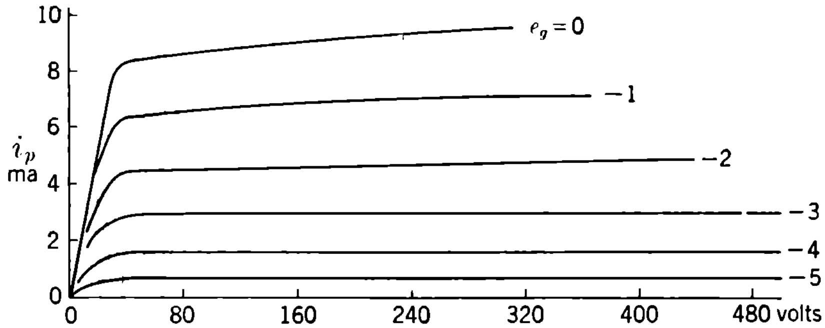


FIG. 35.12. Characteristics of a 6SJ7 pentode.

35.6. Oscillators. Thermionic vacuum tubes offer a convenient means for the generation of oscillations. As discussed in Sec. 34.9, in a circuit which contains resistance or in which there are other means whereby energy is lost the oscillations die out in course of time. If this energy can be supplied to the circuit, sustained oscillations are produced. One method of accomplishing this is shown in the circuit of Fig. 35.13.

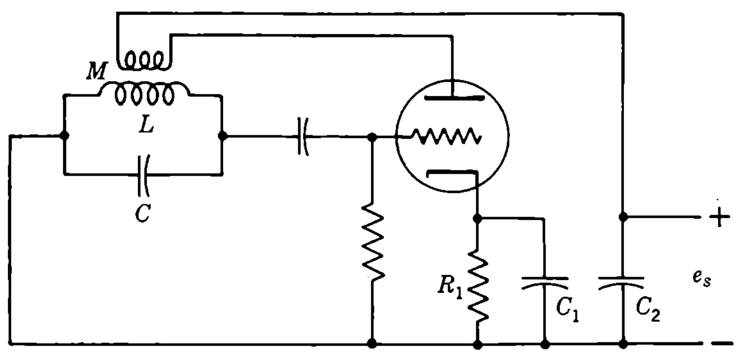


FIG. 35.13. The tuned-grid oscillator.

The oscillating circuit, or *tank* circuit, consists of the inductance L and the capacitance C. An oscillating voltage is applied to the grid of the triode through the coupling capacitor, and there results a change in plate current. The changing plate current produces a change in flux through the inductance L, and hence an emf in the oscillating circuit. The amplitude of the oscillations builds up until the energy lost is equal to the energy supplied by the tube. The necessary energy comes from the plate supply e_s. The proper grid potential is maintained by the potential drop of the steady plate current through the resistance R_1. The fluctuating component of the plate current is by-passed through the capacitors C_1 and C_2. Since the oscillating circuit is connected from grid to cathode, the arrangement is called a *tuned-grid oscillator*.

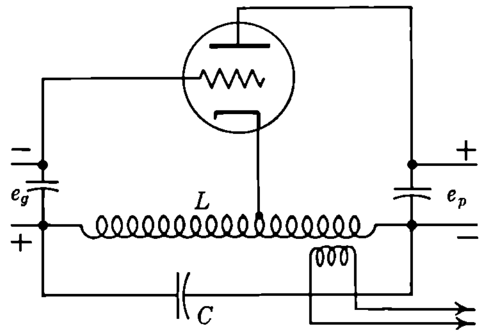


Fig. 35.14. A Hartley oscillator.

There are many variations in the circuits of vacuum-tube oscillators. Each must have a tank circuit and some means of supplying energy to it. One of the more common forms, called a *Hartley oscillator*, is shown in Fig. 35.14. Here the coupling between grid and plate circuits is provided by the mutual inductance between the two parts of the coil. Energy can be taken out of the circuit by means of the coil coupled to the tank circuit.

35.7. Modulation and Demodulation. If an oscillator such as that shown in Fig. 35.14 is used to produce signals for the transmission of

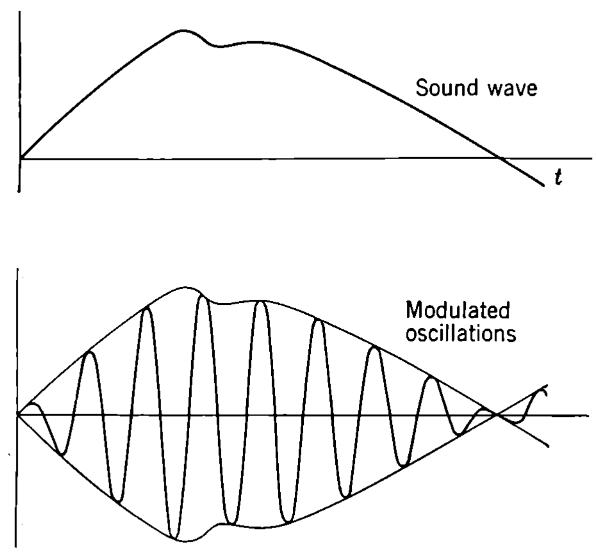


Fig. 35.15. Sound wave and modulated carrier.

intelligence from one place to another, it is necessary to vary the amplitude of the oscillations in some manner. This process is called *modulation*. The simplest method is to turn the plate-supply potential off and on to stop and start the oscillations. The Morse code can then be used to form letters and words.

To transmit speech the amplitude of the oscillations is varied in accordance with the form of the sound waves (see Chap. 38). The form of the sound wave and the modulated oscillations are shown in Fig. 35.15. The wave that is modulated is called the *carrier wave*. Amplitude modulation can be accomplished by means of the circuit of Fig. 35.16, which represents the grid-potential supply for an oscillator such as that shown in Fig. 35.14. A telephone transmitter is connected to the primary of a transformer. In the secondary of the transformer, there is an induced emf Δe_g of an amplitude proportional to the varying sound wave striking the telephone transmitter. The secondary is in series with the steady grid voltage supplied by a battery. The by-pass capacitance in the grid circuit of the oscillator must be large enough to have a small impedance for the carrier frequency, but the capacitance must be small enough not to short-circuit the speech frequencies. Since this modulation scheme varies the grid voltage of the oscillator, it is called "grid modulation." A similar arrangement can be used to vary the plate voltage, and *plate modulation* results. Considerably more power must be employed for plate modulation than for modulation of the grid.

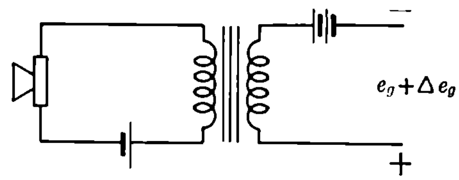


FIG. 35.16. Circuit for modulation of an oscillator.

It is profitable to consider a modulated wave analytically. Suppose the carrier frequency is f_1 and the amplitude A. The amplitude is modulated by a wave of frequency f_2, so that

$$A = A_0 \cos 2\pi f_2 t + B_0 \tag{35.13}$$

The carrier wave of instantaneous voltage e is then

$$e = A \cos 2\pi f_1 t = A_0 \cos 2\pi f_1 t \cos 2\pi f_2 t + B_0 \cos 2\pi f_1 t \tag{35.14}$$

A trigonometric transformation yields

$$e = \tfrac{1}{2}A_0 \cos 2\pi(f_1 - f_2)t + \tfrac{1}{2}A_0 \cos 2\pi(f_1 + f_2)t + B_0 \cos 2\pi f_1 t \tag{35.15}$$

The modulated wave is thus composed of three simple waves with frequencies $(f_1 - f_2)$, $(f_1 + f_2)$, and f_1. The first two waves are called the *side bands*.

Since the amplitude of the wave is varied during the modulation, the process is called *amplitude modulation*. A *frequency-modulated* wave is obtained by varying the frequency of oscillation instead of the amplitude. The resultant wave form can be expressed as

$$e = A \cos [2\pi t(f_1 + f_a \cos 2\pi f_2 t)]$$

To recover the original speech wave form from the modulated carrier the process of *demodulation*, or *detection*, is employed. In the simplest case a diode can be used as shown in Fig. 35.17, which is similar to a rectifier circuit. If a voltage wave form like that in the lower part of Fig. 35.15 is applied to the terminals, the current is zero over the negative half of the wave. The values of R and C must be so chosen that the capacitor has an impedance low compared with R at the carrier frequency but large compared with R at the modulation frequency. The potential difference across R is then the original modulating wave.

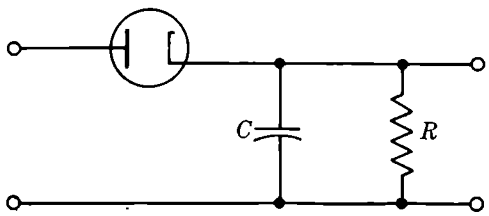


FIG. 35.17. Circuit for diode detection.

A triode can also be used for detection in the very similar circuit of Fig. 35.18. The grid voltage is chosen to be large and negative where the characteristic curve is most nonlinear. The RC network must satisfy the same conditions as for diode detection.

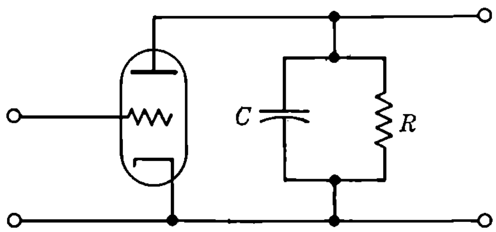


FIG. 35.18. Detection with a triode.

The action of the triode detector is understood most easily by consideration of another characteristic curve of the tube. If the plate current is

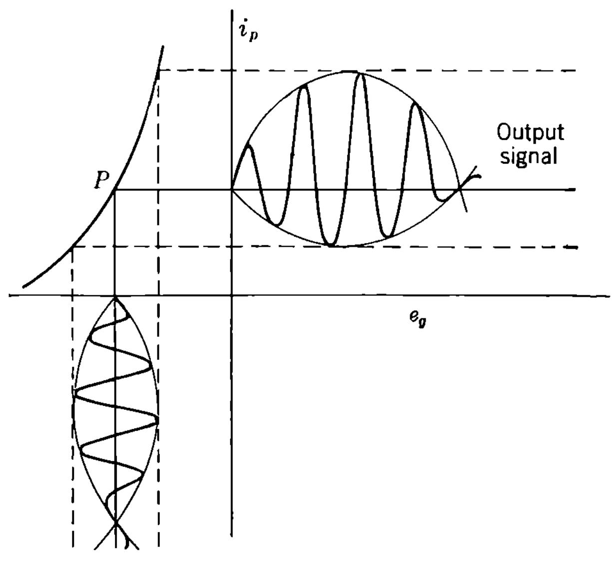


FIG. 35.19. Characteristics of the triode detector.

plotted against the grid potential as a parameter, a family of characteristic curves results. One of these is shown in Fig. 35.19. The operating point is shown as P, and the variations of the grid potential are indicated as the input signal. If vertical lines are drawn to the characteristic curve and then horizontal lines extended to the right, the wave form of the output signal is easily found graphically. A symmetrical input signal is transformed to an unsymmetrical output signal, and the action of the RC filter results in the reproduction of the original modulating wave.

35.8. Radio Transmission and Reception. An electric circuit in which an oscillatory current is flowing can lose energy by *radiation*. Electro-

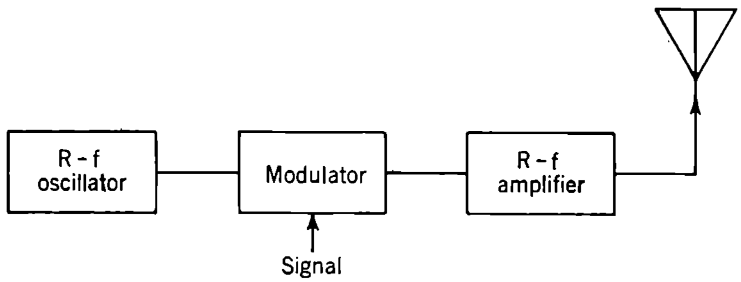


FIG. 35.20. Radio transmitter.

magnetic energy can exist in space, even in the absence of currents or metallic conductors. Just as the vibrations of a string produce sound, which transfers energy from the string to a sound detector, so an oscillating charge produces *electromagnetic radiation*, which takes energy from the circuit in which the charge flows. Electromagnetic energy travels with the velocity of light (Sec. 39.2) and can be used to transmit intelligence from one place to another.

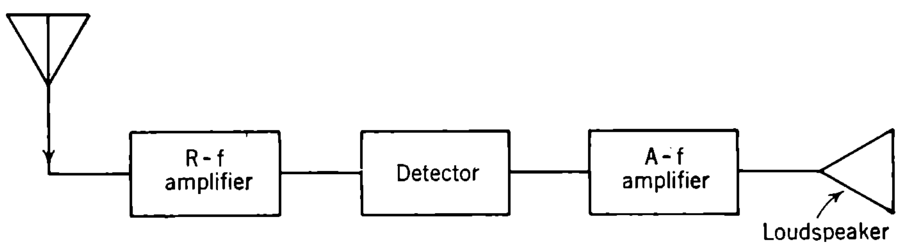


FIG. 35.21. Radio receiver.

The circuit elements described in the preceding sections can be combined to form a complete communication system. An example of such a system is afforded by a radiobroadcasting station and a home radio receiver. Figure 35.20 is the block diagram of a circuit that might be used by a broadcasting station, and Fig. 35.21 shows a typical receiver. The range of frequencies used for broadcasting extends from 0.5 to 1.5 Mc/sec. In both the transmitter and receiver shown in the figures, radio-frequency (r-f) amplifiers are used. The amplifier in the transmitter serves to increase the power level over that generated by the oscillator; in the receiver the r-f amplifier increases the power in the signal so that the detector operates efficiently. The detector is followed by an audio-frequency (a-f) amplifier, which amplifies the range of frequencies encountered in speech or music, from perhaps 100 cps to 8,000 cps. The quality of reception is largely determined by the width of the band of frequencies that are amplified and converted to sound in the loudspeaker.

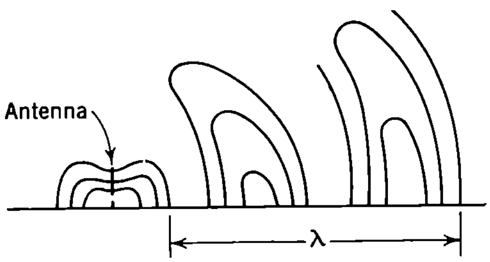


FIG. 35.22. The electric field near a broadcasting antenna.

Electromagnetic waves are produced electrically by means of *antennas*. A transmitting antenna often consists of a single vertical radiator, such as a steel tower, insulated from the earth. The electric field near such an antenna is shown in Fig. 35.22. Since the earth is not a perfect conductor, the electromagnetic waves are attenuated, not only by spreading out from the source, but also by heating losses in the earth. It should be noted that the field at the earth's surface is almost vertical, and consequently the most efficient antenna for reception is a single vertical wire.

Radio transmission is effective over much longer distances at higher frequencies than are used for broadcasting. Although the attenuation by the earth is greater at higher frequencies, electromagnetic waves are

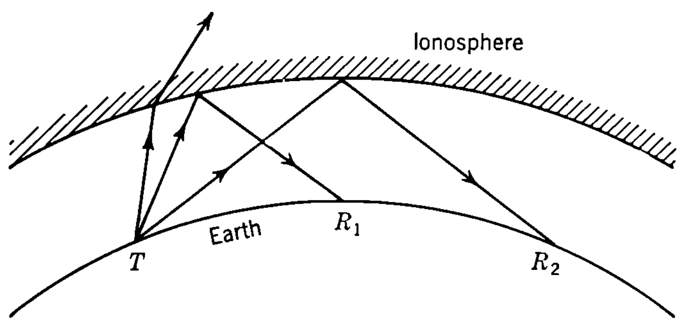


FIG. 35.23. The reflection of high-frequency waves from the ionosphere.

reflected from the *ionosphere* and reach large distances. At altitudes of approximately 100 km above the earth, free electrons and positive ions are present in a region known as the *Heaviside layer*, or the ionosphere. This layer acts as a good reflector for electromagnetic waves, as indicated in Fig. 35.23. Waves from the transmitter at T reach receivers R_1 and R_2 by the paths indicated. If the wave is incident on the ionosphere too close to normal, it is not reflected and a region exists between T and R_1

where the signal intensity is very low. The extent of this area is known as the *skip distance*.

Television signals are transmitted at still higher frequencies. More information must be transmitted per second to form a television picture with sound than for sound alone, and consequently a large bandwidth is necessary. A broadcasting station uses a band of only about 10 kc/sec whereas for television about 6 Mc/sec is needed. To accommodate several stations in a given area, the carrier frequency must be large compared with the bandwidth. Carrier frequencies between 50 and 200 Mc/sec are now in use. To transmit and receive such high frequencies effectively more complicated antennas are required than for broadcast frequencies. The antennas usually take the form of one or more pairs of conductors which produce electric fields like that of an electric dipole but which oscillate at the carrier frequency. Other conductors are arranged as reflectors for the purpose of focusing the beam of radiation. Because the frequency is so high, television signals are not reflected from the ionosphere but pass through it. The effective range of a television transmitting station is therefore rather limited. It may be shown to be approximately $\frac{4}{3}$ times the distance from the transmitting antenna to the horizon. To increase the range it has been suggested that transmitting stations be located in airplanes and the programs relayed from the ground. Seven or eight such airplanes would provide nearly complete coverage of the country. A brief description of a television system is given in Sec. 36.3.

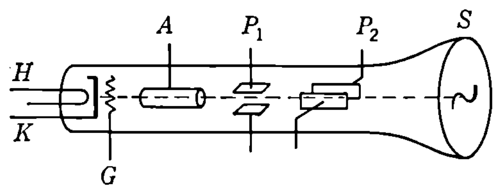


Fig. 35.24. A cathode-ray tube.

35.9. Cathode-ray Tubes. A thermionic tube of a type entirely different from those previously described is the cathode-ray tube shown in Fig. 35.24. A heater H heats the cathode K, which emits thermionic electrons. An anode A in the form of a cylinder, maintained at a positive potential with respect to the cathode, guides the electrons into a beam shown by the dashed line in the figure, and the intensity of the beam is controlled by the grid G. The beam of electrons falls on a screen S, which consists of a thin layer of zinc sulfide coated on the evacuated glass envelope of the tube. When the screen is struck by the electrons it emits green light, or *fluoresces* (Sec. 49.10). The position of the spot on the screen can be controlled by electric fields produced by the two pairs of plates P_1 and P_2. A field across the pair P_1 produces a vertical deflection; one across the pair P_2 makes the spot move horizontally. The deflection of an electron beam was discussed in Sec. 25.10. The name "cathode ray" is an old name for a beam of electrons.

One important application of a cathode-ray tube is as the indicator in an oscilloscope, a useful and interesting device for making visible the wave form of a fluctuating potential difference. If the potential difference to be observed is applied to the plates P_1 and a potential that increases linearly with time is applied to the horizontally deflecting plates P_2, the spot produced by the electron beam traces out the wave form. For instance, the variations of the potentials indicated in Fig. 35.25 produce the wave form shown on the tube in Fig. 35.24. If the saw-tooth wave form is repeated at the same frequency as the alternating potential, the trace on the screen is repeated each cycle and remains stationary.

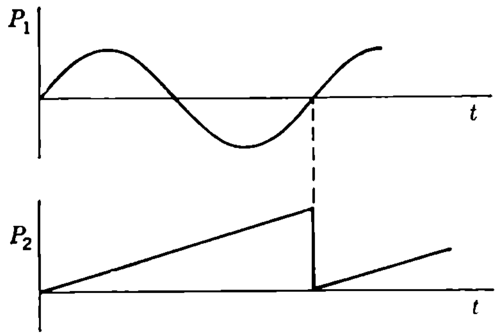


FIG. 35.25. The potentials applied to the plates of a cathode-ray tube in an oscilloscope.

Cathode-ray tubes are also made in which the electron beam is deflected by magnetic fields instead of electric fields. The magnetic fields are produced by pairs of coils outside the glass envelope of the tube.

***35.10. Radar.** During the Second World War the application of radio waves to detect and locate objects invisible in darkness or in fog was developed to a high degree of perfection. *RA*dio *D*etection *A*nd *R*anging, from which the name *radar* is derived, is a process quite similar to the locating of objects by visible light. Electromagnetic radiation of frequencies 100 to 10,000 Mc/sec can be focused into a sharp beam by an antenna of a reasonable size. A radiator placed at the focus of a parabolic mirror, for example, results in a sharp beam of radiation similar to the beam of light produced by a searchlight. The electromagnetic radiation is reflected from any object placed in the beam, and the reflected energy can be received by a radio receiver located near the transmitter. The direction of the reflecting object is then the same as the direction in which the beam from the transmitting antenna is pointed.

To find the distance, or range, of the reflecting object a time measurement is made. The electromagnetic energy is transmitted in the form of short pulses, of 1 to 10 μsec duration. The waves reflected from the object to be located are then received in short pulses, also. The difference in time between the transmitted pulse and the received echo pulse gives immediately the distance of the object, since the velocity of the pulses is known. If an interval of time t elapses, the distance d of the object is

$$d = \frac{ct}{2} \tag{35.16}$$

where c is the velocity of light. An interval of 100 μsec thus corresponds to a distance of 15,000 m.

The presence of an echo signal is detected, and the delay time is measured by the pattern produced on a cathode-ray tube. In the simplest system the antenna remains pointing in a fixed direction. The signals received are applied, after amplification, to the vertically deflecting plates, and a saw-tooth wave such as is shown in the lower part of Fig. 35.25 is applied to the plates for horizontal deflection. The frequency of the saw-tooth wave is made equal to the repetition frequency of the transmitted pulses. The pattern on the cathode-ray tube appears as shown in Fig. 35.26. The distances d_1 and d_2 of the echo pulses from the transmitted pulse on the screen are proportional to the distances of the reflecting objects. The screen of the cathode-ray tube can therefore be calibrated directly in miles or other convenient units.

In more complex radar systems the antenna is continuously and slowly turned in direction, so that objects at any azimuth can be located. A cathode-ray tube

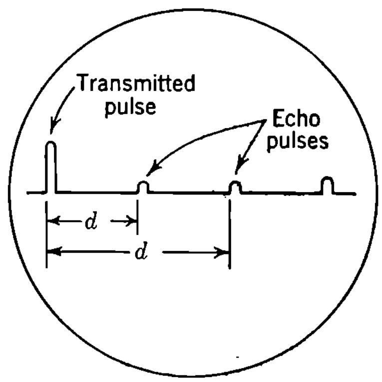


Fig. 35.26. Cathode-ray-tube screen in a simple radar.

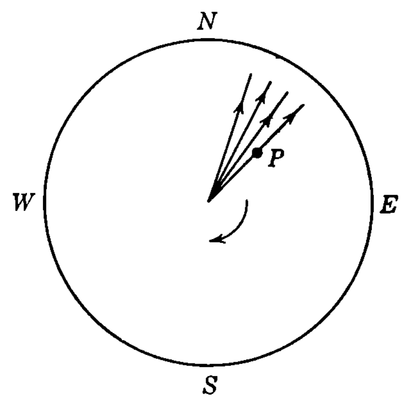


Fig. 35.27. The sweep traces on a plan-position indicator (PPI).

is again used as an indicator, but in a more complicated fashion. The proper potentials are applied to the deflecting plates to make the electron beam traverse with a constant velocity a series of radial lines about the center of the cathode-ray tube, as shown in Fig. 35.27. The direction of the sweep changes in synchronism with the direction of pointing of the antenna. The electron beam therefore completely scans the tube screen while the antenna scans in all directions. The azimuth of the antenna corresponds to the angle of the sweep on the tube. When no echoes are present, the grid of the cathode-ray tube is at a negative potential and the electron current is so reduced that no fluorescence is visible. When a signal is received, the potential of the grid is increased and a spot of light appears, as at P in Fig. 35.27. The distance of the spot from the center of the tube face gives the distance of the reflecting object. A tube used in this way is called a *plan-position indicator* (PPI). If a radar of this type is carried in an airplane, a map of the surrounding country is reproduced on the indicator tube.

PROBLEMS

1. Find the saturation current density for thermionic emission from a metal for which $\phi = 5$ volts and $A = 60$ amp/cm² at $T = 273°K$; 1000°K; 2000°K.

2. Compare the thermionic emission at 1500°K of a metal with a work function of 5 volts to one with $\phi = 3$ volts if A has the same value for both materials.

3. What is the average current through a half-wave rectifier if the peak current is 100 ma? What is it for a full-wave rectifier?

★4. What are the plate current and plate potential for a 6C5 triode with $e_g = -2$ volts, $e_s = 150$ volts; $R_L = 20{,}000$ ohms?

★5. What is the plate current in a 6C5 triode connected as in the circuit diagram of Fig. 35.28?

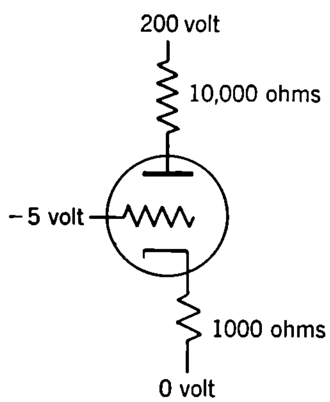


Fig. 35.28. Circuit for Problem 5.

6. The a-f amplifier of a radio receiver amplifies frequencies up to 8,000 cps. If the radio frequency is 1 Mc/sec, what range of frequencies should the r-f amplifier amplify?

★7. What is the length in meters of a radar pulse with a duration of 1 μsec?

★8. A five-stage amplifier has a total gain of 10^5. What is the gain per stage?

★9. A triode amplifier has $\mu = 10$ and $R_p = 80{,}000$ ohms. What gain does it have with a plate load resistor of 40,000 ohms?

CHAPTER 36

ELECTRONICS—Continued

36.1. X Rays. Electrons oscillating back and forth along a transmitting antenna produce electromagnetic radiation that travels outward in all directions. A beam of electrons, on the other hand, moving with a constant velocity produces no radiation. It is the change in velocity, or the acceleration, of charged particles that generates radiation. If a beam of fast-moving electrons is suddenly stopped by allowing the beam to strike a metal target, radiation does occur, called *X radiation*, or *X rays*. Radiation also occurs when a beam of electrons is deflected, as in a cathode-ray tube, or when electrons are accelerated or decelerated by electric fields. The intensity is so weak in these cases that the radiated energy cannot be detected. In a betatron, however, the centripetal acceleration of the electrons produces appreciable radiation.

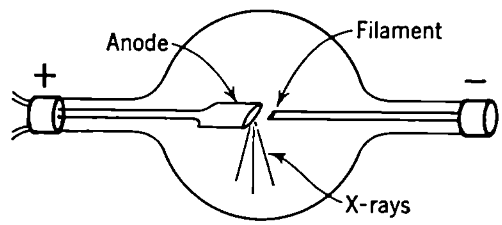


FIG. 36.1. A Coolidge X-ray tube.

The frequency of the X rays can be determined by diffraction methods described in Chap. 47. It is found that radiation of a range of frequencies is produced but that there is a well-defined upper limit to the frequency. The limiting frequency is found to be independent of the electron current but proportional to the energy W of the electrons. Hence

$$W = eV = hf_{max}, \tag{36.1}$$

where V is the potential that accelerated the electrons, f_{max} the frequency limit, and h the constant of proportionality. The constant h is the same Planck's constant that was encountered in Chap. 11. Its value is $h = 6.62 \times 10^{-27}$ erg sec. If $V = 10{,}000$ volts, $f_{max} = 2.4 \times 10^{18}$ sec^{-1}.

X rays of high frequency have the property of penetrating objects that are opaque to light, and their usefulness in medicine and many other fields is well known. They are also called *Roentgen rays* after the German physicist, W. K. von Roentgen (1845–1923), who discovered them in 1895. X rays are now produced in tubes of the type invented by Coolidge (Fig. 36.1). The source of electrons is a tungsten filament, and the anode of the tube is water-cooled to dissipate the electron energy that is not converted to radiation. Betatrons also can produce X rays of extremely high frequency, and hence great penetrating power, if the beam of electrons is allowed to strike a metal target. These very penetrating X rays are useful not only in the study of the interaction of matter and radiation but also for the inspection of large metal castings for flaws.

36.2. The Photoelectric Effect. The absorption of an electromagnetic wave by an antenna imparts velocity to electrons there, and currents flow. If the frequency of the wave is sufficiently high, a new phenomenon occurs. High-frequency radiation falling on a metal is found to cause the ejection of electrons from the metal. This is known as the photoelectric effect. The ejected electrons have a distribution of kinetic energies with an upper limit. If the intensity of the radiation is altered, the number of photoelectrons is changed but the distribution in energy remains the same. If, on the other hand, the frequency of the radiation is increased, the upper limit of energy is increased also. The radiation behaves as if it were composed of discrete bundles of energy, each bundle of energy being communicated as a whole to the electron. A. Einstein first explained the photoelectric effect in the following way: He postulated that the energy in each bundle, or *quantum*, is related to the frequency of the radiation by the equation

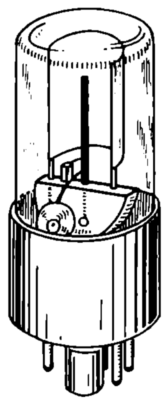


FIG. 36.2. A phototube; solid black line is anode.

$$W = hf \tag{36.2}$$

where h is Planck's constant. Quanta of radiation are also known as *photons*. The maximum energy that a photoelectron can have is hf less the energy necessary to remove the electron from the metal.

Hence

$$\tfrac{1}{2}mv_{\max}^2 = hf - e\phi \tag{36.3}$$

where ϕ is the work function[1] of the metal. Subsequent experiments have verified Einstein's postulates in detail.

Quanta of radiation also afford an explanation of the production of X rays. If an electron must produce at least one quantum of a given frequency, X rays of the maximum frequency are produced when the whole energy of the electron is converted to one quantum of radiation, and Eq. (36.1) follows directly. In X ray production the energy $e\phi$ is usually negligible compared with the electron energy eV.

Photoelectric tubes, or phototubes, are used in many electronic applications. They consist simply of a cathode of low work function and an anode to collect the electrons (Fig. 36.2). Since the electron

[1] See Sec. 35.1.

current is quite low, precaution must be taken to ensure good insulation. The photoelectric efficiency of a cathode, *i.e.*, the number of electrons produced per incident photon, varies with the frequency and with the cathode material. The sensitivity, expressed in amperes per unit light intensity, varies similarly. A curve showing this variation for a 929 phototube is given in Fig. 36.3. The sensitivity falls to zero at a frequency of 4.5×10^{14} cps. From Eq. (36.3) with v_{max} set equal to zero the work function ϕ of the surface is found to be 1.96 volts.

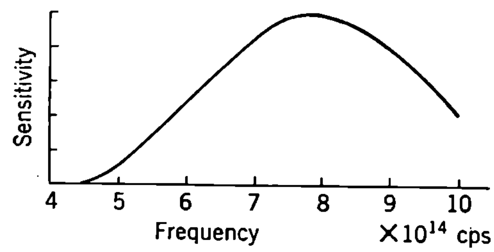


FIG. 36.3. The sensitivity of a 929 phototube.

Phototubes are used in the reproduction of sound for motion pictures. The side of the strip of film, or sound track, is exposed and blackened to an amount proportional to the instantaneous pressure in the sound wave. To reproduce the sound, light is sent through the sound track of the moving film and onto the cathode of the tube, as in Fig. 36.4. The variation of the photoelectric current through the resistor R produces a varying potential difference that is amplified and converted to sound in a loud-speaker.

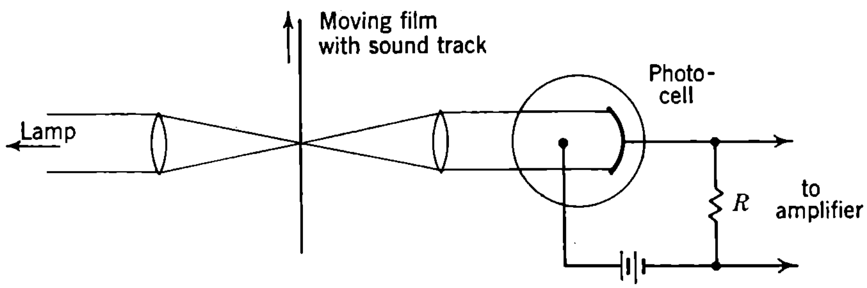


FIG. 36.4. The reproduction of sound with a phototube.

36.3. Television. A system to transmit visual images from one location to another requires three essential elements: (1) means of converting the image into a series of electrical impulses; (2) means for transmission of the electrical impulses; (3) a method of reconverting the electrical impulses into another visual image. The transmission of the electrical impulses can be done by conventional radio transmitters and receivers, and a cathode-ray tube can be used to produce a visual image. The conversion of the image to a series of electrical signals is done, in one system of television, by a tube known as an "iconoscope." The tube is shown schematically in Fig. 36.5. The image is projected on the screen S by a lens. The screen is composed of many tiny photoelectrically sensitive particles deposited on a thin sheet of mica. On the other side of the mica is a metal plate from which the electrical signals can be taken. When light falls on the screen, photoelectrons are emitted and the particles become positively charged. An electron beam, formed by a cathode K and two anodes A_1 and A_2, is swept across the screen in a series of horizontal lines by means of deflecting coils not shown in the figure. As the electron beam strikes a charged photosensitive particle, the particle is discharged and a pulse of current passes through the resistance R. The potential variations across R are amplified and used to modulate a radio transmitter. At the receiving station the electron beam of the cathode-ray tube moves in synchronism with the beam in the iconoscope, the synchronization being maintained by special synchronizing signal

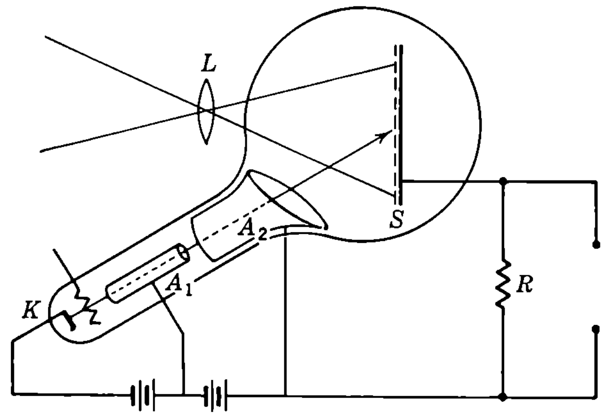


FIG. 36.5. The iconoscope.

pulses. The intensity of the cathode-ray-tube spot is varied by means of the grid potential to correspond to the light and dark portions of the image.

***36.4. Ionization by Collision.** In a tube traversed by a beam of electrons which is partially evacuated, the electrons interact with the molecules of the residual gas. If the electrons have sufficient energy, a collision between an electron and a gas molecule may result in the ejection of an electron from the molecule and the formation of a positive ion. Suppose the residual gas is argon. When the electrons possess more kinetic energy than 15.7 ev, the reaction during a collision can be described by the equation

$$e + \mathrm{A} + \mathrm{KE} \rightarrow 2e + \mathrm{A}^+ + \mathrm{KE} \qquad (36.4)$$

The kinetic energy on the right is divided between the two electrons. If there is a field in the region where the collision occurs, the two electrons move one way and the positive argon ion the other. Consequently the current in the electron beam is increased. The presence of the gas results in a current amplification.

This gas amplification from ionization by collision is utilized in several electronic devices. A small quantity of gas is introduced into some photoelectric tubes. The tube characteristic is altered as shown in Fig. 36.6. The rise of current at high potential differences across the tube is the result of gas amplification. Amplification factors of 5 to 10 can be attained with stable operation.

If a uniform electric field is present, the increase dN in the number N of electrons by ionization is proportional to the number present and is given by

$$dN = \alpha N\, dx \tag{36.5}$$

where α is a quantity dependent on the field, the pressure, and the nature of the gas. Equation (36.5) can be integrated, and there results

$$N = N_0 e^{\alpha x} \tag{36.6}$$

where N_0 is the number of electrons at $x = 0$. The increase with distance is consequently very rapid. The coefficient α also increases rapidly with the electric field. It is called the Townsend coefficient.

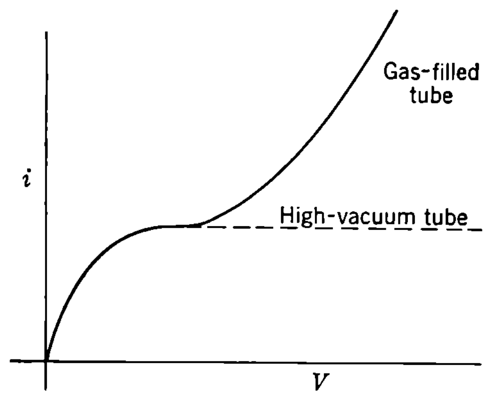


FIG. 36.6. Characteristic of a gas-filled phototube.

A second application of gas amplification is the *Geiger-Mueller counter* (see Fig. 36.7). If a fine wire stretched along the axis of a metal cylinder is at a positive potential with respect to the cylinder, a high electric field exists in the neighborhood of the wire. The wire and cylinder are contained in a glass envelope filled with gas to a pressure of about 10 cm Hg. If a single ion pair is formed within the volume of the cylinder, the electron is drawn to the wire. In the high-field region, gas amplification takes place that may be as large as 10^{10} or 10^{11}, and the resultant pulse of current through the resistor is easily detected and recorded by an electronic circuit. Geiger-Mueller counters are widely used to record the presence of ionization produced by fast-moving particles or by quanta of radiation.

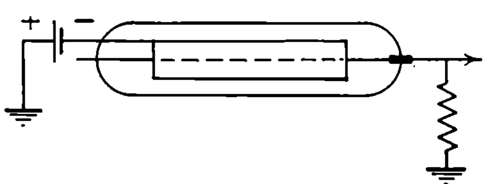


FIG. 36.7. A Geiger-Mueller counter.

If a diode or triode contains gas, the characteristics are altered but the device has other useful properties. If the gas amplification is large, the current carried by the electrons originally emitted from the cathode is a small part of the total current. Within the tube both positive and negative charges are present. The net charge at any place is small, and no space-charge effects exist. Much larger currents can be passed through the tube at low potential. Diodes containing mercury vapor are extensively used as rectifiers for obtaining large currents.

A triode containing gas can be made in which the starting of the current and the gas amplification can be controlled by the potential of a grid. Such a device is called a *thyratron*. Once the large current has started to flow in this tube, it is usually necessary to reduce the plate potential to stop it.

The rapid change in current with potential that is characteristic of gas amplification has been applied to obtain more efficient triode detectors. Such devices are little used today.

***36.5. Secondary-electron Emission.** Another phenomenon that can be used to amplify a current of electrons is the emission of electrons from a surface struck by an electron beam. The number of so-called *secondary* electrons produced in this way may be larger than the number of electrons bombarding the surface, and amplification is obtained. All surfaces show the phenomenon to some extent, but the number of secondary electrons is particularly large from beryllium and its alloys if the energy of the primary bombarding electrons is in the range 500 to 1,500 ev. At lower and at higher energies the number is less.

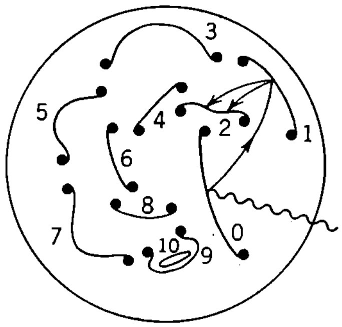


Fig. 36.8. The electrodes in an electron-multiplier tube.

Amplification by secondary emission is employed to increase the current in a phototube. A series of electrodes are used, and the electrons are accelerated between each pair of electrodes. The structure of one such tube is shown in Fig. 36.8; it is called an *electron-multiplier tube.* Photoelectrons are emitted from the electrode 0, and secondary emission occurs from electrodes 1 to 9; the final current is collected by the ring electrode 10. An amplification of between 10^5 and nearly 10^7 is obtainable with this tube.

***36.6. Self-sustaining Discharges.** At sufficiently high potentials a discharge can be maintained between two electrodes in a gas at reduced pressure without a supply of electrons from a heated cathode. The additional process that takes place is the emission of electrons from the cathode of the tube when the cathode is bombarded by positive ions. Such a discharge is called a *glow discharge.* Glow

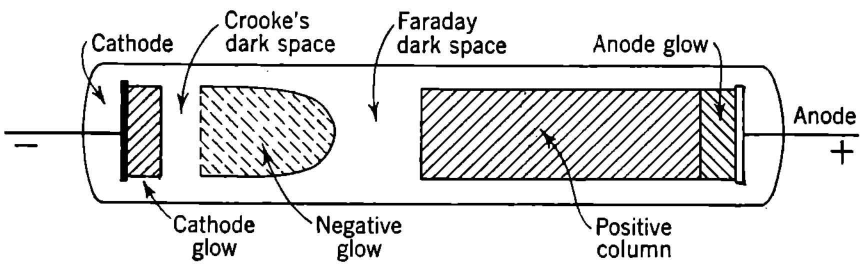


Fig. 36.9. The structure of a glow discharge.

discharges can exist only when both the current in the discharge and the potential across the discharge tube are above certain minimum values.

The structure of a glow discharge is quite complicated, and the appearance is usually striking. The various parts have been given names, as indicated in Fig. 36.9. The distance from the cathode to the positive column depends on the pressure and kind of gas but is nearly independent of the length of the tube. Nearly all the light comes from the positive column, and the length that this column can attain is familiar to everyone who has seen neon signs. Sometimes the positive column is broken up into alternate light and dark bands called *striations*.

A *fluorescent lamp* is a discharge tube with a coating on the inner surface of the tube that absorbs the ultraviolet light from the discharge and reemits it as visible light. The luminous efficiency is much greater than that of incandescent lamps. Since fluorescent lamps must be operated at low voltage, a filament to supply electrons is used at each end.

36.7. Photovoltaic Cells. Not all electronic devices employ the motion of electrons in a complete or partial vacuum; some involve the motion of electrons in solids. An example of such a device is known as a *photovoltaic cell*. A thin layer of cuprous oxide, Cu_2O, is coated on a disk of copper by heating. If light shines on the oxide, electrons are ejected from it, not into the surrounding air, but into the copper. The oxide layer thus becomes positively charged and the copper negatively charged. The light therefore generates an emf. If contacts are made to the oxide layer and to the copper, the photovoltaic emf produces a current. Other substances show this effect also, for example, a layer of selenium on iron.

Photovoltaic cells are commonly employed in exposure meters. A current meter is connected between the terminals of the cell, and the current indicated is proportional to the light intensity falling on the oxide layer. The meter can therefore be calibrated directly in light intensity.

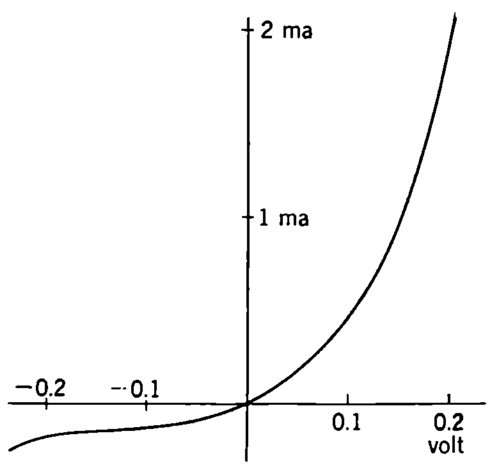


FIG. 36.10. The characteristic of a crystal rectifier.

***36.8. Semiconductors.** Many substances have electric conductivities intermediate in value between that of a pure metal and that of a good insulator. Examples are silicon, germanium, cuprous oxide, and selenium. Such substances are known as *semiconductors*. The conductivity of a semiconductor is usually the result of a few free electrons, rather than of many as in a pure metal. The number of such free electrons, and hence the conductivity, depends greatly upon the impurities in the material, the strains, and other factors that do not influence the properties of a good conductor to a large extent. Semiconductors are useful because the contact area between a metal and a semiconductor has a resistance that depends on the *direction* of the current across the contact. Such a contact can therefore be used to rectify an alternating current. The combination of cuprous oxide and copper that shows the photovoltaic effect also makes a good rectifier. A large area of contact can be employed, and the resistance to the current in one direction can be made small. Many devices employ copper-oxide rectifiers, which are particularly useful at low potential differences.

An excellent rectifier can also be made with a fine tungsten point touching a crystal of silicon or germanium. Since such a device is extremely small, it is useful for the rectification of currents of very high frequency. Crystal rectifiers have been successfully used in microwave radars operating at frequencies as large as 30,000 Mc/sec. Figure 36.10 shows the characteristic curve of a silicon crystal rectifier; this curve should be compared with that in Fig. 35.1. These devices are satisfactory only for small currents.

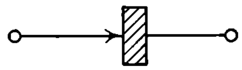


FIG. 36.11. Circuit symbol for a rectifying contact.

Contacts that rectify are indicated in electric circuits by the symbol shown in Fig. 36.11.

***36.9. Piezoelectricity.** Certain crystals that are insulators show electrical properties of another sort. If a potential difference is applied between the two sides of a crystal of quartz, the size and shape of the crystal change slightly, the nature of the change being dependent on the orientation of the crystal axes with respect to the field. The converse of this effect also occurs. If a crystal is subjected to a force that changes the shape of the crystals, a separation of charge takes place and a potential difference between parts of the crystal occurs. These phenomena constitute what is called *piezoelectricity*.[1] If an alternating potential difference is applied to a quartz crystal, the crystal undergoes forced oscillations. If the frequency of the alternations is correct, the phenomenon of resonance occurs. The internal friction of quartz is very low, and with the proper precautions the resonance can be made very sharp. A quartz crystal can thus be used as a standard of frequency. Clocks regulated by oscillating quartz crystals can be made accurate to within a small fraction of a second over a year of operation.

***36.10. Electrical Noise.** We have already noted that the electrons in a metal are moving in a manner similar to the molecules of a gas. The electrons possess large velocities of thermal agitation. Across any cross section of a resistor, electrons are moving in both directions, and the current that would result if this motion were in one direction only would be tremendously larger than any current passing through the resistor caused by an external emf. On the average, of course, the internal net current is zero. It is, at any one instant, however, not exactly zero, but flowing first one way and then the other. The ends of a resistor left unconnected in a circuit are therefore charged for short intervals of time and a potential difference V exists across the resistor. On the average V is zero; but again, at any one instant, it may have a small value positive or negative. As a measure of the fluctuations of V we can employ as usual the rms value $\sqrt{\overline{V^2}}$. It can be shown that the rms value of the potential difference depends only on the resistance R of the resistor and on the temperature T and is given by the formula

$$\sqrt{\overline{V^2}} = 2\sqrt{RkT(f_1 - f_2)} \qquad (36.7)$$

The constant k is Boltzmann's constant, and f_1 and f_2 are the limiting frequencies

[1] Greek *piezo*, pressure.

for which the apparatus that measures V is sensitive. For example, f_2 might be zero, and f_1 perhaps 10 kc/sec.

If an amplifier is connected to the terminals of a resistor and the output signal applied to a loud-speaker, sounds that can only be described as noise are heard. The fluctuating potential difference given by Eq. (36.7) is one source of this noise, which is called *thermal noise* (or *Johnson noise* after the physicist who made most careful measurements on its magnitude). In a similar fashion the plate current in a vacuum tube also fluctuates, since discrete amounts of charge are transferred by each electron, and *shot noise* is produced.

The presence of electrical noise of this sort constitutes a fundamental limitation to the sensitivity of amplifiers and the properties of all communication systems. Often, of course, interference occurs from other sources, such as thunderstorms or fluctuating power supplies, but that arising from the discreteness of the electronic charge cannot be avoided.

36.11. The Electron Microscope. A final illustration of the application of electronic motions is afforded by the electron microscope. If a

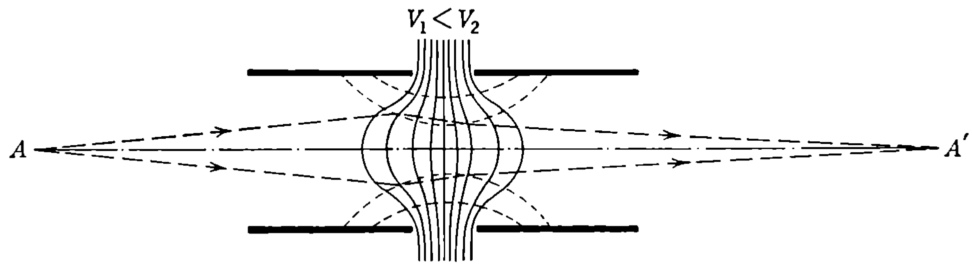


FIG. 36.12. An electron lens formed of two coaxial cylinders. The equipotential lines are solid; the lines of force dashed.

beam of electrons passes through a thin object, the beam is scattered in all directions. Each portion of the object scatters electrons in a manner characteristic of its structure, a part with a dense collection of atoms scattering more than a less dense part. The object thus modifies an electron beam in exactly the same manner as a light beam is modified by a translucent substance. It is possible, moreover, to focus a beam of electrons by means of electric and magnetic fields in the same way as a beam of light is focused by lenses. The electron beam can then be made visible with a fluorescent screen. By a combination of lenses a microscope can be constructed, and the image on the screen shows the structure of the object, magnified many times. We can thus "see" with electrons.

The simplest electron lens consists of two coaxial cylinders such as are shown in Fig. 36.12. The potential V_1 of the first cylinder is below the potential V_2 of the second cylinder, and both V_1 and V_2 are higher than the potential of the cathode, where the electrons start. In the figure the equipotential surfaces are shown as solid lines. The surfaces are figures of revolution about the axis AA'. The lines of electric force are normal to these equipotential surfaces and are shown as dotted lines. If an electron traverses the lens along the axis, the acceleration that it experiences is always parallel to the velocity and no deflection occurs. An electron traveling from A along the path indicated in the figure is accelerated toward the axis as well as along the axis as it approaches the lens. As the electron leaves the lens, it experiences an acceleration with a component away from the axis. Since the acceleration along the axis

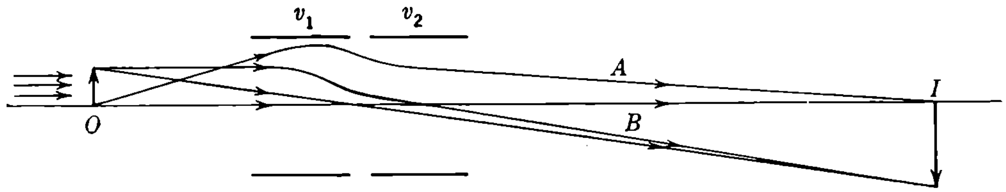


Fig. 36.13. The formation of a magnified image by an electron lens.

is always present, the electron spends less time in the right half of the lens than in the left half. The force away from the axis acts for a shorter time than that toward the axis, and the electron is deflected so that it strikes the axis again at the point A'. Electrons coming from the point A are focused to the point A'. In a similar manner a point near A is focused to another point near A', and the lens forms a magnified image of an object as indicated in Fig. 36.13. If the object at O is bombarded by a parallel beam of electrons from the left, it scatters the electrons and each point in the object acts as a point source of electrons. If a fluorescent screen is placed at I, a magnified electron image is made visible.

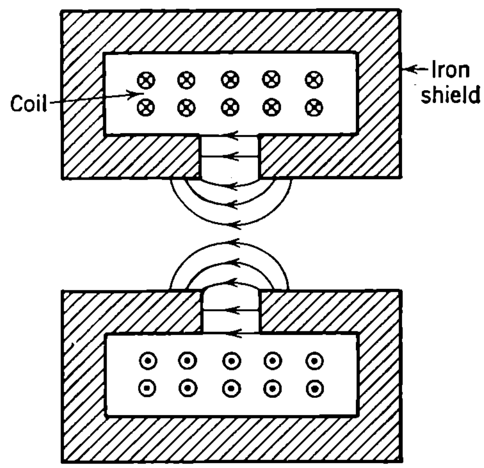


Fig. 36.14. A magnetic electron lens.

Electrons can also be focused by a magnetic field. An iron-clad coil of wire carrying a current, with a gap in the iron, produces a magnetic leakage field, as shown in Fig. 36.14. An electron coming from the left tends to spiral around the lines of induction and consequently receives a deflection toward the axis just as in the electrostatic lens.

In the electron microscope a combination of a heated cathode and an electron lens is used to "illuminate" an object with a parallel beam of electrons. Two more lenses, similar to the objective and the eyepiece of an ordinary microscope, then form a magnified image on a fluorescent screen. Since electrons impinging directly on a photographic emulsion expose it, the fluorescent screen can be replaced by a photographic plate and the image can thus be recorded directly. An important feature of an electron microscope is its high resolving power. This is discussed in Chap. 47.

PROBLEMS

1. It is desired to produce X rays with a maximum frequency of 10^{17} cps. What energy in electron volts should the bombarding electrons have?

2. What is the energy in one quantum of radiation at the following frequencies: 60, 10^6, 10^{15}, 10^{18} cps?

3. A phototube cathode has a work function of 4 volts. What is the maximum velocity of the photoelectrons from light of frequency of 3×10^{15} sec^{-1}?

★4. An electron ionizes by collision in a field strength for which $\alpha = 0.1$ cm^{-1}. How far must it travel to produce 100 ion pairs?

★5. An electron-multiplier tube has a gain of 10^6 in 11 stages. What is the gain per stage? If the gain per stage is increased by 10 per cent, what is the resultant total gain?

★6. What is the rms noise voltage generated in a resistor of 1 megohm over a frequency band of 1 Mc/sec?

7. A betatron accelerates electrons to 100 Mev, and they strike a target. What is the maximum frequency of the X rays produced?

8. The efficiency of a phototube is such that 1 electron is produced for every 40 photons falling on the cathode for light of frequency of 10^{15} sec^{-1}. What current is produced when 1 microwatt of such light strikes the sensitive surface?

9. A gas-filled photocell has an amplification of 10. If the ionization by collision takes place over a distance of 1 cm, what is the value of Townsend's coefficient α?

10. The maximum energy of electrons ejected from a photosensitive surface is 1.5 ev. If the incident light has a frequency of 2×10^{15} sec^{-1}, what is the work function of the surface?

CHAPTER 37

WAVE MOTION

37.1. Introduction. The phenomenon of wave motion is encountered in every branch of physics. Water waves are the most readily observed; but there are also sound waves in gases, liquids, and solids, waves in vibrating wires, rods, and diaphragms, radio waves, light waves, etc. Of recent years there have been many interesting and important developments with respect to the subject of sound, coming from research in telephony, radio, supersonics, acoustics of rooms, and military applications. Most of these advances in our knowledge of sound have resulted from the great developments in the field of electronics.

Waves originate in the displacement of some portion of an elastic medium from its normal position, causing the layers of the matter to oscillate about their equilibrium points. Because of the elastic properties of the substance, the disturbance is transmitted from one layer to the next, and so the wave form progresses through the medium. It is important to note that the layers, or elements, of the medium oscillate, or rotate, only in limited paths—there is no bulk movement of matter with the wave motion. For example, in the case of water waves the particles of water on the wave crest are moving in the direction of the wave, but in the trough the water particles are oppositely directed. In their participation in the passage of waves along the surface the particles of the water rotate only in elliptical paths while the wave form or configuration progresses.

By means of wave motion, energy may be transferred from one point to another; but the entire region, from the source of the energy to the receiver of the wave impulses, must be filled with a medium capable of transmitting such waves. Now one can produce a single wave pulse—for example, by striking a single blow against a taut rope. Another example would be the single shock wave resulting from an explosion. Or the periodic vibrations of a tuning fork or a telephone receiver can create similar periodic vibrations of the particles of the surrounding media, and a *train of waves* results. The single disturbance or the successive periodic vibrations are transmitted at a uniform speed determined by the proper elastic modulus and the density of the medium. Each layer transmits its energy, consisting of both kinetic energy of motion and potential energy due to its displacement from its equilibrium position, to the next layer in line. There is thus an energy flow from the vibrating source through the wave train to the receiver.

The wave pulse traveling along the rope is an example of a *transverse* wave, one in which the particles of the medium are vibrating perpendicularly to the direction of propagation of the wave. There may, of course, be a train of such waves sent down the rope, because of a periodically applied disturbance at one end, and these waves together with their reflections from the other end will produce visible *standing waves* in the rope (Sec. 37.10).

When a vibrating tuning fork or diaphragm sends waves into the surrounding elastic medium, usually air, the vibrations take place in the direction of propagation of the waves. No wave form is visible, there being merely periodic changes in the density of the medium. At the momentary positions where the density is slightly increased the layers of the medium are moving in the same direction as the disturbance. These positions are called *condensations*. Midway between these are positions of *rarefaction*, where the density is slightly less than normal and the layers are moving counter to the direction of propagation. Waves made up of such condensations and rarefactions are known as *longitudinal* waves. When they fall on the ear and excite the auditory nerve, we have the sensation of sound.

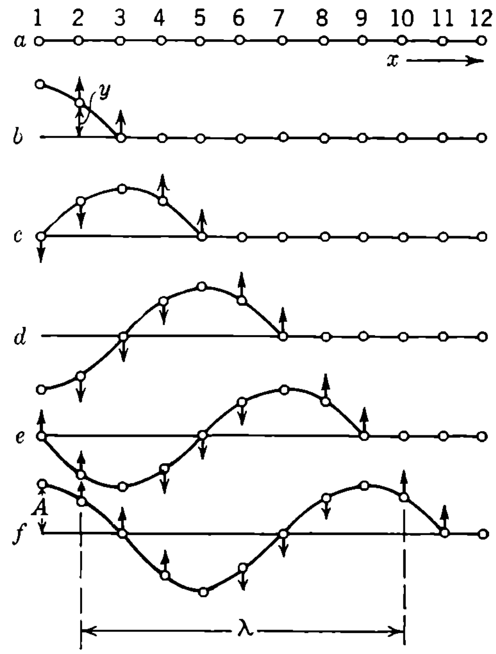


FIG. 37.1. The development of a simple harmonic wave in a row of elastically connected particles.

37.2. Waves from SHM. Since Hooke's law is almost always obeyed, the vibrations of the source will in most cases be of the simple harmonic variety. The vibrations of the particles of the elastic medium transmitting the waves will also be simple harmonic. If periodic vibrations are of more complicated form, these may be considered to be made up of a number of simple harmonic components (Sec. 37.3). Let us consider, then, a simple transverse wave set up in a row of particles of equal mass that are connected elastically. Suppose particle 1 to be given a transverse simple harmonic vibration. This will be communicated to particles 2, 3, etc., successively, and the wave form will develop as indicated in Fig. 37.1. In the wave as drawn, the phase difference between successive particles is one-eighth of a period of the SHM. The wave configuration has the shape of a sine or cosine curve always associated with simple harmonic vibrations.

The period P and amplitude A of the wave are the same as those of the individual particles in their transverse simple harmonic vibration. The wavelength λ is the distance between any two particles, such as 2 and 10 in Fig. 37.1f, which are in the same phase of vibration. In Fig. 37.1 it is seen that while one particle, such as 2, has made one complete vibration, the wave has advanced a distance λ. Denoting this constant speed with which the wave advances by v, we have then $\lambda = vP$. Since the frequency $f = 1/P$, this may also be written

$$v = \lambda f \tag{37.1}$$

The speed v is a function only of the properties of the elastic medium composed of these mass particles and is independent of the values of λ or f. Equation (37.1) is a perfectly general relation holding for all types of wave motion.

In Chap. 12 it was shown that the displacement y of a particle executing a simple harmonic vibration may be represented by

$$y = A \sin\left(\frac{2\pi t}{P} + \delta\right) \tag{37.2}$$

in which δ is the *phase constant*, or *phase at time* $t = 0$. If $y = 0$ when $t = 0$, then $\delta = 0$, and Eq. 37.2 becomes

$$y = A \sin\frac{2\pi}{P}t \tag{37.3}$$

If this is the equation representing the vibration of our reference particle, then for any other particle down the line that leaves its equilibrium position at a time t' later,

$$y = A \sin\frac{2\pi(t - t')}{P} = A \sin 2\pi\left(\frac{t}{P} - \frac{t'}{P}\right) \tag{37.4}$$

where $-(2\pi t')/P$ is the phase difference between this particle and the reference particle. Now, if x be the distance of this particle from the reference particle, it is seen from Fig. 37.1 that

$$\frac{t'}{P} = \frac{x}{\lambda} \tag{37.5}$$

Therefore we may rewrite Eq. (37.4) in the usable form

$$y = A \sin 2\pi\left(\frac{t}{P} - \frac{x}{\lambda}\right) \tag{37.6}$$

which is the equation of the transverse waves of amplitude A and period P being propagated down the row of elastically bound particles in the positive x direction. If we hold t constant and plot the values of y for successive increments in x, we obtain the wave configuration in space as in Fig. 37.1*f*. If, on the other hand, we hold x constant and allow t to increase, we may compute the transverse displacement at any instant of that particle which is vibrating with simple harmonic motion with a phase constant $-(2\pi x)/\lambda$.

The equation for a similar wave traveling in the negative x direction is

$$y = A \sin 2\pi \left(\frac{t}{P} + \frac{x}{\lambda}\right) \tag{37.7}$$

Since $\lambda = vP$, Eq. (37.6) may be written

$$y = A \sin \frac{2\pi}{P}\left(t - \frac{x}{v}\right) \tag{37.8}$$

It is to be emphasized that the constant velocity v with which the wave train proceeds has no relation to the velocity v_y of any of the mass particles in their vibratory motion. The latter we obtain from Eq. (37.6) by differentiation with respect to t (partial differentiation, holding x constant; such a partial derivative is denoted by $\partial y/\partial t$).

$$v_y = \frac{\partial y}{\partial t} = \frac{2\pi A}{P} \cos 2\pi \left(\frac{t}{P} - \frac{x}{\lambda}\right) \tag{37.9}$$

Similarly, $\partial v_y/\partial t$ gives the acceleration a_y of the particle at any instant in its transverse vibratory motion.

$$\begin{aligned} a_y = \frac{\partial v_y}{\partial t} &= -\frac{4\pi^2}{P^2} A \sin 2\pi \left(\frac{t}{P} - \frac{x}{\lambda}\right) \\ &= -\frac{4\pi^2}{P^2} y \end{aligned} \tag{37.10}$$

37.3. Complex Vibrations. Fourier Analysis. It is possible for a number of different vibrations to be traveling simultaneously through a medium. The resultant effect of the several coincident vibrations at any point may be obtained by plotting the separate vibrations with proper relative amplitudes and phases against the time and then adding algebraically the ordinates for each value of t. The important case for sound waves is where the frequencies of the several vibrations are all integral multiples of one fundamental frequency. These higher frequencies are called *harmonics* of the fundamental and will usually have smaller amplitudes.

In Fig. 37.2 are drawn three component vibrations with frequencies in the ratio 1:2:3 and amplitudes in the ratio 1:½:⅓. Note that t, not x, is plotted along the horizontal. The resultant vibration is shown as a heavy line. Even for this relatively simple case the resultant vibration begins to have a somewhat complicated contour, but it is periodic, its frequency being that of the longest component vibration in

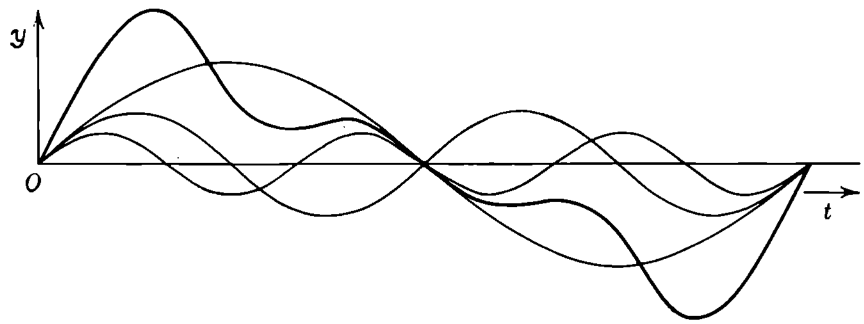


FIG. 37.2. The resultant of three superposed vibrations having frequencies in the ratio 1:2:3 and amplitudes in the ratio 1:½:⅓.

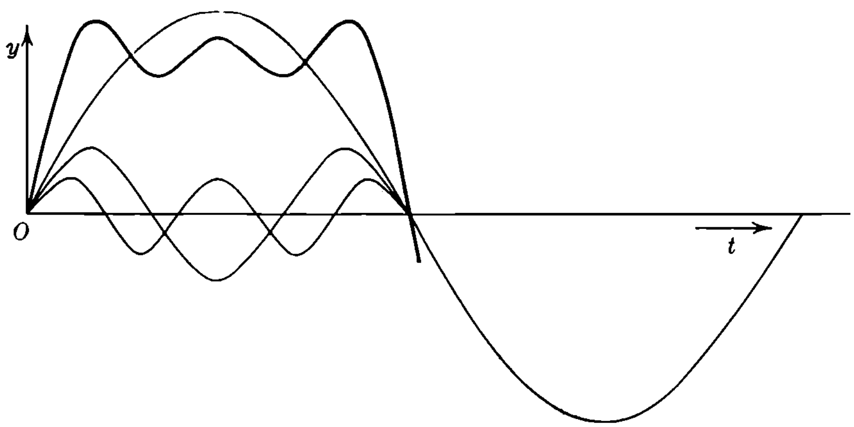


FIG. 37.3. The addition of three component vibrations with frequencies in the ratio 1:3:5 and amplitudes as 1:⅓:⅕ gives a resultant (heavy line) incipiently of square wave form.

the group. If the phase relations among the three component vibrations are different from those shown, the resultant vibration will have a different appearance but will have the same periodicity. The algebraic formula for the resultant vibration sketched in Fig. 37.2 is

$$y = A \sin \omega t + \frac{A}{2} \sin 2\omega t + \frac{A}{3} \sin 3\omega t$$

As a further example of the superposition of several vibrations, assume three component vibrations to have frequencies in the ratio 1:3:5 and amplitudes as 1:⅓:⅕. This situation is plotted in Fig. 37.3. If additional component waves of frequencies 7, 9, 11, . . . and amplitudes $\frac{1}{7}$, $\frac{1}{9}$, $\frac{1}{11}$, . . . times those of the fundamental were included, it would be found that the resultant would increasingly take on the appearance of a square wave as more components were added. The student should as an exercise plot at least one such resultant vibration with a different set of assumed relations among the amplitudes, frequencies, and phases of the three components (Probs. 2 and 3). It is also easy to verify graphically that, if the components differ in amplitude and phase but *not* in period, the resultant is a sine wave form of the same period but with some intermediate value of the phase.

Since we may build up a complex periodic vibration by the superposition of a number of simple vibrations, the question naturally arises as to whether the reverse process may be carried out—the decomposition of a complex vibration into component sine or cosine vibrations. The French mathematician, J. Fourier (1778–1830), was the first to show that this can be accomplished. According to *Fourier's theorem*, any smooth periodic function may be represented as the sum of a number of sine and cosine functions with frequencies that are multiples of one basic frequency. The resultant displacement of any particle in the medium transmitting the complex periodic wave is thus given by an equation of the form

$$y = A_1 \sin \omega t + A_2 \sin 2\omega t + A_3 \sin 3\omega t + \cdots + B_0 + B_1 \cos \omega t + B_2 \cos 2\omega t + B_3 \cos 3\omega t + \cdots \quad (37.11)$$

in which the periods and wavelengths of the components are in the ratio $1:\frac{1}{2}:\frac{1}{3}:$ . . . . The Fourier series representing a square wave alluded to in our discussion of Fig. 37.3 would be

$$y = A \sin \omega t + \frac{A}{3} \sin 3\omega t + \frac{A}{5} \sin 5\omega t + \cdots \quad (37.12)$$

If the periodic vibration is a fairly simple one, this mathematical analysis into the proper sine and/or cosine function components is not too difficult.

The best way to observe a complex sound wave is with a cathode-ray oscilloscope (Chap. 36). The sound is produced before a microphone and changed into an electric potential difference whose time variations are synchronous with those of the sound wave, and then the potential difference is applied to the pair of plates in the oscilloscope to give the electron beam its vertical deflections. To keep the wave form at rest on the oscilloscope screen the "sweep" frequency must be adjusted to be that corresponding to the periodicity of the complex wave.

37.4. Velocity of a Transverse Wave in a String. Continuing our discussion of one-dimensional transverse waves, we shall develop an expression for the velocity v with which a transverse wave pulse travels along a stretched string. An example might be the pulse set up in a violin string by light plucking or that sent along a taut rope by giving one end a quick sidewise jerk. We shall assume that the string, wire, or rope is perfectly flexible and uniform and that the displacements are small, and we shall neglect the effect of reflections at the ends. Our first proof of the expression for v is included because of its being an interesting exercise, using physical principles already emphasized in this text. A better proof, using calculus, follows in Sec. 37.5.

In Fig. 37.4 is depicted the wave pulse between a and b proceeding from right to left in the stretched string, with velocity v. Let us imagine the entire string to be moved from left to right with this same speed so that the wave form remains fixed while the particles composing the string successively round the curve. Such a stationary pulse can actually be produced in a long, loose pulley belt. A small section of the wave of length Δl may be considered to be an arc of a circle of radius r as indicated. If d is the linear density of the string, the mass of this element is $d\,\Delta l$. The tension F in the string is a tangential pull at each end of this small segment of the string. The components of these two forces along the vertical are equal, and their sum has the value

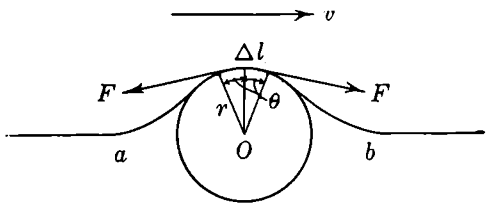


FIG. 37.4. A wave pulse in a stretched string.

$$2F \sin \theta = 2F \times \frac{1}{2} \frac{\Delta l}{r} = F \frac{\Delta l}{r}$$

These forces supply the centripetal acceleration directed toward O. By Chap. 7 the centripetal force acting on a mass $d\,\Delta l$ moving in a circular path of radius r with velocity v is $(d\,\Delta l\,v^2)/r$. We therefore have

$$F \frac{\Delta l}{r} = \frac{d\,\Delta l\,v^2}{r}$$

and thus

$$v = \sqrt{\frac{F}{d}} \tag{37.13}$$

showing that the velocity of a transverse wave in a string is a function only of the tension F and the linear density of the string. If F is in dynes and d in gm/cm, v is expressed in cm/sec. If F is in poundals and d in lb/ft, v is in ft/sec.

***37.5. The General Wave Equation.** Equations (37.6) and (37.7) are solutions in integrated form of a differential equation known as the *wave equation*. This equation may be obtained from Eq. (37.8) by differentiating twice both with respect to t and x. We must take *partial derivatives;* for since y is a function of the two independent variables t and x, we hold one of these constant while differentiating with respect to the other variable. From Eq. (37.8),

$$\frac{\partial y}{\partial t} = \frac{2\pi A}{P} \cos \frac{2\pi}{P} \left(t - \frac{x}{v}\right) \qquad (37.14)$$

and

$$\frac{\partial^2 y}{dt^2} = -\frac{4\pi^2 A}{P^2} \sin \frac{2\pi}{P} \left(t - \frac{x}{v}\right)$$

or

$$\frac{\partial^2 y}{dt^2} = -\frac{4\pi^2}{P^2} y \qquad (37.15)$$

Differentiating Eq. (37.8) twice similarly with respect to x, holding t constant, we obtain

$$\frac{\partial^2 y}{\partial x^2} = -\frac{1}{v^2}\frac{4\pi^2}{P^2} y \qquad (37.16)$$

Combining Eqs. (37.15) and (37.16) gives

$$\frac{\partial^2 y}{\partial t^2} = v^2 \frac{\partial^2 y}{\partial x^2} \qquad (37.17)$$

which is the equation in differential form of a transverse wave traveling in the x direction with velocity v.

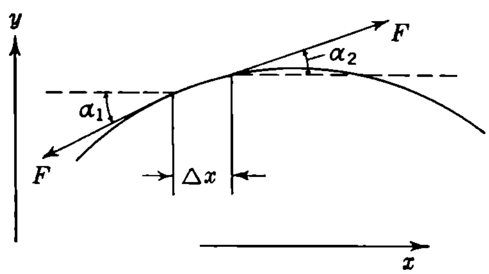


FIG. 37.5. Portion of the wave pulse in a stretched string.

Let us consider once again the wave in the stretched string. In Fig. 37.5 let Δx be an element of the string in the wave pulse. This element has an acceleration $\partial^2 y/\partial t^2$ in the y direction because of the difference in the y components of the two forces of tension F acting tangentially at the ends of the element. This difference is

$$F \sin \alpha_1 - F \sin \alpha_2 = F(\sin \alpha_1 - \sin \alpha_2)$$
$$= F\left[\frac{\partial(\sin \alpha)}{\partial x}\right] \Delta x \qquad (37.18)$$

Now, for small angles, $\sin \alpha = \tan \alpha$ and $\tan \alpha = \partial y/\partial x$. Therefore Eq. (37.18) may be written

$$F \sin \alpha_1 - F \sin \alpha_2 = F\left(\frac{\partial^2 y}{\partial x^2}\right) \Delta x \qquad (37.19)$$

This unbalanced force in the y direction may be set equal to the mass $d\,\Delta x$ of the element times its acceleration $\partial^2 y/\partial t^2$ along y. Therefore

$$F\left(\frac{\partial^2 y}{\partial x^2}\right) \Delta x = d\,\Delta x \frac{\partial^2 y}{\partial t^2}$$

or

$$\frac{\partial^2 y}{dt^2} = \frac{F}{d}\frac{\partial^2 y}{\partial x^2} \qquad (37.20)$$

Comparing Eq. (37.20) with Eq. (37.17) we obtain $v = \sqrt{F/d}$ just as in our deduction in Sec. 37.4.

One reason for referring to Eq. (37.17) as the *general* wave equation is that it can be satisfied by y being any function of $x \pm vt$. We have already mentioned that most sound waves are not just of the sinusoidal variety, although, to be sure, they may be analyzed into a number of simple sine and cosine terms. But for y in this equation one could substitute any periodically varying quantity such as the pressure variation in longitudinal sound waves in a fluid or the longitudinal displacements of the particles of the fluid. We proceed to use the equation in our discussion of sound waves in a fluid.

Worked Example. Calculate the total kinetic and potential energy per unit length of a transverse wave in a stretched string.

The E_{kin} per unit length of mass d is

$$E_{\text{kin}} = \frac{1}{2} d \left(\frac{\partial y}{\partial t}\right)^2 = \frac{1}{2} d \frac{4\pi^2 A^2}{P^2} \cos^2 \frac{2\pi}{P}\left(t - \frac{x}{v}\right) \tag{37.21}$$

Since the restoring force per unit length is $d(\partial^2 y/\partial t^2)$, which equals, by Eq. (37.15), $-d(4\pi^2/P^2)y$, the E_{pot} per unit length is

$$E_{\text{pot}} = \int_0^y d \frac{4\pi^2}{P^2} y \, dy = d \frac{4\pi^2}{P^2} \frac{y^2}{2} = \frac{1}{2} d \frac{4\pi^2}{P^2} A^2 \sin^2 \frac{2\pi}{P}\left(t - \frac{x}{v}\right) \tag{37.22}$$

Therefore

$$\begin{aligned} E_{\text{kin}} + E_{\text{pot}} &= \frac{1}{2} d \frac{4\pi^2}{P^2} A^2 \left[\cos^2 \frac{2\pi}{P}\left(t - \frac{x}{v}\right) + \sin^2 \frac{2\pi}{P}\left(t - \frac{x}{v}\right)\right] \\ &= \frac{1}{2} d \frac{4\pi^2}{P^2} A^2 = d \times 2\pi^2 f^2 A^2 \end{aligned} \tag{37.23}$$

The result that the total energy varies directly as the square of both the frequency and the amplitude is general, holding for all types of wave motion.

37.6. Sound Waves in a Fluid. In a fluid, either a liquid or a gas, sound waves are longitudinal. That is, the particles of the medium vibrate back and forth in the direction in which the waves are moving. In an extended medium the waves will expand radially in three dimensions, but to facilitate the analysis we shall consider our medium to be confined in a very long tube. The medium we suppose to be homogeneous, isotropic (same properties in all directions), and elastic. Since the only modulus of elasticity possessed by a fluid is a bulk modulus, compressional waves are the only kind a fluid can transmit.

Let us suppose that the tube pictured in Fig. 37.6 is closed at the left end with a piston that is made to shuttle back and forth periodically with a short stroke. The tube is filled with the elastic fluid medium and extends indefinitely to the right. We represent the successive layers of the fluid by the vertical lines. As the piston moves forward from B to B', the adjacent layers of the fluid are compressed and a compression pulse starts down the tube. On the return stroke of the piston from B' to B the particles of the medium move backward with it, expansion occurs, and a rarefaction starts down the tube toward the right. As indicated by the arrows in Fig. 37.6, the particles are always moving in the condensations c in the direction of propagation of the wave train, while in the rarefactions r the particles are moving in the reverse direction. During one complete stroke of the piston from B to B' and back to B the wave disturbance will have moved forward to the right a distance equal to one wavelength λ. The maximum forward velocity of the particles is at c in the middle of a condensation, but in the entire compression extending for half of a wavelength the particles are moving in the direction in which the wave train is proceeding. In the other half of each wave all the particles are moving backward, with the maximum velocity at r, the

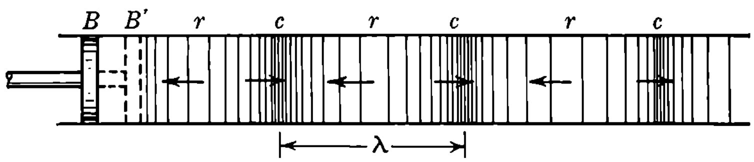


FIG. 37.6. The vibrating piston produces alternate condensations and rarefactions which travel to the right, with the layers of fluid in the condensations moving in the direction of propagation, while in the rarefactions the layers are in the reverse direction.

center of the rarefaction. These motions are superposed on the random thermal motion of the molecules of the fluid medium and are actually small mass movements of layers of the substance. As we shall show at the end of Sec. 37.7, however, there is some relation between the velocity v with which the waves are propagated in a gas and the kinetic-theory rms velocity of its molecules.

It can be shown that these compressional waves travel through the fluid with a velocity v given by

$$v = \sqrt{\frac{M}{\rho}} \tag{37.24}$$

where M is the bulk modulus of elasticity of the fluid and ρ is its density. Recalling that $M = p/(\Delta V/V)$, we can check Eq. (37.24) dimensionally as follows[1]:

$$v = \sqrt{\frac{[MLT^{-2}][L^3]}{[L^2][M]}} = \sqrt{[L^2T^{-2}]} = [LT^{-1}]$$

Similarly it can be shown that the velocity of propagation of compress-

[1] Note that M in the dimensional equation stands for mass, not modulus.

ional waves in a solid rod is

$$v = \sqrt{\frac{Y}{\rho}} \tag{37.25}$$

where Y is Young's modulus.

Worked Example. Let us compute the velocity of sound in water and in steel, using the last two equations. The bulk modulus for water (Table 4.1, page 54) is about 2.2×10^{10} dynes/cm², while, at 4°C, ρ is, of course, 1 gm/cm³. Substituting in Eq. (37.24),

$$v = \sqrt{(2.2 \times 10^{10})/1} \text{ or } 1.48 \times 10^5 \text{ cm/sec} = 1{,}480 \text{ m/sec.}$$

In water above 4°C, M increases with temperature while ρ decreases; therefore v increases when the water temperature rises above this value.

Young's modulus for steel is approximately 2×10^{12} dynes/cm² while 7.8 gm/cm³ may be taken for the density. From Eq. (37.25)

$$v = \sqrt{\frac{2 \times 10^{12}}{7.8}} \text{ cm/sec} = 5.06 \times 10^5 \text{ cm/sec} = 5{,}060 \text{ m/sec}$$

37.7. Velocity of Sound in Gases. The application of Eq. (37.24) to gases, especially air, merits more detailed consideration. If the compressions and expansions in the layers of gas were isothermal and Boyle's law held, then, as we have seen in Sec. 19.2, the bulk modulus should equal the equilibrium pressure of the gas (usually atmospheric pressure). In that case

$$v = \sqrt{\frac{p}{\rho}} \tag{37.26}$$

If the gas is air under standard conditions, according to Eq. (37.26)

$$\begin{aligned} v &= \sqrt{\frac{1.013 \times 10^6 \text{ dynes/cm}^2}{0.001293 \text{ gm/cm}^3}} = 2.80 \times 10^4 \text{ cm/sec} \\ &= 280 \text{ m/sec} \end{aligned}$$

whereas the experimental value of the velocity of sound, at 0°C and a barometric pressure of 76 cm Hg, is 331.7 m/sec. Newton was aware of this discrepancy but was unable to account for it.

Actually the compressions and expansions involved in the passage of audible sound waves through a gas constitute an adiabatic process (Sec. 19.6), as was first pointed out by Laplace in 1816. In the compressional parts of each wave the gas is slightly heated, while in the rarefactions the expansion produces a slight cooling. But the regions which are heated and cooled are so much larger than the mean free path of the gas molecules, and the time available is so short, that heat conduction cannot effect temperature equalization. We have already shown (Sec. 19.6) that the bulk modulus of elasticity of a gas for adiabatic compressions is given by $M = \gamma p$, where γ is the ratio of the specific heat at constant pressure to the specific heat at constant volume of the gas and p is the normal pressure. We thus have for the speed of sound in a gas

$$v = \sqrt{\frac{\gamma p}{\rho}} \tag{37.27}$$

The value of γ for air is 1.403. Multiplying the value 280 m/sec calculated from Eq. (37.26) by $\sqrt{1.403} = 1.187$, we have $v = 332.0$ m/sec, in very close agreement with the experimental value. If we were to use lb/ft^2 for p and lb/ft^3 for ρ, we must remember to multiply p by 32 ft/sec^2, and v would be in ft/sec.

From the equation of state for an ideal gas we know that

$$\frac{p}{\rho} = \frac{RT}{m}$$

where m is the mass of 1 mole of the gas. Equation (37.27) may therefore be written

$$v = \sqrt{\frac{\gamma RT}{m}} \tag{37.28}$$

Since, for a given gas, γ, R, and m are constants, it is to be concluded that v is independent of the pressure, provided that the pressure is atmospheric or below so that the gas closely approximates an ideal gas. This means that quite high in our atmosphere the velocity of sound for equal temperature is the same as on the ground. And if v_1 is the velocity of sound at the absolute temperature T_1, v_2 the velocity in the same gas at T_2, then, from Eq. (37.28), $v_1 = \sqrt{\gamma RT_1/m}$ and $v_2 = \sqrt{\gamma RT_2/m}$. Dividing, we have

$$v_1 = v_2 \sqrt{\frac{T_1}{T_2}} \tag{37.29}$$

In developing Eqs. (37.24) and (37.25) for the velocity of longitudinal sound waves we have purposely mentioned the vibration back and forth of layers of the medium rather than of the individual molecules. The latter have at all times their random thermal velocities, upon which these displacements are superposed. In a given layer, of thickness Δx, that is shuttling back and forth and suffering compressions and expansions, molecules will be moving into and out of the layer because of their incessant thermal motion. But this constant replacement of molecules by others need not concern us here, for these compressions involve the slight vibration (usually much smaller than the mean free path of the molecules) of regions large as compared with the diameter of a molecule. Nevertheless there should be some connection between the velocity of compressional waves in a gas and the rms velocity of its molecules (Sec. 20.3), for the transfer of momentum between molecules must affect the elastic modulus as well as the molecular speeds between collisions. You would therefore expect that the thermal velocity of the molecules would always be somewhat greater than the velocity of compressional waves in the same gas. This is indeed true, for the molecular velocity is $\sqrt{3p/\rho}$, while the value of γ for Eq. (37.27) is never larger than 1.67 for any gas. The two velocities are, however, of the same order of magnitude.

***37.8. Water Waves.** In the case of waves on the surface of a liquid the particles execute a periodic elliptical motion. If from the impact of an object, force of the wind, etc., a portion of the liquid is either heaped above or depressed below the normal surface level, what restoring forces come into play? Obviously the force of gravity must be the principal restoring force when the waves are large. But if the curvature of the waves is large, as for ripples, surface tension should be the important force tending to restore the surface to its normal form.

It can be shown that the velocity of propagation of surface waves is

$$v = \sqrt{\frac{g\lambda}{2\pi} + \frac{2\pi\sigma}{\lambda\rho}} \tag{37.30}$$

where ρ and σ are the density and surface tension, respectively, of the liquid. When λ is large, obviously only the first term under the radical is important, whereas if λ is small the second term predominates. For wavelengths in water longer than about 10 cm, the surface-tension term may be neglected, while for wavelengths less than 3 mm the term $g\lambda/2\pi$ may be neglected. If λ lies between these two values, both terms should be included. Considering the variation of velocity with λ, we note that there must be a wavelength of an intermediate value for which the velocity of water waves is a minimum. This minimum value of v is 23 cm/sec at a λ of 1.8 cm.

If small capillary waves or ripples of known frequency are made with a vibrating stylus in the surface of a liquid of known density ρ, measurement of the wavelength in the train of ripples will give one enough information, using Eqs. (37.1) and (37.30), to compute the surface tension σ.

37.9. Interference of Waves. Beats. In Sec. 37.3 we considered how vibrations of complex form result from the superposition (interference) of sinusoidal vibrations with various amplitude, frequency, and phase relations with one other. Similarly there can be *interference* of wave trains in space, with the production of a resultant amplitude, at every point along the path of the waves, varying from zero to the arithmetical sum of the amplitudes of the component waves.

An interesting case of interference of sound waves occurs when two close sources, say two tuning forks, whose frequencies f_1 and f_2 differ by only a very small amount, are vibrating simultaneously. We detect the two emitted wave trains at a given point in space. In Fig. 37.7 we represent two such vibrations, using the device of plotting the longitudinal vibrations as transverse, starting in phase with each other at O. The resultant of these two vibrations of the same amplitude and nearly equal frequencies is seen to have an amplitude varying slowly and periodically between the values zero and $2A$, where A is the amplitude of either of the components. Now the loudness of a sound depends on the

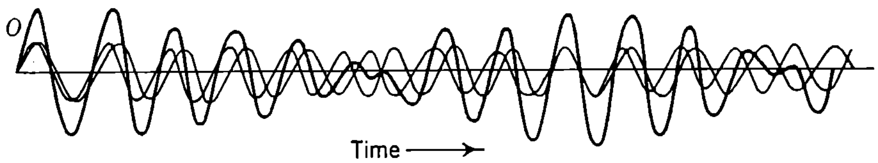


FIG. 37.7. The resultant (heavy line) of two vibrations of nearly the same frequency varies in amplitude with the *beat* frequency equal to the difference in the frequencies of the two component vibrations.

amplitude of the vibrations, and hence the resultant sound wave will fluctuate in loudness periodically. It is evident that the number of times the two vibrations will be in phase with each other per second is equal to the difference in their frequencies. There will therefore be $f_1 - f_2$ *beats* per second.

To examine this analytically, assume that the displacements η_1 and η_2 of the separate vibrations at any one point in the medium are given by the simple harmonic relations

$$\eta_1 = A \sin 2\pi f_1 t \qquad \eta_2 = A \sin 2\pi f_2 t$$

The resultant displacement is

$$\eta = \eta_1 + \eta_2 = A[\sin 2\pi f_1 t + \sin 2\pi f_2 t]$$

and since there is the trigonometric relation

$$\sin a + \sin b = 2 \sin \tfrac{1}{2}(a + b) \cos \tfrac{1}{2}(a - b)$$

we may rewrite this resultant in the form

$$\eta = \left[2A \cos 2\pi \frac{f_1 - f_2}{2} t\right] \sin 2\pi \frac{(f_1 + f_2)}{2} t \qquad (37.31)$$

According to this expression the resultant has a frequency $(f_1 + f_2)/2$, or the mean of the frequencies of the two component vibrations. The portion in the square brackets is a variable amplitude factor, the variation with time being of frequency $(f_1 - f_2)/2$. Whenever the cosine function is 1 or -1, this amplitude factor will be a maximum. These maxima, or beats, will occur then at a rate $2(f_1 - f_2)/2$ per second.

In tuning two musical instruments use is made of this phenomenon of beats, for when the beat note is eliminated, one knows that the two tones are of exactly the same frequency. Beats can be distinguished by the ear as separate pulsations up to almost 5 per second. As the difference in the frequencies of the two notes increases gradually beyond this value, the resultant sound becomes to the ear more and more of a discord, up to the point where the two frequencies bear some fairly simple ratio to one another. The combination then becomes harmonious.

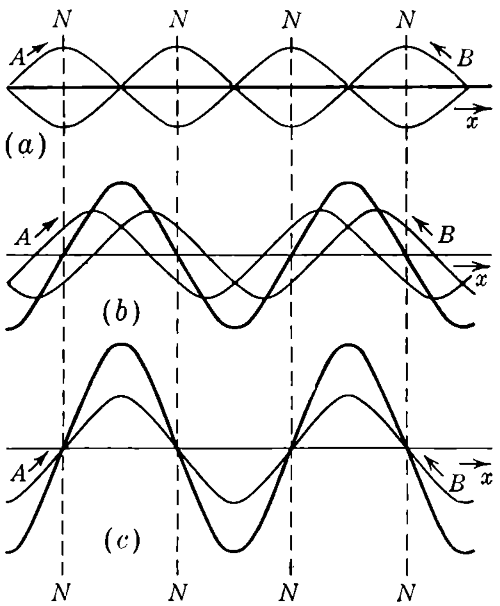


FIG. 37.8. The resultant of two similar wave trains moving in opposite directions is a stationary wave with nodes fixed at the positions N.

37.10. Stationary Waves. When two wave trains of the same period and amplitude moving in opposite directions are carried simultaneously by the particles of a medium, *stationary*, or *standing*, *waves* result. These may be of the transverse variety in a stretched string, longitudinal stationary sound waves in a fluid, or stationary surface waves. In fact this is a universal, easily produced phenomenon for all types of wave motion. Figure 37.8 depicts two such wave trains, A moving to the right, B to the left. In (*a*) the two wave trains are in phase opposition, the resultant being zero displacement throughout, as represented by the heavy horizontal line. In (*b*) of Fig. 37.8 each wave train has advanced one eighth of a wavelength, while in (*c*) each has advanced an additional eighth of a wavelength. It is to be noted that at the points marked N there is always zero displacement. These positions, known as the *nodes* of the stationary wave, are half a wavelength apart. Between these points of complete interference the particles of the medium are always in vibration; but all these vibrating particles pass through the equilibrium position at the same instant, and therefore there is no phase difference between them. The positions of maximum displacement midway between the nodes are called *antinodes*, or *loops*. Between two adjacent nodes the particles are always moving simultaneously in the same direction, while on opposite sides of a node the displacements at any one instant are always in opposite directions.

Let the displacements in wave train A be represented by

$$y_1 = A \sin 2\pi \left(\frac{t}{P} - \frac{x}{\lambda}\right)$$

Then the equation for wave train B traveling in the opposite direction is given by

$$y_2 = A \sin 2\pi \left(\frac{t}{P} + \frac{x}{\lambda}\right)$$

The resultant is,

$$y = y_1 + y_2 = A \sin 2\pi \left(\frac{t}{P} - \frac{x}{\lambda}\right) + A \sin 2\pi \left(\frac{t}{P} + \frac{x}{\lambda}\right)$$

On employing the trigonometric relation

$$\sin (a \pm b) = \sin a \cos b \pm \cos a \sin b$$

this equation for y may be put in the form

$$y = \left(2A \cos 2\pi \frac{x}{\lambda}\right) \sin 2\pi \frac{t}{P} \qquad (37.32)$$

Equation (37.32) is an analytical representation of all the facts we have mentioned about stationary waves. Notice that the sine function

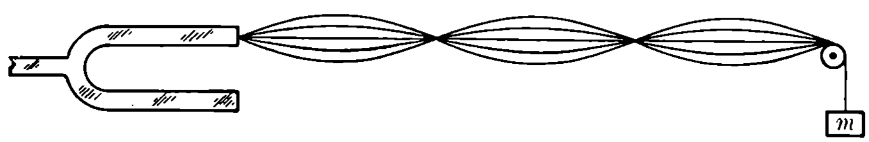


FIG. 37.9. Stationary waves in a string (Melde's experiment).

involves the time but not the distance x. The amplitude factor in parentheses is a function of x, the coordinate in the direction the waves are moving. When $x = \lambda/4$, $3\lambda/4$, $5\lambda/4$, . . . , the cosine function is zero and we have the points of zero amplitude, the nodes, spaced a half wavelength apart. For $x = \lambda/2$, λ, $3\lambda/2$, . . . the cosine function is $+1$ or -1, and hence at the points of maximum vibration, the antinodes, the amplitude is $2A$. Finally the sign of $\cos 2\pi(x/\lambda)$ changes at the nodal points.

Stationary waves in a stretched string may be studied in Melde's experiment, in which a long string under tension is attached to the prong of an electrically driven tuning fork (Fig. 37.9). The train of transverse waves having the frequency of the tuning fork is sent down the string and reflected from the far end with but little diminution in amplitude. Since the velocity of propagation of the wave depends upon the tension in the string according to Eq. (37.13), for fixed frequency f the wavelength λ may be changed by varying the mass m. For certain values of the tension F, λ will be such that standing waves are observed with the string vibrating in 1, 2, 3, . . . loops. Let us compute the tension needed to give standing waves with four loops, in a string 1 m long, weighing 0.5 gm, and hanging from a fork vibrating 200 times per second. Since $\lambda = 50$ cm, $v = 10{,}000$ cm/sec, and hence from Eq. (37.13)

$$F = (10{,}000 \text{ cm/sec})^2 \times \frac{0.5}{100} \text{ gm/cm}$$
$$= 5 \times 10^5 \text{ dynes} = 510 \text{ gm}$$

37.11. Wave Fronts and Rays. Huygens' Principle. If a train of waves is filling a medium, a surface connecting particles of the medium all of which are in the same phase of vibration constitutes a *wave front.* In the case of sound waves in air, the vibrating source being quite small, such a surface will be spherical; and if the distance from the source is large, a small portion of this spherical wave front may be considered as a plane wave. If the medium is homogeneous and isotropic, the direction of propagation is always at right angles to the wave front. A line normal to the wave fronts indicating the direction of motion of the waves is called a *ray.* In discussing various wave phenomena it is sometimes best to consider the behavior of the wave fronts, but if we are interested only in directional effects we may confine our attention to the rays.

The behavior of the wave fronts as a train of waves proceeds unimpeded through a uniform medium, bends around obstacles (*diffraction*), rebounds from a surface (*reflection*), or changes direction when entering a new medium (*refraction*) may be predicted by using a simple method first proposed by Huygens (1629–1695), a contemporary of Newton. If the position of the wave front is known at some given time, then with the aid of Huygens' principle the position of the wave front at any later time may be determined. This principle states that *every point on a wave front may be considered to be a new source of disturbance from which spherical wavelets issue. The envelope of these secondary wavelets is the new wave front at a time r/v later,* where r is the distance the wavelets go in this time at the velocity v.

This is illustrated in Fig. 37.10, where ABC is a section of a spherical wave surface that originated at O. To construct the wave front a time t later, we form arcs of radius $r = vt$ from any points on ABC. The surface $A'B'C'$ tangent to all these arcs is the required wave front. It might be thought that, if all points along the surface ABC are new sources of disturbance, another wave ought to be propagated back toward O. Detailed analysis shows, however, that Huygens' principle in this simple form gives correct answers only if we ignore its prediction of a back wave.

Diffraction, or the bending of the wave disturbance into the region behind an obstacle, is easily explained by using Huygens' principle. In Fig. 37.11 a spherical wave front from O is pictured as it strikes an aperture BC. On taking points along the wave surface between B and C as new sources it is evident that some disturbance is to be found in the regions apparently blocked by the stops defining the opening. Whether or not this bending of the waves around the sides of obstacles is observable depends upon the relative dimensions of the waves and the obstacle or opening. If the length of the waves and the width of the opening are comparable in size, the bending around the corner will be complete and there will be no part of the region beyond the opening not filled with the disturbance. This is the case for ordinary sounds passing through open windows and doors, since for sound waves in air a frequency of 264 vibrations per second, for example, has a wavelength greater than 4 ft. For such sounds a hill or a very large building would cast a quite sharp sound shadow. With visible light waves of wavelength, say, 0.0005 mm, one must use openings of very small width and observe with care in order to detect diffraction effects (Chap. 47).

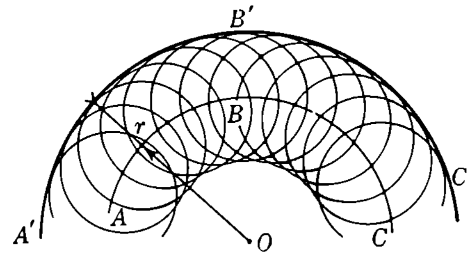


FIG. 37.10. Given the spherical wave surface ABC, any succeeding wave surface as $A'B'C'$ may be found by Huygens' construction.

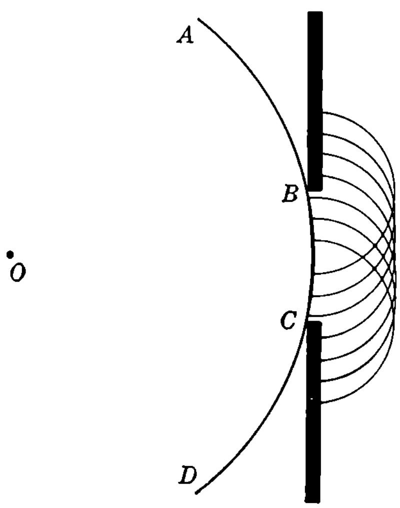


FIG. 37.11. Since every point between B and C on the wave front is a new source, some disturbance is found behind the obstacles.

37.12. Reflection of Waves. The law of reflection for a wave front incident upon a smooth obstructing surface may be obtained by application of Huygens' principle. In Fig. 37.12, AC represents a plane wave incident obliquely on the surface IJ, and with the end A of the wave just touching the surface. In the time it takes for the C end of the wave to travel the distance CE to the surface, the disturbance at A would have proceeded to H, if the surface were not there, and the new wave front would then be HE. But we may take A as being a new center of disturbance; and if we draw a circle of radius $AF = CE$, we know that the reflected wave from A will be represented by this circle at the instant when the other end of the wave arrives at E. Next consider some intermediate point B on the incident wave. The disturbance from B proceeds to D on the reflecting surface. D becomes a new center of disturbance, and drawing at this point a circle of radius DG we know that this represents the wave from this center at the instant when the C end of the wave has arrived at E. The line EF tangent to these spherical wave disturbances arising from A and D represents the reflected wave front at this instant.

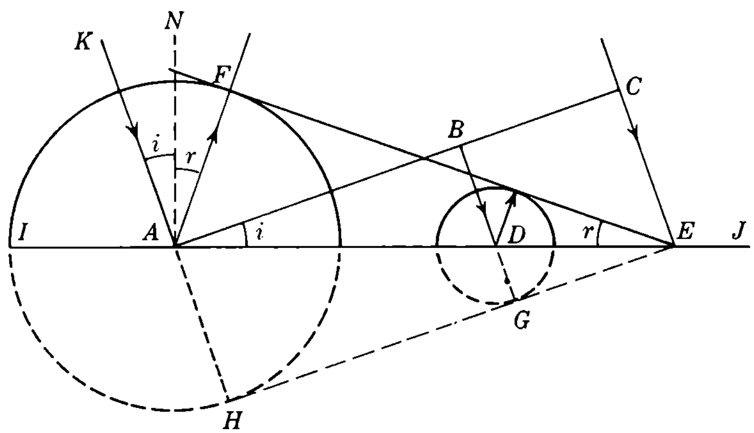


Fig. 37.12. The reflection of a plane wave.

Since the right triangles ACE and EFA are equal, the angles i and r between the incident and reflected wave fronts and the reflecting surface, respectively, are equal. But because the rays are perpendicular to the wave fronts, KA and AF also make these same equal angles with the normal AN to the surface. This proves the *law of reflection*, that the *angle of incidence* i equals the *angle of reflection* r.

The change of phase of a wave when it is reflected is a matter of some importance. In Fig. 37.13 is pictured a single transverse pulse in a rope, undergoing reflection at the fixed end of the rope. The wave form traveling up the rope involves displacements to the left. Because of the elastic reaction of the fixed end of the rope the displacement there becomes reversed, and the reflected pulse therefore starts down, with complete reversal of phase as indicated. To demonstrate this analytically, we represent the displacements in the up wave pulse by

$$y_1 = A \sin 2\pi[ft - (x/\lambda) + \phi_1]$$

those in the down pulse by

$$y_2 = A \sin 2\pi[ft + (x/\lambda) + \phi_2]$$

where ϕ_1 and ϕ_2 are the necessary phase constants. Taking $x = 0$ at the fixed end of the rope, and since the resultant displacement $y = y_1 + y_2$ must be zero at this point, we have

$$\sin 2\pi(ft + \phi_1) = -\sin 2\pi(ft + \phi_2)$$

which can be true only if $\phi_2 - \phi_1 = \pi$. When the wave traveling down the rope arrives at its free end, however, this end, being under no restraint, will move out with increased displacement. This larger amplitude of the free end, in the same sense as that in the incident wave, starts a reflected wave of the *same* phase back up the rope, and if the analysis above were carried out for reflection at the free end it would show that $\phi_1 = \phi_2$.

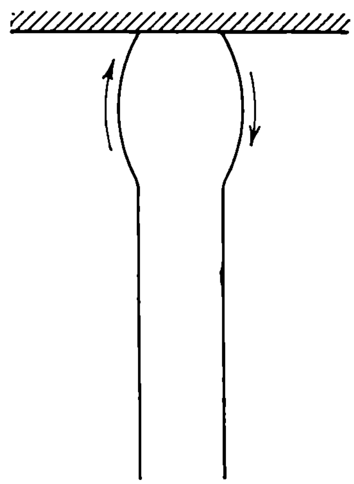


FIG. 37.13. A wave pulse in the rope is reflected at the fixed end with a change of phase of 180°.

Similar phase effects occur when compressional waves in a fluid in a tube are reflected from the end of the tube. Suppose that a compression moving down the tube from left to right is reflected from the closed end. On recalling that in a compression the layers of particles of the fluid are moving in the same direction as the wave, the situation is quite analogous to the impact of a small steel sphere on a much larger mass of steel. Just as the ball will rebound, so there will be a reversal of the direction of motion in the compressed layer at the closed end, and the compression starts back moving from right to left (Fig. 37.14*a*). If the compression arrives at the open end of a tube (Fig. 37.14*b*), however, that compressed layer of the fluid continues to move forward from left to right because of the lack of constraint and the expansion causes a diminution of the pressure at that point. A rarefaction, with the layers of particles still moving from left to right, therefore moves back down the tube from right to left. At first thought this reflection of a compression from an open end as a rarefaction might seem to represent a change of phase of 180°. But as compared with a reflection from a fixed end there is *a delay of half a period* in the reflection of this wave of opposite phase. Hence, in the reflection of a compressional wave from the fixed end of the tube, a denser medium, there is a reversal of phase, or change of phase, of 180°.

In the reflection from the open end, or a less dense medium, however, there is zero change of phase, for the direction of displacement of the layers of the fluid does not change.

In a similar manner we could show that a rarefaction should be reflected from a denser medium as a rarefaction but from a less dense medium as a compression. Also it should be pointed out that these are but examples of a general rule holding for all types of wave motion: *Reflection from a denser medium involves a phase change of π radians, while reflection from a less dense medium involves no change in phase.* We shall have occasion to refer to these changes for sound waves in the following chapter and for light waves in Chap. 46.

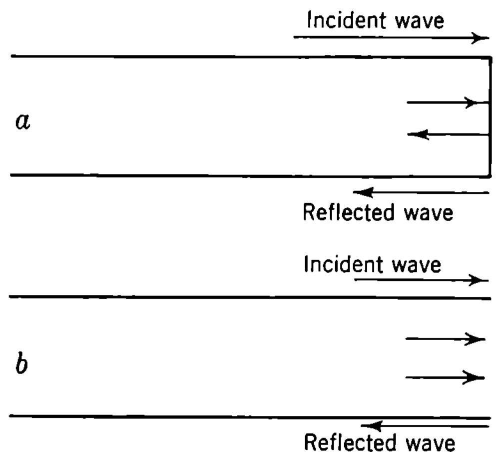


FIG. 37.14. A longitudinal wave in a fluid is reflected from the fixed end of the tube with reversal of phase, but with no phase change at an open end.

37.13. Refraction of Waves. When a wave front passes obliquely from one medium into another in which it travels with a different velocity, its direction will change. This is the phenomenon of *refraction*. In Fig. 37.15 suppose AC to represent a section of a plane wave front incident at the angle i on the boundary between two media. Let the velocity of propagation be v_1 in the upper

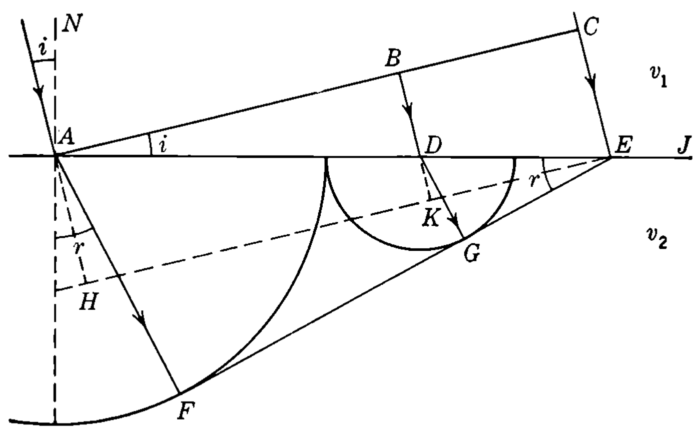


FIG. 37.15. The refraction of a plane wave upon passing obliquely with increase of velocity into a different medium.

medium and v_2 in the lower medium, with $v_2 > v_1$. This would be the case for sound waves if the lower medium is a liquid or solid, with air above it.

At the instant pictured, the point A of the wave has just reached the surface J. Now the wave will usually in part be reflected back into medium 1, but we assume that in part it passes on into medium 2. By Huygens' principle, A becomes a new source of disturbance; and in the time it takes for the wave to travel the distance CE in medium 1, the disturbance emanating from A will travel a distance $(v_2/v_1)CE = AF$ in medium 2. We draw a circle of radius AF centered at A. In the same time interval the disturbance from the point B on the wave AC will go to D on the interface between the two media and then a distance $DG = (v_2/v_1)DK$ in medium 2. We draw an arc of this radius from D as the center. The line EGF from E tangent to these two arcs represents the refracted wave at the end of this time interval. The angle r that this refracted wave front makes with the surface is known as the *angle of refraction.*

The fact that the ratio $\sin i/\sin r$ is a constant for any two media is known as Snell's law.[1] To prove this relation, we note, referring to Fig. 37.15, that

$$\frac{\sin i}{\sin r} = \frac{CE}{AF} = \frac{CE}{(v_2/v_1)CE} = \frac{v_1}{v_2} \qquad (37.33)$$

For light waves the velocity v_2 in the solid or liquid medium will be less than v_1, with the result that, for light waves passing from air into a solid or liquid, $\sin r < \sin i$.

PROBLEMS

1. Show that Eq. (37.8) may be written in the alternative forms

$$y = A \sin 2\pi f[t - (x/v)],\quad y = A \sin (2\pi/\lambda)(vt - x),\quad y = A \sin [2\pi ft - (2\pi x/\lambda)].$$

2. Plot the resultant wave form for three component sine waves of the same period but with amplitudes in the ratio $1:\frac{1}{2}:\frac{1}{3}$ and phases of 0, $\pi/2$ and π. Comment on the nature of the resultant.

3. Four superposed sinusoidal vibrations have frequencies in the ratio 1:2:3:4 and amplitudes in the ratio $1:\frac{1}{2}:\frac{1}{3}:\frac{1}{4}$, but the even- and odd-numbered components have opposite signs. Plot the resultant vibration. Note the near approach to a "saw-tooth" form.

4. If the equation for a transverse wave in a stretched string is $y = 2 \sin 2\pi [(t/0.02) - (x/20)]$, with lengths expressed in centimeters and time in seconds, compute the amplitude, wavelength, frequency, and velocity of the wave. Also, calculate the maximum transverse velocity and acceleration of a particle of the string.

5. A simple harmonic wave moving in the positive x direction has an amplitude of 3 cm, frequency of 100 vibrations per second, and velocity of propagation of 30 m/sec. Calculate the displacement y at $x = 140$ cm from the origin at $t = 3$ sec, the wavelength, the particle velocity $\partial y/\partial t$, and the particle acceleration $\partial^2 y/\partial t^2$.

[1] Snell (1591–1626) actually did not state the law as the ratio of the sines of the two angles. The law was first expressed in this form by Descartes (1596–1650).

6. A steel wire 1 m long and of mass 5 gm is stretched with a tension of 10 kg. What is the velocity of a transverse wave in this wire?

7. If a string 60 ft long weighs 0.04 lb, with what tensile force in pounds must it be stretched in order that a transverse wave would travel its length in 0.1 sec?

8. The string mentioned in Prob. 4 has a linear density of 0.01 gm. Compute the total energy per centimeter of that transverse wave.

9. Calculate the velocity of sound in hydrogen and in air at 20°C. By how much should the velocity of sound in air increase for each degree rise in temperature?

10. In echo sounding to determine the depth of the water, compressional wave pulses are sent to the bottom, and the elapsed time for the reflected waves to return to the ship is determined. If this time is measured as 1.35 sec, and assuming that the bulk modulus for sea water is the same as that for pure water and that the average water temperature is about 4°C, calculate the depth of the ocean.

11. If the velocity of a sound wave in a gas is 400 m/sec at 17°C, what will be the velocity at 47°C and double the pressure?

12. Compute the tension to give standing waves with three loops in a string 90 cm long, weighing 0.4 gm, and hanging from a fork vibrating 150 times per second.

13. A copper wire of 1 mm^2 cross section is under the tension produced by a 1-kg weight. Calculate (*a*) the velocity of a transverse wave, (*b*) the velocity of a longitudinal wave in this wire.

14. A plane sound wave in air strikes a water surface at an incident angle of 10°. What angle does the refracted wave in the water make with the surface? Assume the temperature of the air to be 0°C and that of the water 4°C. Will the sound be refracted into the water for every angle of incidence?

15. The density of helium gas under standard conditions is 0.000178 gm/cm^3, and the ratio of its specific heats is 1.6. Calculate the velocity of sound in helium at a pressure of 74 cm of mercury and a temperature of 27°C. By how much would this velocity change if the pressure became 75 cm of mercury but the temperature remained constant?

***16.** Show that the minimum value of v and the corresponding value of λ for water waves as given in Sec. 37.8 are correct.

CHAPTER 38

PRODUCTION AND RECEPTION OF SOUND

38.1. The Nature of Sound. In the preceding chapter on Wave Motion considerable attention was paid to sound waves. We emphasized that sound originates in vibrating bodies and that the vibrations are transmitted through the surrounding elastic medium, usually air, as a wave motion of the longitudinal type. When the compressional sound waves are passing through air, the amplitude of the vibratory motion of the layers or particles is surprisingly small, being only about 10^{-8} cm for a sound that is barely audible. Since a fluid offers no elastic resistance to a shearing force, *i.e.*, its shear modulus of elasticity is zero, it cannot transmit transverse waves.

The velocity of sound in air, about 342 m/sec according to Eq. (37.27), is not large compared with velocities familiar today and is insignificant in comparison with the velocity of light (3×10^8 m/sec). Rifle bullets travel with a muzzle speed of two or more times the speed of sound, and we are soon to have stratosphere supersonic airplanes flying at speeds of 700 m/sec, or about double the speed of sound. V-2 rockets attain a speed of 1,000 m/sec. The sound from a distant, swift airplane moving across your line of sight always comes from a point in the sky well behind the airplane. Measurement of many acoustic phenomena such as reverberations, echoes, and resonances involves the velocity with which the sound waves are propagated. This velocity may be easily measured in the laboratory when propagation occurs in any solid, liquid, or gas by a measurement of stationary waves in the material, using a source of known frequency.

Sound waves, of course, exhibit the properties of reflection, refraction, and diffraction characteristic of all forms of wave motion. Such behavior of sound waves may be rendered visible by the method of spark photography. The waves passing before the photographic plate are illuminated for a brief interval by a timed electric spark. The slight increase in density in the condensation in the wave front causes some refraction of the light and thus produces a shadow in the photograph. Figure 38.1 is such a photograph, taken by Foley and showing the reflection of a spherical sound wave from a plane surface. Figure 38.2 is a photograph taken by this same technique and demonstrating the refraction of a sound wave in passing through a lens of carbon dioxide gas, as well as the reflection from the face of the lens and the diffraction at its edge.

The diffraction of sound waves has already been mentioned in Sec.

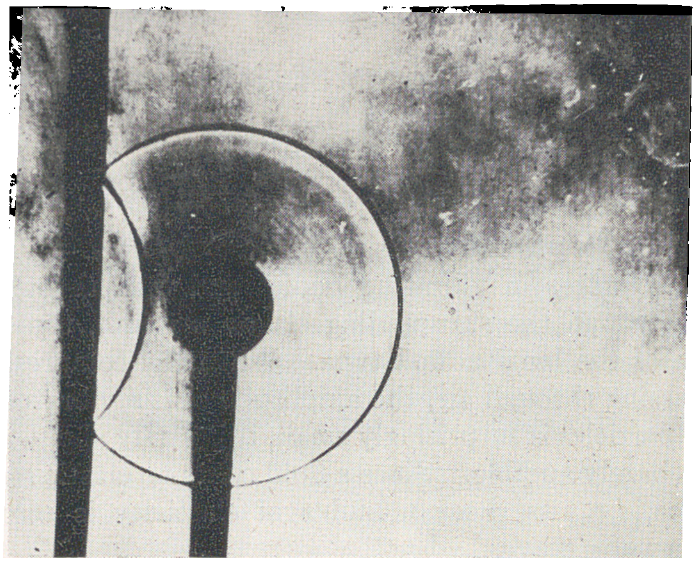


FIG. 38.1. Reflection of spherical sound wave from a plane surface. (*After Foley.*)

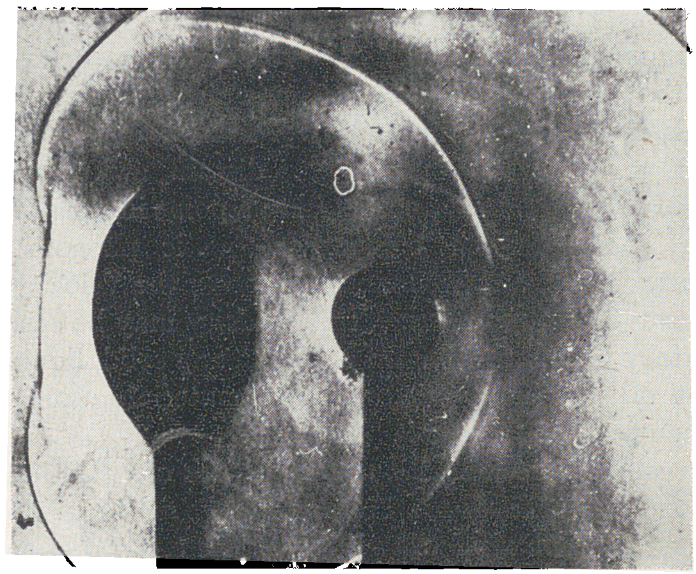


FIG. 38.2. Sound wave through lens of carbon dioxide gas, showing the original wave (upper right corner of photograph) and the reflected, refracted, and diffracted waves. (*After Foley.*)

37.11. The bending of sound waves around corners is the rule, since the wavelength is about as great as the size of the opening or the obstruction. Audible sounds range in wavelength from 17 m (20 vibrations per second) to 1.7 cm (20,000 vibrations per second). For the shortest wavelengths, as well as for supersonic waves, one would expect from Huygens' construction the sharp sound shadows that are observed.

38.2. Musical Sounds. The difference between a noise and a musical sound, or *tone*, is that in the former a continuous range of frequencies is present, whereas a musical tone is composed of one or more discrete frequencies. Sometimes, however, the distinction between a noise and a musical sound is not sharp. If you drop a stick of wood on the table, for example, you would call the resulting sound a noise. But if you drop onto the table in quick succession four suitably chosen sticks of wood, you may get the notes of a musical chord.

Musical tones are characterized by three properties, *pitch*, *loudness*, and *quality*. The *pitch* of a tone depends upon the number of vibrations per second—the greater the frequency, the higher the pitch. *Loudness* is associated with the intensity of the sound, which is proportional to the power expenditure in the vibrating source. The quality of a musical sound is directly related to the shape of the wave form, which, in turn, is affected by the number and intensity of higher frequencies or *overtones* accompanying the fundamental tone.

38.3. Pitch of a Sound. Doppler Effect. That the pitch of a tone rises as the number of vibrations per second in the source increases is a matter of common observation. When an electric motor picks up speed, the sound produced by the increasing vibration rate of the motor components rises in pitch. As a rotary power saw cuts through a board, the number of impulses given to the air, and hence the frequency and pitch of the resulting sound, varies directly with the speed of the saw. And the connection between frequency and pitch may be shown directly with a siren, which consists essentially of a variable-speed disk having regularly spaced holes around its periphery. As the disk rotates, an air blast plays against this row of holes, so that a succession of compression pulses is imparted to the surrounding air. The pitch of this sound rises steadily as the speed of the disk is raised. Incidentally, if the holes in the siren disk are not uniformly spaced, the irregular train of condensations in the surrounding air constitutes a noise, not a musical tone.

As already noted, the human ear can detect tones of frequency as low as 20 per second and as high as 20,000 per second, although this upper limit is lowered by age and by diseases of the ear. Supersonic vibrations as high as 500 Mc/sec or 5×10^8 cps can be produced by oscillating quartz crystals in suitable electronic circuits.

If there is a relative motion between the sound source and the observer along the line joining them, the pitch of the sound as heard by the observer is changed. This is known as the *Doppler effect*, first elucidated by C. Doppler in 1842 for the analogous optical case. If the source is in motion toward the observer, the pitch of the tone heard is higher than when the source is stationary; but if the source is receding, the observer hears a sound of lower pitch than when the source is at rest. Similar changes in the pitch heard occur if the observer is in motion toward or away from the source. Nowadays, with our high-speed trains, automobiles, and airplanes, this phenomenon is very common.

a. Source in Motion. Let the frequency of the source be f, its velocity *toward* the observer be v', and the velocity of propagation of the waves be as usual denoted by v. Since the f waves produced in each second are crowded into the space $v - v'$ (Fig. 38.3*a*), the wave length of the sound

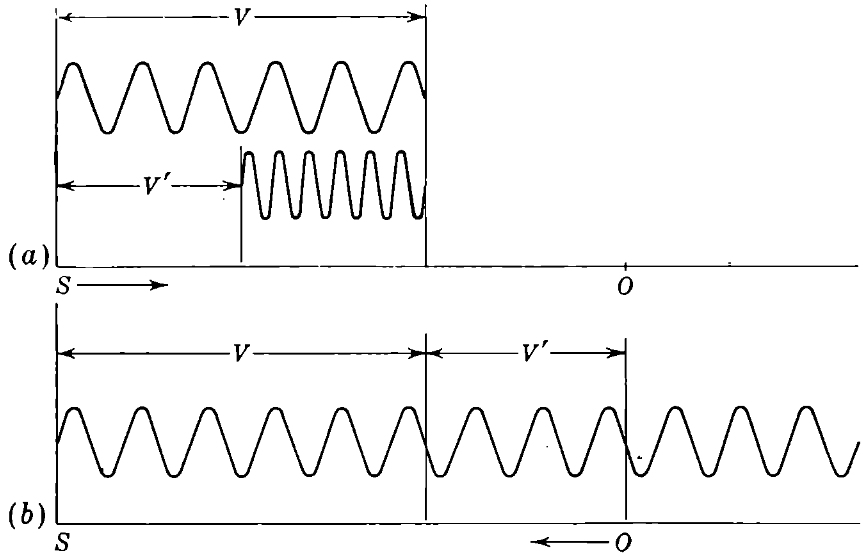


FIG. 38.3. Doppler effect. (*a*) f waves crowded into space $v - v'$. (*b*) Observer receives waves in space $v + v'$ each second.

will be shortened from $\lambda = v/f$ to $\lambda' = (v - v')/f$. The pitch $f' = v/\lambda'$ of the sound heard is, then,

$$f' = f\frac{v}{v - v'} \qquad (38.1)$$

If the *source is receding*, the sign of v' should be reversed, with the result that

$$f' = f\frac{v}{v + v'} \qquad (38.1')$$

b. Observer in Motion. When the *observer* is moving with velocity v' *toward* a stationary source (Fig. 38.3*b*), there is no crowding of waves but the ear receives each second, not only the f waves in the interval v, but also an additional number v'/λ in the distance v'. Therefore, the increase in frequency heard is $f(v'/v)$, and the pitch f' of the sound heard is

$$f' = f\left(1 + \frac{v'}{v}\right) \qquad (38.2)$$

If the *observer* is moving with velocity v' *away from* a stationary source, the sign of v' should be reversed.

$$f' = f\left(1 - \frac{v'}{v}\right) \tag{38.2'}$$

Equation (38.1) may be expanded as follows:

$$f\frac{v}{v - v'} = f\left(1 - \frac{v'}{v}\right)^{-1} \cong f\left(1 + \frac{v'}{v} + \cdots\right)$$

if the binomial theorem is used. When v'/v is small the terms indicated by dots may be neglected, and the result is identical with Eq. (38.2). But as v' approaches v in value, by Eq. (38.1), $f' \to \infty$ while, from Eq. (38.1′), $f' \to f/2$. If the observer is in motion, with v' nearing the velocity v, by Eq. (38.2) $f' \to 2f$; but when Eq. (38.2′) applies, $f' \to 0$. And jet airplanes are already flying at speeds greater than v!

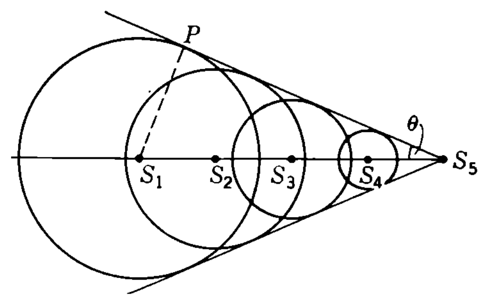


FIG. 38.4. A source moving with a velocity greater than the velocity of sound creates a wave front making an acute angle with direction of motion.

In Fig. 38.4 suppose S_1, S_2, S_3, S_4, S_5 to be the positions in successive time intervals of a body moving with a velocity v' *greater* than the velocity of sound v. When the body arrives at S_5, the spherical wave fronts of the sound that originated at the other positions as the body passed those points are as shown. That is, the sound wave from S_1 will have traveled a distance S_1P while the body has moved the distance S_1S_5. By Huygens' principle these waves will reinforce each other along the tangent surface represented by PS_5 but will interfere at all other points. The angle θ made by this wave front with the line of motion of the body is given by

$$\sin \theta = \frac{S_1P}{S_1S_5} = \frac{v}{v'} \tag{38.3}$$

A familiar example of such a wave front is the bow wave formed by a ship moving through quiet water. All shock waves are of this type.

Worked Example. Two trains, each moving at 60 miles/hr, are receding from each other. One of the locomotives whistles, the frequency of its sound being 200 per second. What is the frequency of this sound as heard by an observer on the rear platform of the other train? On assuming v = 1,100 ft/sec, calculating first f' from Eq. (38.1), and then using this as the f in Eq. (38.2′), the result is

$$f' = 200 \times \frac{1{,}100 - 88}{1{,}100 + 88} = 170 \text{ vibrations per sec}$$

38.4. Sound Intensity. The *intensity I of a sound* is defined as *the average time rate of transfer of energy per unit area* of a surface perpendicular to the direction of propagation. Let us assume that we are dealing with a train of plane compressional waves of the simple harmonic variety, traveling in the x direction. In Sec. 12.9 it was shown that the energy of an oscillating particle is $2\pi^2 f^2 A^2 m$. The energy W per unit volume is n times as great if the number of oscillators per unit volume is n. But $mn = \rho$, and hence

$$W = \rho \times 2\pi^2 f^2 A^2 \qquad (38.4)$$

where ρ is the density of the fluid. Therefore, since the energy flow takes place with the velocity v with which the disturbance is propagated, the energy passing per second through each unit area of a surface perpendicular to the x direction, or the intensity I of the sound, is

$$I = \rho\, 2\pi^2 f^2 A^2 v \qquad (38.5)$$

According to this result the intensity of a sound varies directly as the square of both the amplitude and the frequency of the vibration.

As the waves proceed through the fluid medium there will always be a certain amount of *damping*, or loss of energy from the waves to the fluid. Damping may be explained as due to viscosity and to heat conduction from the condensations to the rarefactions. Also, at certain rather high frequencies there may be marked absorption of the sound energy into the internal motions of the molecules.

How does the intensity of a sound vary with the excess pressure associated with the waves? To answer this question we use Eq. (37.24) and employ the value of the modulus M derived in Sec. 19.2,

$$v = \sqrt{\frac{M}{\rho}} = \sqrt{-\frac{p}{(\Delta V/V)\rho}}$$

where V denotes the original volume and ΔV the change in volume produced by the excess pressure p. Now the change in volume that occurs in the fluid in the transmission of these compressional waves arises from the displacement η of the fluid layers from their equilibrium positions. Therefore the strain $\Delta V/V$ is solely longitudinal and equal to $\Delta\eta/\Delta x$, or, in the limit, to $\partial\eta/\partial x$. Substituting $\partial\eta/\partial x$ for $\Delta V/V$ in the equation for v and solving for p, we have

$$p = -v^2\rho\partial\eta/\partial x \tag{38.6}$$

For these simple harmonic waves,

$$\eta = A \sin 2\pi f\left(t - \frac{x}{v}\right)$$

Hence
$$\frac{\partial\eta}{\partial x} = -\frac{2\pi Af}{v} \cos 2\pi f\left(t - \frac{x}{v}\right)$$

Equation (38.6) may therefore be written

$$p = 2\pi Avf\rho \cos 2\pi f\left(t - \frac{x}{v}\right) \tag{38.7}$$

We thus see that the maximum excess pressure p_{max} in the sound waves is given by

$$p_{max} = 2\pi vf\rho A = \frac{2\pi\rho v^2}{\lambda} A \tag{38.8}$$

Equation (38.7) may then be written

$$p = p_{max} \cos 2\pi f\left(t - \frac{x}{v}\right) \tag{38.9}$$

which may be considered to be the equation of a *pressure wave*. The factor p_{max} is often referred to as the *pressure amplitude*. Note that this pressure wave is 90° out of phase with our original displacement wave, for it is a cosine function, whereas we represented the displacement wave by a sine function. Combining Eqs. (38.5) and (38.8) we obtain

$$I = \frac{p_{max}^2}{2\rho v} \tag{38.10}$$

which shows that the intensity varies as the square of the excess pressure. Since the frequency does not occur in this formula, it is evident that the intensities of sounds of different frequencies may be compared by just measuring the associated excess pressures. The acoustical engineer therefore prefers to measure pressure variations rather than amplitudes.

The values of p_{max} range all the way from about 2×10^{-4} dynes/cm² for the faintest sound that can be detected by a good ear[1] up to about 200 dynes/cm² for a sound so loud that it begins to be painful. For ordinary

[1] This 2×10^{-4} dyne/cm² value is the American Standards Association (ASA) reference level for pressure amplitude.

conversational tones, p_{max} may be about 1 dyne/cm². For sound waves in air at 20°C at normal pressure of 76 cm Hg and for $\lambda = 1$ m, we may calculate from Eq. (38.8) the amplitude A of the vibrations of the particles of the medium.

$$A = \frac{1 \text{ dyne/cm}^2 \times 100 \text{ cm}}{2\pi \times 0.00121 \text{ gm/cm}^3 \times (3.44 \times 10^4 \text{ cm/sec})^2} = 1.12 \times 10^{-5} \text{ cm}$$

This verifies our statement in Chap. 37 that the vibration amplitudes in sound waves in a gas are very small.

The intensity of these same sound waves is

$$I = \frac{(1 \text{ dyne/cm}^2)^2}{2 \times 0.00121 \text{ gm/cm}^3 \times 3.44 \times 10^4 \text{ cm/sec}} = 1.2 \times 10^{-2} \text{ erg sec}^{-1} \text{ cm}^{-2} = 1.2 \times 10^{-9} \text{ watt/cm}^2$$

This is a small amount of power per square centimeter of the wave front, but the human ear can detect sounds of very much lower intensity than this. If we use the value $p_{max} = 2 \times 10^{-4}$ dyne/cm² for the lower limit of audibility, the corresponding value of I is 5×10^{-17} watt/cm², which shows that the ear is an extremely sensitive organ. For comparison, we note that a mosquito in flight develops a power of about 10^{-4} watt! The ear is in fact almost sensitive enough to hear the noise produced by the thermal motions of molecules.

Finally we should mention that the intensity of a sound decreases with distance r from the source because the total energy radiated per second is passing through an ever larger surface as r grows. If the sound source may be assumed to be a point emitting uniformly in all directions, the total energy in these expanding *spherical waves* passing per second through any spherical surface concentric about the source is constant (neglecting damping). Since the surface area of a sphere is $4\pi r^2$ the energy transferred through unit area thus varies as $1/r^2$. It follows from Eq. (38.5) that the amplitude of the waves decreases as $1/r$. If therefore we denote the amplitude of the waves by A'/r, then the equation for a *harmonic spherical wave* is

$$\eta = \frac{A'}{r} \sin 2\pi f\left(t - \frac{r}{v}\right)$$

38.5. Quality of a Sound. A *pure tone* has a wave form of the simple harmonic variety. Such a tone is produced by a vibrating tuning fork, but the sounds from most sources have complex wave shapes. As already discussed in Sec. 37.3, complex periodic vibrations result from the superposition of a number of simple harmonic vibrations having frequencies always some integral multiple of a lowest *fundamental* frequency. Any higher frequency components accompanying the fundamental of a musical sound are known as *overtones*. They are often, but not always, *harmonics* of the fundamental tone. As was first shown by the great scientist H. von Helmholtz in 1859, the *quality of any tone depends upon the number, the intensities, and the frequencies of all the overtones present with the fundamental.*

Since the phase relations between the component frequencies do affect the contour of the complex wave, it might be thought that such phase shifts might change the quality of the tone. The investigations of Helmholtz and others more recent seem to show, however, that the relative phases of all the component frequencies have little to do with the quality of the resultant tone. It must be remembered that the sensations of pitch, loudness, and quality are psychological quantities, all interrelated, and not subject to very precise measurement. That is not to say that the physical quantities frequency, intensity, and wave form cannot be measured exactly, for they can be. Sound has the two aspects; subjectively we refer to the auditory sensation in the brain, while objectively we refer to the physical features of the wave motion.

Many interesting experiments on the judgments of pitch and quality of musical tones when certain component frequencies are eliminated by electric filters employed in the telephone system have been reported by Harvey Fletcher of the Bell Telephone Laboratories.[1] In general the elimination of the fundamental frequency has no effect on the pitch of a musical sound and changes the quality but little. If, however, higher overtones are eliminated, the musical quality of the tone is definitely affected. Even when groups of several components either of low or of high frequencies are eliminated the pitch does not change, although the quality may be markedly changed. When the note G (396 vibrations per second) was sounded with a violin and then all frequencies from 1,000 up were filtered out, the listeners reported that the violin quality was gone from the tone. Similarly, when an organ pipe sounded C (264 vibrations per second) and then all component frequencies above 750 per second were eliminated, the hearers said that the sound was "dull."

What makes a vowel "ee" recognizable as a particular sound, no matter whether uttered by a deep bass voice of pitch about 90 cycles/sec or by a high-pitched soprano voice of 250 cycles/sec? The harmonic analysis of this vowel sound spoken at various pitches shows always a group of high-frequency components clustering around 2,300 per second as well

[1] H. Fletcher, "Speech and Hearing," Van Nostrand, 1929.

as a component at about 375 cycles/sec (Fig. 38.5). If the high-frequency group of overtone frequencies were filtered out, there would be a distinct change in the quality of this vowel sound. The "ah" in *father* has the components around 900 cycles increased in intensity. These form the ninth, tenth, and eleventh harmonics for the bass voice, but form the third and fourth harmonics for the soprano voice speaking the vowel "ah." That is, whenever this vowel is spoken, the oral cavity takes such a shape that it resonates to a frequency of about 900, thus reinforcing that part of the total sound coming from the vocal cords.

Since of all the vowel sounds and the tones of various musical instruments each has particular overtone frequencies or a characteristic

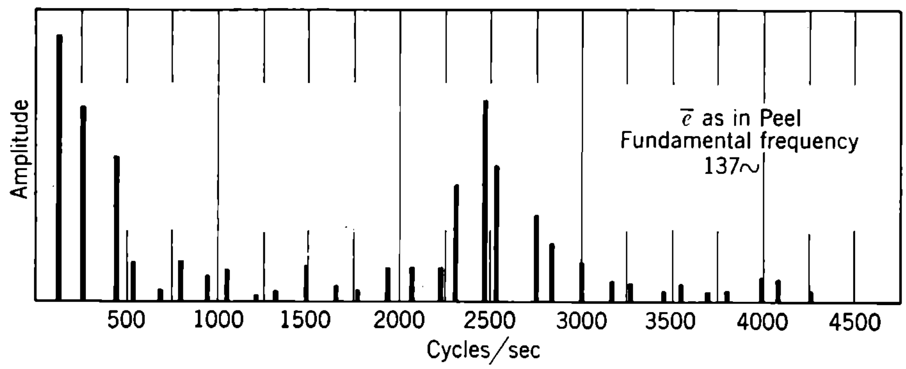


FIG. 38.5. Sound spectrum of vowel "ee" at pitch of 137 cps (*after Harvey Fletcher*). The components around 375 and 2,300 cps are characteristic of this vowel sound, no matter what the fundamental pitch.

"spectrum" of component frequencies, it ought to be possible to imitate any of these sounds by proper mixing of a set of pure tones covering the necessary frequency range. This indeed has been done, first by D. C. Miller, using a set of organ pipes, and more recently by Fletcher, at the Bell Telephone Laboratories, with a group of electronic circuits that enable him to mix in any proportions 100 pure tones spaced over the principal audible range.

38.6. The Vibration of Strings. As sources of musical sounds the stringed instruments such as the violin and piano form an important group. When a string that is under tension is struck or bowed or plucked, the disturbances travel to the two fixed ends, where reflection occurs with change of phase and with but little diminution in amplitude. At each end, then, the incident and reflected waves, being of the same frequency, tend to produce stationary waves with nodes spaced at half-wavelength intervals (Sec. 37.10). The reflected waves are re-reflected from the opposite end, of course; and in order that the two systems of nodes coincide the length of the string must be an integral multiple of half wavelengths. Only those frequencies satisfying this condition persist and produce standing waves in a vibrating string. These are the so-called *free vibrations* of a string, as opposed to the *forced vibrations* that might be impressed upon it by some applied periodic force.

When a string stretched between two supports is lightly bowed in the middle, it will vibrate in the simplest manner, in one segment with a node at each end (Fig. 38.6*a*). The note produced is a pure tone, the *fundamental tone* of the string. Its wavelength is $2l$, where l is the length of the string. The string merely vibrates as a whole between the two extreme positions, and twice in each complete vibration it will have its normal, straight-line appearance.

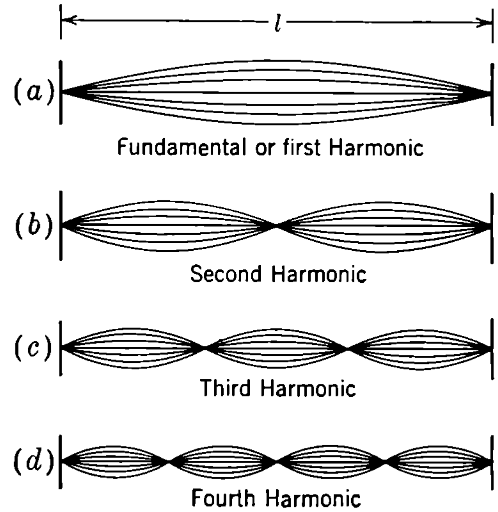


FIG. 38.6. Standing waves in a stretched string. These forms may all be present simultaneously.

A string may also vibrate in 2, 3, 4, etc., segments as indicated in Figs. 38.6*b* to *d*. Since the distance between adjacent nodes is always $\lambda/2$, the wavelengths are $2l/2$, $2l/3$, $2l/4$, respectively, for these three overtone vibrations. Employing the general relation $v = \lambda f$, we have for these four lowest frequencies of the string

$$f = \frac{v}{2l}, \frac{2v}{2l}, \frac{3v}{2l}, \frac{4v}{2l}$$

Thus, these overtones are exact multiples, or *harmonics*, of the fundamental frequency. It is customary to speak of the fundamental as the *first harmonic*, or *first partial;* the first overtone as the second harmonic, or second partial, etc.

Since the velocity v of these waves is $\sqrt{F/d}$ (Sec. 37.4), where F is the tension in the string and d is its linear density, we have

$$f = \frac{N}{2l}\sqrt{\frac{F}{d}} \tag{38.11}$$

For a given string under fixed tension F, this equation gives the frequencies of the successive harmonics if N has the values 1, 2, 3, 4, ... By adjusting the tension a string is tuned to produce exactly the correct fundamental pitch. The violinist increases the fundamental frequency of a string by effectively shortening its length with his finger. And since the fundamental varies inversely as the square root of the linear density, the strings for the low-pitched notes are the heaviest, while, for the high-pitched notes, fine, lightweight strings are used.

It is of interest to see how the standing-wave patterns for vibrating strings may be obtained by solving the wave equation [Eq. (37.20)]. If the string is to execute simple harmonic vibrations, every particle of the string must oscillate with the same frequency. Also, the displacement y at any point along the string must be represented by the product of a function such as $\sin(\omega t + \alpha)$ by a function of x alone, say $Y(x)$. Thus,

$$y = Y(x) \sin(\omega t + \alpha)$$

Therefore
$$\frac{\partial^2 y}{\partial t^2} = -Y(x)\omega^2 \sin(\omega t + \alpha) = -\omega^2 y$$

and
$$\frac{\partial^2 y}{\partial x^2} = \frac{d^2 Y}{dx^2} \sin(\omega t + \alpha)$$

Substituting in Eq. (37.17),

$$-\omega^2 Y = v^2 \frac{d^2 Y}{dx^2} \tag{38.12}$$

for which a solution is (cf. Sec. 12.2)

$$Y = A \sin\left(\frac{\omega x}{v} + \beta\right) \tag{38.13}$$

Now the two ends $x = 0$ and $x = l$ are always nodes. Therefore $Y(0) = 0$, so that $\beta = 0$. Also $Y(l) = 0$, so that, from Eq. (38.13), $\sin(\omega l/v) = 0$, or $(\omega/v)l = (2\pi f/v)l = N\pi$, where $N = 0, 1, 2, \ldots$. Hence,

$$f = (N/2l)v = (N/2l)\sqrt{F/d}$$

which is Eq. (38.11).

In any stringed instrument the sound is very largely emitted by the body or frame, which vibrates in resonance with the string. A sustained tone is produced by bowing a string, the bow alternately gripping and slipping on the string. The pleasing tones from these instruments have complex wave shapes, *i.e.*, they are rich in harmonics. Therefore a string must be capable of vibrating in a number of modes simultaneously. The waves of various lengths pass through one another, each setting up its own pattern of nodes and antinodes. Just which overtones are present depends largely on where the string is bowed or struck, for obviously this point cannot be a node. Usually the striking or bowing is done at about one-eighth the length from one end, thus eliminating the seventh and ninth harmonics. Since the basis of harmony is the simplicity of the ratios of the frequencies that are sounded together (cf. Sec. 38.11), these particular frequencies do not give harmonious combinations.

***38.7. The Vibration of Air Columns.** Another important class of musical instruments, *wind instruments*, make use of the free vibrations of columns of air in tubes with rigid walls. Examples of such instruments are the pipe organ, horn, clarinet, and flute. Some of these, *e.g.*, the clarinet, have vibrating reeds. With others, such as the cornet, the lips of the performer generate the vibrations. In all of them the tube, whose effective length is usually controlled by opening or closing holes placed at proper intervals, acts as a *resonator* to reinforce, sustain, and control these vibrations. And since in a vibrating air column there is already a considerable mass of air in motion, no additional resonating body is needed to increase the loudness of the tones. The overtones are harmonics, which is an essential characteristic of vibrating systems if they are to be useful as musical instruments. The *percussion instruments*, such as the drum, the bell, and the cymbal, are exceptions to this statement; incidentally, it may be noted that for most vibrating bodies the normal modes of vibration are indeed inharmonic.

As the prototype of all wind instruments, consider a jet of air blowing gently against the edge of the open end of a pipe, as in Fig. 38.7. This will raise the pressure slightly at this end, and a pulse of condensation, with the air particles moving in the direction of the wave, starts down the pipe. At the closed end this condensation will be reflected as a condensation (phase change of 180°). When this reflected condensation or pressure wave returns to the open end of the pipe, it will throw the air jet away from the edge, if, indeed, it has not already been forced out by the increased pressure in this region. This expansion of the returned condensation at the open end starts a rarefaction traveling down the pipe, with the air particles moving upward, opposite to the direction of propagation of the rarefaction. At the closed end the rarefaction will be reflected as a rarefaction back to the open end of the pipe, where, because of the lessened pressure, the air jet is drawn back inside the edge of the pipe. This starts another condensation down the pipe, and so on. During one vibration of the air jet across the edge of the pipe, a condensation and the following rarefaction each travels twice the length of the pipe. The emitted sound will have, then, a wavelength equal to four times the length of the pipe. The latter merely serves as a means of regulating the period of vibration of the air jet and hence the pitch of the sound. The energy comes from the jet. Since the closed end of the pipe is a point of no motion, it is a node, while the open end, being a point of maximum motion, is an antinode of the stationary wave pattern.[1]

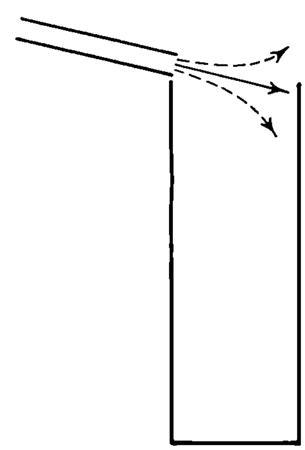


FIG. 38.7. The jet of air vibrates back and forth across the edge of the pipe at a rate controlled by the reflected pulses from the other end.

Similar considerations apply if the pipe is open at both ends. A condensation is reflected from the far end as a rarefaction, however, and when it arrives back

[1] In this discussion we are considering stationary *displacement* waves. The corresponding *pressure* waves would have antinodes when the displacement waves have nodes, and vice versa.

at the jet end, where the jet will already have been forced outside the edge of the pipe by the rise in pressure from the blowing, it will pull the jet back inside the edge. Thus the next condensation will be started down the pipe for the beginning of the next cycle. During one complete vibration of the air jet over the edge the disturbance will have traveled twice the length of the pipe. This length is therefore the wavelength of the emitted sound, and thus if the pipe has the same length as the closed pipe the sound will have double the frequency, or be the octave, of the note from the closed pipe. For this fundamental note the stationary wave pattern in the air column consists of an antinode at each end and a node in the middle of the open pipe.

If the rate of flow of the air from the jet is increased, various overtones will be produced; in fact the blowing must be very gentle indeed if the first overtone is not to be produced with greater intensity than the fundamental. For the *closed* pipe, the open end is always an antinode, and the closed end always a node. Consequently, for the first overtone a second node exists at one-third of the length from the open end, as indicated in Fig. 38.8*b*. The wavelength of this overtone is $4l/3$, if l is the pipe length. For the second overtone of the closed pipe, one node must be one-fifth of the length from the open end, so that $\lambda = 4l/5$. For a closed pipe, then, the fundamental and overtones have $\lambda = 4l/1,\ 4l/3,\ 4l/5, \ldots$ and therefore frequencies $v/4l,\ 3v/4l,\ 5v/4l, \ldots$. Notice that the even-numbered harmonics are missing for a closed pipe.

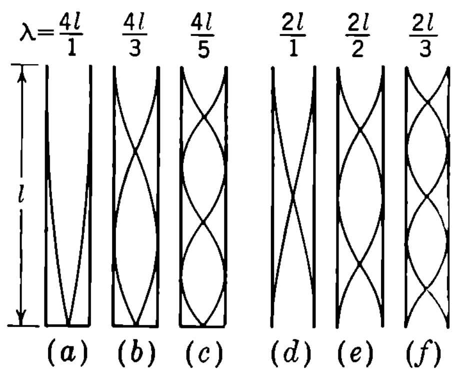


FIG. 38.8. Stationary wave patterns showing positions of nodes for the first three harmonics of closed and open pipes.

The stationary-wave patterns for the fundamental and first and second overtones for the *open* pipe are sketched in Figs. 38.8*d* to *f*. Since the distance between two adjacent loops or nodes is always $\lambda/2$, it follows that the wavelengths of these three partials are $2l/1$, $2l/2$, and $2l/3$. The frequencies of the fundamental and overtones for an open pipe must then go as $v/2l,\ 2v/2l,\ 3v/2l, \ldots$. Thus it is seen that all the harmonics are present in this case. The tones from open pipes are therefore of richer quality than those from closed pipes, the presence or absence of the strong second harmonic accounting for most of the difference.

We have assumed that the reflection at the open end of a pipe takes place exactly in the plane of the end. This is in fact not true, for the reflection occurs only after some expansion into the region beyond the end of the pipe. Since in a large-diameter pipe a greater mass of air vibrates than in a small-diameter pipe, we should expect that the larger inertial effect would carry the expansion farther in the case of the larger pipe. Experiment verifies this conclusion and shows that the end correction to be added is about three-tenths of the diameter.

Worked Example. The longest pipe commonly used in a stop of pipes in an organ is 8 ft. If it is an open pipe, 5 in. in diameter, and if v = 1,100 ft/sec, compute the pitch of the fundamental note. Adding two end corrections, l = 8.25 ft. Therefore, f = (1,100 ft/sec)/2 × 8.25 ft = 66.7 vibrations per sec. Without the corrections, f = 68.8 vibrations per sec.

***38.8. The Vibration of Rods.** A rod may be made to vibrate either longitudinally or transversely. To produce longitudinal vibrations it should be stroked with a rosined cloth; the exact mode of vibration will depend both on the stroking and on the manner in which the rod is clamped. If clamped at the end (Fig. 38.9*a*) the rod's fundamental mode has a node at this point and an antinode at the free end, so that $\lambda = 4l$ as for the closed organ pipe. The resemblance to the closed-pipe vibrations is apparent also in the first overtone, for which the additional node comes at one-third l from the free end, with the result that $\lambda = 4l/3$. To generalize, only the odd harmonics are produced by a rod clamped at one end. If, on the other hand, the rod is clamped in the middle or at points $l/4$ in from each end (Figs. 38.9*b* and *c*), these points are nodes and both ends are antinodes. The first of these, *b*, corresponds to the fundamental mode of open pipes and with $\lambda = 2l$, the second, *c*, gives the harmonic note with $\lambda = l$. As with the open pipe, both even and odd harmonics may be produced by a rod so mounted, with both ends free.

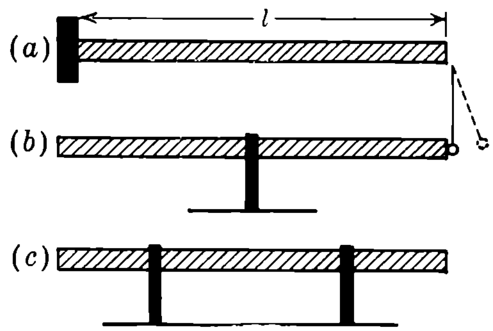


FIG. 38.9. Methods of clamping a rod for longitudinal vibrations.

One way of visibly demonstrating the existence of longitudinal vibrations in a rod is to have a light ball suspended as a pendulum in front of and in contact with one end of the rod. When the latter is stroked with the rosined cloth, the ball will fly out because of the impacts from the rod (cf. Fig. 38.9*b*).

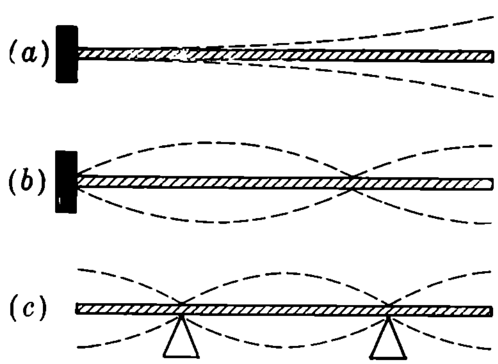


FIG. 38.10. Transverse vibrations of rods.

Transverse vibrations are also possible in a solid rod. If clamped at one end and struck a blow at the free end, a rod will vibrate as in Fig. 38.10*a* when giving off its fundamental tone. When struck more sharply, it may vibrate in its first overtone (Fig. 38.10*b*), but such overtones for transverse vibrations of rods are not integral multiples of the fundamental tone. If supported in two places but with free ends, the rod will vibrate transversely, as shown in Fig. 38.10*c*. With an additional support between these two, the rod may produce its first overtone, but again this is not a harmonic.

If the rod, vibrating as in Fig. 38.10, is bent at the center, the two nodes move closer together; and if it is bent completely into a U shape, we have the familiar tuning fork (Fig. 38.11). When it is vibrating in its fundamental mode, the two nodes are close together at the base and the stem vibrates up and down. If the tuning fork is properly constructed and is not struck too violently, there are no overtones present in its vibration. This fact, that a tuning fork produces a pure tone, makes it a useful device as a frequency standard. Tuning forks are usually made either of steel or of a light magnesium alloy, the latter being a material that prolongs the vibrations because of smaller internal friction. If the fork is mounted on a resonating wooden box having one end closed and with a length approximating $\lambda/4$ for the fork, the tone will be very considerably augmented by the resonance effect discussed below.

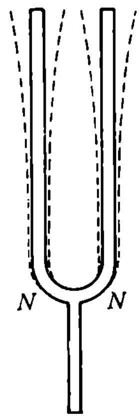


FIG. 38.11. A tuning fork has but one mode of vibration; thus giving a pure tone.

★38.9. The Vibrations of Plates and Membranes. A thin metal plate or a membrane supported at its outer edge will in general, when struck a blow, vibrate in a complicated fashion. In a plate such as the diaphragm of a telephone receiver or transmitter the restoring force called into play by the distortion of the surface arises from its stiffness. In a membrane such as a drumhead or the diaphragm of a condenser microphone the restoring force is caused by the tension in the surface. The two-dimensional waves undergo multiple reflections from the boundaries, thus setting up a standing-wave pattern, which usually involves a number of overtone frequencies along with the fundamental.

The nodes for these two-dimensional stationary waves are *nodal lines.* They may be demonstrated by supporting the plate or membrane horizontally, sprink-

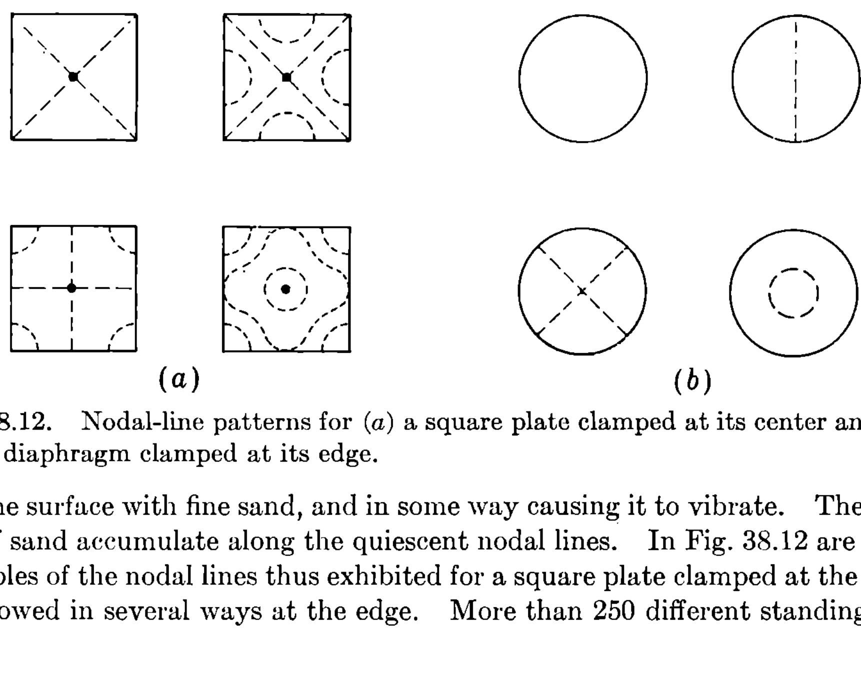


FIG. 38.12. Nodal-line patterns for (*a*) a square plate clamped at its center and (*b*) a round diaphragm clamped at its edge.

ling the surface with fine sand, and in some way causing it to vibrate. The particles of sand accumulate along the quiescent nodal lines. In Fig. 38.12 are shown examples of the nodal lines thus exhibited for a square plate clamped at the center and bowed in several ways at the edge. More than 250 different standing-wave patterns in plates have been demonstrated by this technique. For circular membranes or plates supported at the edge, this boundary is, of course, a nodal line, and for the fundamental mode the entire surface vibrates symmetrically as a whole, with maximum amplitude at the center. The overtones have nodal lines that are diameters and/or circles concentric about the center of the surface (Fig. 38.12). These overtone frequencies are not harmonics.

A bell may be regarded as a round plate bent into the characteristic bell shape. In Fig. 38.13, looking up at the mouth of the bell, the lowest mode of vibration is indicated. Actually, however, overtones are always produced when a bell is struck, and some of these are more intense than the fundamental. The pitch of the sound from a vibrating bell is some *combination or difference tone* (Sec. 38.12) involving these overtone frequencies.

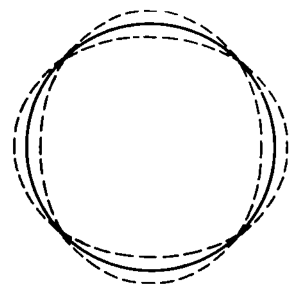


FIG. 38.13. Lowest mode of vibration of a bell.

38.10. Resonance. We have already referred to the phenomenon of *resonance* in connection with the mounting of strings and tuning forks on bodies that by resonating with the applied vibrations from the source greatly amplify the intensity of the sound. As a matter of fact, many important examples of resonance are encountered, not only with sound waves, but also with light and radio waves and in vibrating structures generally (Chap. 12). Whenever the *forced vibrations* in a body have nearly the same frequency as its natural *undamped vibrations* (Sec. 12.14), then by virtue of *resonance* large amplitudes may result for these sympathetic vibrations.

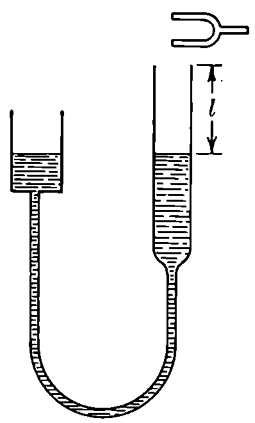


FIG. 38.14. Reinforcement of the sound from the fork occurs when the air column above the water is of length $\lambda/4$, $3\lambda/4$, ...

As a clear illustration of resonance, suppose that two identical tuning forks mounted on quarter-wavelength boxes on rubber pads stand so that the open ends of the boxes face each other. When one of the forks is vibrated, then quickly damped, it is found that the second fork, because of resonance with the train of waves from the first fork, has started to vibrate audibly. If the second fork is then damped, it is noticed that the first fork is again audibly in vibration, but of course with much less intensity since the input energy is rapidly being dissipated.

Resonance between the sound waves from a tuning fork and the stationary waves in an air column affords a quick and fairly accurate method of measuring the velocity of sound in air. The air column is adjusted in length by raising or lowering the water level, as indicated in Fig. 38.14. When the fork is sounding and the reservoir is gradually lowered, a marked reinforcement of the sound will be heard for a particular length l_1 of the air column. This is the first resonant length l_1 for which the reflected waves from the water surface arrive back at the mouth at the right instant to be in phase with the vibration of the fork. Clearly l_1 must be the length of air column having as its fundamental frequency the f of the fork. If we neglect the end correction, $\lambda = 4l_1$, and $v = 4l_1f$. With the fork sounding, upon lowering the water level further the second resonant length l_2 will be encountered. For this, $\lambda = 4l_2/3$, and $v = 4l_2f/3$. The error due to neglecting the end correction may be eliminated by using the difference between the first two resonant lengths, $l_2 - l_1 = \lambda/2$. Then,

$$v = 2(l_2 - l_1)f \tag{38.14}$$

Kundt's method of measuring the velocity of sound affords an additional example of resonance. A metal rod R is clamped at its center and carries on one end a piston P, which fits loosely into a glass tube T(Fig. 38.15). An adjustable

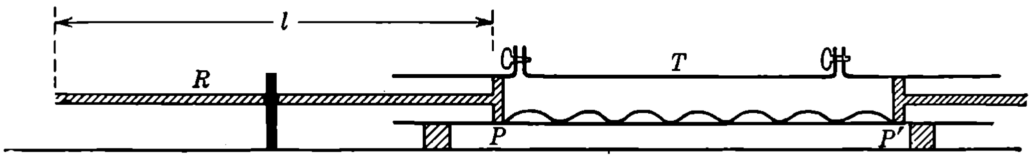


FIG. 38.15. Kundt's method for determining the velocity of sound.

piston P' closes the tube at the far end. The rod is set into longitudinal vibrations by stroking it with a rosined cloth, and the air column T is adjusted in length by means of P' until resonance occurs. The standing waves in the air column are detected with cork dust or lycopodium[1] powder sprinkled along the bottom of tube T. At resonance the dust is strongly agitated and lines up in ridges at the displacement antinodes but is quiescent at the nodes. P' is of course a node, and P is close to a node.

The wavelength λ_R of the waves in the rod is twice the length l of the rod, while in the air inside the tube the wavelength λ_a is twice the average distance d_a between adjacent nodes, as indicated by the cork dust. Now the frequency of the standing waves in the rod and in the air column is the same. Hence

$$\frac{v_R}{\lambda_R} = \frac{v_a}{\lambda_a} \qquad \text{or} \qquad \frac{v_R}{l} = \frac{v_a}{d_a} \tag{38.15}$$

If v_a at the existing temperature is known, the velocity of sound, v_R, in the metal of which the rod is made may thus be determined.

If the air in T is replaced by another gas in which the velocity of sound is v_g and for which the average distance between nodes is d_g, then

$$\frac{v_R}{\lambda_R} = \frac{v_g}{\lambda_g} \qquad \text{or} \qquad \frac{v_R}{l} = \frac{v_g}{d_g} \tag{38.16}$$

[1]From the greek *lykos*, wolf + *pous*, foot; the genus name for the club mosses. The powder is composed of the spores and is very light and easily moved by the sound waves.

***38.11. Musical Scales.** When two or more musical tones are sounded together, the result may be a pleasing sound, in which case the component notes are said to be *consonant,* or it may be a harsh, unpleasant sound, in which case the component notes are said to be *dissonant.* The physical basis for dissonance is the existence of beats (Sec. 37.9) either between the fundamentals or between their overtones. As already mentioned in Sec. 38.9, if the pitches of two tones are so nearly alike as to produce no more than, say, 5 beats per second, the effect is rather pleasing. In fact, on large organs there is a "tremolo stop" in which each of the notes comes from a pair of pipes adjusted to give about 3 beats per second. But as the number of beats becomes larger, they cannot be distinguished as separate beats and the result is a rough, unpleasant sound. When the beat frequency is great enough, however, it will no longer be noticed, and then there is consonance if the two tones have a simple frequency ratio.

The simplest frequency ratio for two notes is 1:2, corresponding to a note and its octave. Other most pleasing combinations are the "fifth" (2:3), the "fourth" (3:4), and the "third" (4:5). If three notes are to be sounded to form a consonant triad and if the interval from the lowest to the highest note in the triad is to be less than an octave, it can be shown that there are only two possibilities, the *major triad* with frequency ratios 4:5:6 and the *minor triad* with ratios 10:12:15. The major and minor scales of modern Occidental music are based on these consonant triads, but it should be emphasized that many other musical scales have been devised. The Arabian scale, for example, has more than 50 intervals.

A major scale of eight notes is formed by three major triads. On taking middle C as the keynote and designating the notes up to the octave by D, E, F, G, A, B, C′, the frequencies of these notes are determined by the three triads CEG, GBD′, and FAC′. In the key of C major the scale of notes, based on the standard pitch of 440 vibrations per second for A, and the frequency ratios are as follows:

Major Scale

	C	D	E	F	G	A	B	C′	D′
Name	do	re	mi	fa	sol	la	si	do	re
Frequency (sec^{-1})	264	297	330	352	396	440	495	528	594
Ratio	9/8	10/9	16/15	9/8	10/9	9/8	16/15		

The ratios 9/8 and 10/9 are called *full-tone* intervals, while the ratio 16/15 is known as a *half-tone* interval.

A minor scale of eight notes based on the same middle C note but using three minor triads yields the same notes as the major scale, except for E, A, and B, with frequencies 316.8, 422.4, and 475.2 cps, respectively. Similar scales composed of three consonant triads may be based on any one of the notes appearing in the key that we have used in our illustration. Each of these keys introduces some new notes, so that, even though the calculations are restricted to major scales, some 35 different notes are needed in each octave interval. Since with keyboard instruments such a large number of notes per octave would be impossible, there was devised a scale of *even temperament*, giving the 12 notes per octave to which we are accustomed in the piano and organ. In this system of *tempering* the scale, successive frequency ratios are each $\sqrt[12]{2}$, or about 1.0595. Two of these steps are taken for a full tone, and one for a half tone. This very simple method of dividing up the octave works remarkably well. The differences between this *equally tempered* scale and the *just scale* are so slight that even the musical ear judges it as satisfactory.

***38.12. Combination Tones.** When two simple tones differing in frequency by a considerable amount are sounded together, a *difference tone* of frequency equal to the difference between the frequencies of the two generating tones may sometimes be heard. Difference tones are to be distinguished from beats. With more difficulty a *summation tone*, whose frequency is the sum of the two frequencies, may be detected. All these are known as *combination tones*. Helmholtz, who contributed so much to the study of sound, was the first to offer the probably correct explanation of these tones, attributing them to the nonlinear response of the ear. In the vibrating mechanism of the ear, especially in the "middle ear" (Sec. 38.14) the restoring force is not a linear function of the displacement. Also, it is asymmetric; *i.e.*, a greater restoring force arises from a displacement in one direction than from a displacement in the reverse direction.

We represent the excess pressure p of the two driving vibrations by

$$p = A\ (\cos \omega_1 t + \cos \omega_2 t).$$

Now the nonlinear response R of the ear may be taken as

$$R = ap + bp^2$$

where a and b are constants. Hence

$$R = aA\ (\cos\ \omega_1 t + \cos\ \omega_2 t) + bA^2(\cos^2\ \omega_1 t + \cos^2\ \omega_2 t + 2\ \cos\ \omega_1 t\ \cos\ \omega_2 t).$$

The three terms in the second parenthesis, the nonlinear part, of this equation are to be interpreted as follows: Since $\cos^2 \omega t = \frac{1}{2} + \frac{1}{2} \cos 2\,\omega t$, the first two of these terms represent the second harmonics of the impressed frequencies f_1 and f_2. For the third term we can make the substitution

$$2 \cos \omega_1 t \cos \omega_2 t = \cos (\omega_1 - \omega_2)t + \cos (\omega_1 + \omega_2)t.$$

We thus see that the required difference and summation frequencies appear in this product term. These difference and summation terms are called side bands (Sec. 35.7) in the electrical case.

38.13. Loudness. The Decibel. In our discussion of sound intensity in Sec. 38.4 we have mentioned that the ear is a remarkably sensitive organ, the lower limit of audibility for a good ear being an intensity of about 10^{-16} watt/cm^2. Yet before the sensation of sound becomes painful, the intensity may rise to about 10^{13} times this value. The loudness of a sound does not increase proportionally with the intensity, however, but more nearly as the logarithm of I. This is an example of the well-known *Weber-Fechner law* in psychology, which states that the magnitude of any sensation is proportional to the logarithm of the stimulus. The minimum fractional increase in intensity that is just perceptible to the ear as a change in intensity should, according to this law, be a constant independent of the intensity. The difference in loudness, technically known as *intensity level*, between two sounds is therefore defined as the logarithm of the ratio of the two intensities, the unit difference being the *bel*.[1] Since this is a rather large unit, one-tenth of it, the *decibel* (abbreviated db), has become standard. The *intensity level* of a sound is then defined by the equation

$$\text{Intensity level} = 10 \log_{10} \left(\frac{I}{I_0}\right) \text{db} \qquad (38.17)$$

where I_0 is a reference intensity arbitrarily taken as 10^{-16} watt/cm^2, which corresponds roughly to the lower limit of audibility.

If the intensity level of a sound increases by 1 db, the increase in loudness is barely perceptible. When I rises to 10^{-4} watt/cm^2 and the sound becomes painfully loud, the intensity level is $10 \log_{10} (10^{-4}/10^{-16}) = 120$ db. In Table 38.1 we give the intensity level for some common noises.

Table 38.1. *Approximate Intensity Levels of Various Sounds*

Noise	*Level, db*
Airplane, near by	120
Inside subway train	100
Noises of busy street	75
Ordinary conversation	70
Quiet automobile	50
Purring cat	25
Rustle of leaves	10
Threshold of hearing	0

For the acoustical engineer, interested in such problems as noise abatement and soundproofing, such description of the intensity level of the sound without regard to frequencies is quite satisfactory. Small pressure amplitudes may be accurately measured with the aid of modern developments in electrical engineering. It happens, however, that the ear is not equally sensitive for all frequencies, so that, at the same intensity level, sounds of different frequencies may be judged to be of quite different loudness.

Figure 38.16 shows for the average human ear both the lower and upper limits of sound intensity level as a function of frequency. Notice that the frequency scale is logarithmic. The course of the lowest curve, giving

[1] The word "bel" was chosen to honor Alexander Graham Bell (1847–1922), the inventor of the telephone.

the threshold of audibility, shows that there is a tremendous difference in sensitivity of the ear for different frequencies. The normal ear has its maximum sensitivity for low intensity levels in the range 2,000 to 4,000 vibrations per second. The highest curve shows that the threshold of pain comes at about the same intensity level for all frequencies. As we

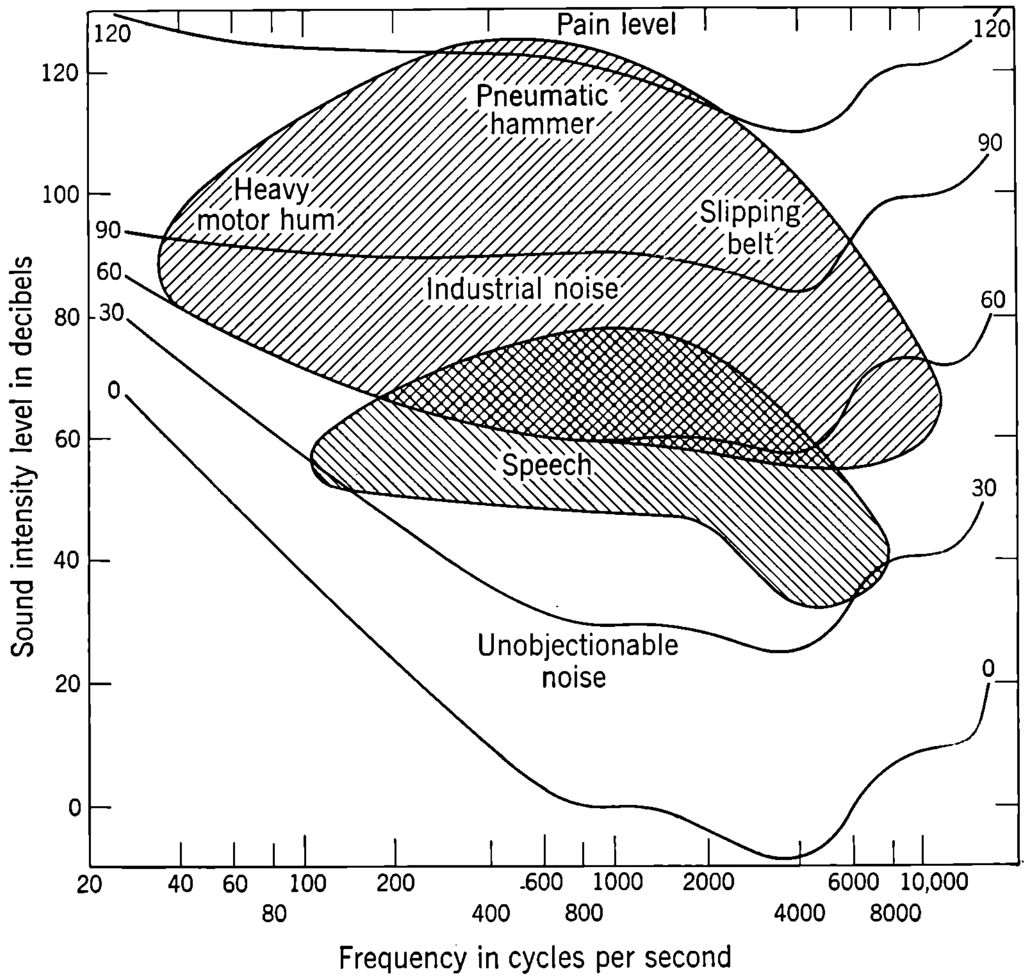


Fig. 38.16. The lowest and highest curves for a normal ear show the threshold of audibility and of pain as functions of frequency. The area between these two curves gives the entire audible range. The numbers labeling the curves represent the decibel level at 1,000 cps.

have previously mentioned, the ear can indeed hear sounds through the range 20 to 20,000 vibrations per second, but at the extreme limits the intensity-level range is small. Unless the intensity level is at least as high as 20 db, even a sound of frequency 200 vibrations per second is not heard, while at 20 vibrations per second the level must be at least 100 db, or close to the threshold of pain. The entire audible range is represented by the area between the two extreme curves. For persons hard of hearing the lower curve is raised by varying amounts at the different frequencies.

The ear is much more sensitive to change in pitch than to changes in intensity. At about 2,500 cycles/sec, for which the ear has the greatest intensity-level range, it is possible to detect a change in frequency of one-fifth of 1 per cent, but a change in intensity must be at least 5 per cent to be detectable. At very low frequencies, however, the change in intensity must be tenfold before a change in loudness can be detected.

Another interesting fact about our auditory perception is the sense of the direction from which the sound comes. The fact that the sound might be more intense at the ear facing the source could only contribute to this effect for the high-pitched notes, for which the wavelength is considerably shorter than the diameter of the head, since for the long wavelengths the pressure variations must be the same at both ears. Experiment shows that for the lower pitches our judgment of direction comes from the difference in *phase* of the sound at the two ears. Such phase differences produce certain time differences in the impulses in the nerve fibers going from the two ears to the brain and these time differences apparently give us a clue as to the direction from which the sound comes.

***38.14. The Ear.** Figure 38.17 is a diagrammatic section through the human ear, with the inner ear drawn on a larger scale to bring out details. That part of the ear to the left of the eardrum is known as the outer ear. Then comes the small middle-ear portion between the eardrum and the oval window. To the right of the oval window is the inner ear. The outer and middle ear contain air, while the inner ear contains watery fluids. Sound waves arriving at the ear pass down the ear canal and cause the eardrum to vibrate. These vibrations are transmitted through the middle ear by three small bones known as the hammer, the anvil, and the stirrup, the latter acting on the membrane over the oval window. Since the eardrum has an area 15 to 20 times that of the oval window and the leverage of the three middle-ear bones multiplies the force by 3, the pressure changes on the liquid in the inner ear are 40 to 60 times those in the outer ear. Also, these bones in the middle ear have a nonlinear transmission characteristic that is responsible for emphasizing harmonics and combination tones, as mentioned in Sec. 38.12.

In the inner ear is the cochlea, only 6 or 7 mm in diameter at its widest part and twisted into a spiral of 2¾ turns like a snail shell. It is the end organ of hearing. Uncoiled, the human cochlea is 30 to 35 mm long. The separating structure that divides it into two canals consists in part of the flexible basilar membrane, along which are about 23,500 so-called "Corti rods." The rods are moved by vibration of the membrane. At the end of each rod is a hair cell, which stimulates the nerve ending at its base.

There is evidence to indicate that the various regions of the basilar membrane in the cochlea are sensitive to different tones, as indicated in Fig. 38.18. These frequencies thus allocated decrease progressively from the oval window to the apex end of the cochlea, this trend being in agreement, for one thing, with the fact that the vibrating mass of fluid becomes greater as the stimulated region goes toward the apex. Also, it is to be noted that resonant vibrations such as these in the basilar membrane must be highly damped, varying in exact synchronism with the stimulus. Considering that our discrimination of pitch is so

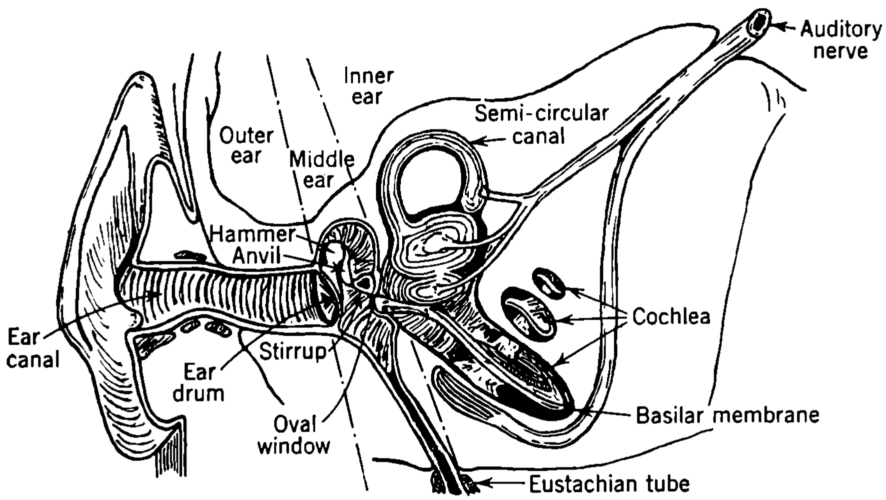


Fig. 38.17. Diagrammatic section through the right ear.

keen, it is indeed remarkable that the resonant mechanisms in this small organ can produce such sensitivitiy.

***38.15. Architectural Acoustics.** The sound arriving at the ear of a listener in a room comes in part directly from the source and in part after reflections from the walls. Now a speaker may utter two to three syllables per second, and a musical instrument produce two or three notes each second, with the consequence

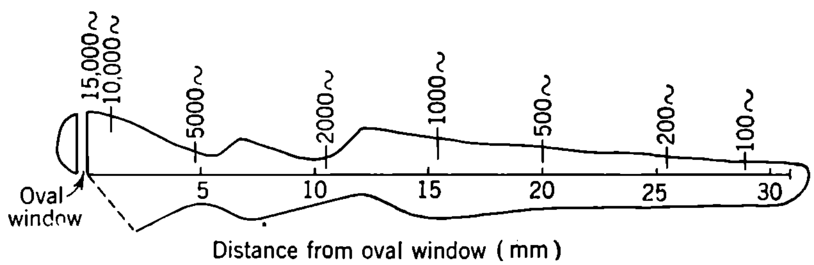


Fig. 38.18. Characteristic frequency regions on the basilar membrane of the cochlea.

that the ear may receive a syllable or note by the direct route at the same instant as the preceding syllable or note after one or more reflections. Reflections will increase the loudness of the sound but may result in making the speech rather unintelligible. The persistence of sound in a room because of reflections from surfaces is known as *reverberation*. By covering the surfaces with porous, sound-absorbing material such as hair felt and draperies the reflected sound may be damped out to any desired degree. Experiment shows that the intensity of the sound undergoing multiple reflections decreases exponentially with the time (like e^{-kt}). The *reverberation time* T is arbitrarily defined as the time required for the intensity to fall to one-millionth of its original intensity or for the intensity level to decrease by 60 db.

This definition of reverberation time for a room was made as a matter of convenience by W. C. Sabine (1868–1919), an American physicist who first placed architectural acoustics on a scientific basis. Before his studies it was common practice to string wires across a hall to disperse the reflected sound waves. Sabine showed the futility of this method and determined by careful measurements with an organ pipe of frequency 512 and a chronograph the coefficients of absorption of various sound-reflecting surfaces and the relation of the reverberation time to the volume and total absorbing power of the hall [Eq. (38.21)].

Suppose that, at each reflection at a wall, a fraction α of the intensity of the sound is absorbed and a fraction $1 - \alpha$ is reflected. The quantity α is called the *absorption coefficient* of the surface. Its average value $\bar{\alpha}$ for the room is obtained by adding the products of the α of each surface times the area of that surface and dividing by the total surface area.

$$\bar{\alpha} = \frac{A_1\alpha_1 + A_2\alpha_2 + \cdots}{A_1 + A_2 + \cdots} \tag{38.18}$$

Some average absorption coefficients are given in Table 38.2.

Table 38.2. *Absorption Coefficients for f = 512 cps*

Open window	1.000
Glass	0.025
Carpet	0.20
Celotex	0.36
Draperies, heavy	0.4–0.6
Hair felt, 1 in. thick	0.78
Plaster	0.025
Wood, varnished	0.03
Adult person	0.44
Upholstered seat	0.30

It is arbitrarily considered that the sound travels an average distance $4V/A$ between reflections (cf. Prob. 16), where V is the volume of the room and A is the total area of its surfaces. Having a velocity v, the average number of reflections these sound waves make in a time t is $Avt/4V$. Therefore the intensity I after this time is,

$$I = I_0(1 - \bar{\alpha})^{Avt/4V} \tag{38.19}$$

where I_0 is the value of the initial intensity. If t is set equal to the reverberation time T, $I = 10^{-6} \times I_0$. Then

$$10^{-6} = (1 - \bar{\alpha})^{AvT/4V}$$

Taking natural logarithms of both sides, $\ln(10^{-6}) = (AvT/4V)\ln(1 - \bar{\alpha})$

or

$$T = \ln(10^{-6}) \times \frac{4V}{Av \ln(1 - \bar{\alpha})} \tag{38.20}$$

Since $\bar{\alpha}$ is usually considerably less than 1 (Table 38.2), we may expand $\ln(1 - \bar{\alpha})$ as a series and retain only the first term, $-\bar{\alpha}$. Then, approximately,

$$T = 0.165 \frac{V}{a} \tag{38.21}$$

where $a = \bar{\alpha}A = A_1\alpha_1 + A_2\alpha_2 + \ . \ . \ .$ is the total absorbing power, V is in cubic meters, and A is in square meters. If V is in cubic feet and A in square feet, the numerical coefficient in Eq. (38.21) becomes 0.049.

In a room whose surfaces are completely covered with sound-absorbing material, the absence of reflections makes sounds dull and lifeless. The reverberation time may be made too short, then, even for speech. Also, since the damping is so great in a perfectly absorbing room, a person would have to produce about 100 times the normal power for ordinary conversation in order to be heard distinctly 30 ft away. But if the reverberation time is too long, the strong reflected sound tends to make speech unintelligible. Making a compromise between these two effects, it has been found that the optimum reverberation time varies from about 1 sec for a large room to about 2 sec for a large auditorium. These values of T are also about correct for music, for which a certain amount of reverberation is desirable in order to blend the notes.

***38.16. Supersonics.** The terms *supersonics* and *ultrasonics* are used for sound having frequencies above 20,000 cycles/sec. These high frequencies are usually produced by a quartz crystal plate mounted between metal electrodes and set into elastic oscillations by resonance with an applied alternating electric field (piezoelectric effect). The natural elastic frequency of the crystal varies inversely as its thickness and also depends on a characteristic modulus of elasticity and the density of the crystal. It is possible to produce supersonic vibrations up to 500 Mc $\sec^{-1}$ (1 Mc = 10^6 cycles) in this way. At this very high frequency the wavelength is but 6×10^{-5} cm in air and 2.4×10^{-4} cm in liquids, or about the length of the waves of visible light. The *magnetostriction effect,* in which a rod of ferromagnetic material in a suitable, rapidly alternating magnetic field vibrates in resonance with the alternating current, is also used to produce supersonic vibrations up to about 300,000 cps.

A beam of supersonic waves passing through a liquid in a cell produces periodic density changes in the direction in which the waves are progressing.

In water, supersonic waves of frequency around 40,000 cps travel many kilometers before their intensity has dropped to half because of absorption, but at 1 Mc this distance decreases to 40 m. In submarine signaling, supersonic waves of frequency about this 40,000-cps value are used for talking from ship to ship, the quartz vibrations being modulated by the speech frequencies. Another application is the detection of submerged submarines by the echo principle. The supersonic signals are suitably pulsed, the transmitter being changed, in the interval between pulses, into a receiver to register any reflected waves. In this way the time taken for the sound to travel to the object and back, and hence the distance of the object, may be found. Essentially this same procedure may be used in echo sounding to determine the depth of the sea. By using supersonic waves the operation is silent, the echo sounding may be continuous, soundings may be taken in shallow water, and the accuracy of measurement is greater than if audible wavelengths are used.

When supersonic waves are passed through a mass of metal, noticeable reflection or absorption of the sound occurs at any blowholes or cracks in the interior. The detection of such flaws in large pieces of metal prior to their failure in use is a matter of considerable importance.

Supersonic waves transform colloidal suspensions, liquid mixtures, and gels into very fine, stable emulsions. Aerosols (mist, dust, smoke, etc.), on the other hand, are coagulated by supersonic vibrations. Various biological effects of supersonic waves have been discovered: microorganisms and small animals in water are killed, red blood corpuscles are destroyed, yeast cells lose their power of reproduction, and milk may be sterilized. All such effects are perhaps in part due to the local generation of heat, but in greater measure they probably are produced by the large acceleration of the particles of the liquid in their very rapid vibratory motion.

PROBLEMS

NOTE: In all problems the velocity of sound may be assumed to be 1,100 ft/sec or 340 m/sec unless its calculation is required.

1. The Diesel engine of a streamliner sounds its horn of frequency 150 vibrations per second while traveling at a speed of 75 miles/hr. What change of frequency would an observer on the station platform hear as the train passes him?

2. Two musicians some distance apart both sound middle C (264 vibrations per second); but because they are slightly out of tune with each other, an observer between detects 4 beats per second. How fast must the observer travel from one musician toward the other so that he will no longer notice the beats?

3. Compute the amplitude of the vibrations in air at 20°C for the faintest sound that can be heard and also for a sound at the threshold of pain. Take the pressure amplitudes as 2×10^{-4} dyne/cm² and 200 dynes/cm², respectively. $f = 1{,}000$ vibrations per sec.

4. The E string of a violin is 36 cm long and has a mass of 0.15 gm. With what tension must it be stretched if tuned to 660 vibrations per sec for its fundamental?

5. What must be the tension in a string 50 cm long and of mass 1 gm to give a frequency of 132 per second?

6. What should be the length of an organ pipe that is closed at one end if it is to have a frequency of 264 per second? What are the frequencies of its first two overtones?

7. Two open pipes of lengths 60 in. and 61 in. are sounded together. How many beats per second are produced by the two fundamental notes?

★8. The first two lengths of air column, above an adjustable water surface for resonance with a tuning fork of frequency 528 per second, are 6.0 in. and 18.5 in. Calculate the velocity of sound.

★9. The dust piles in a Kundt's tube were an average distance of 10 cm apart for air and 8.3 cm for CO_2 gas. Compute the velocity of sound in CO_2.

★10. Calculate the frequencies of an ocatve of notes of a major scale based on E = 330 as the keynote.

11. Denoting the relative frequencies of the first 10 harmonics of a note by the integers 1 through 10, show from a consideration of the ratios between the various frequencies that 7 and 9 give the least harmonious combinations.

12. The intensity level of a sound is increased by 5 db. If the original intensity was 10^{-8} watt/cm², what is the new intensity?

13. For a certain sound in air the waves are essentially plane and have a pressure amplitude of 1 dyne/cm². Compute the intensity of this sound, the number of decibels it is above the lower limit of hearing, and the amount of energy passing through each square meter of a wave front per minute.

14. Two sounds have intensities of 5 and 100 μwatts/cm², respectively. How many decibels is the one sound above the other?

★15. A lecture hall with a volume of 200,000 ft³ is found to have a reverberation time of 2 sec. What is the total absorbing power of all the surfaces in the hall? If the area of all the sound-absorbing surfaces is 30,000 ft², what is the average absorption coefficient?

★16. Using the formula 4 × (vol.)/(surface area) for the average distance that sound travels between reflections in a room, calculate this distance for a sphere and a cube in terms of the radius and edge, respectively. Give the maximum and minimum distances for each of these cases.

17. A sound source moves with velocity v_1', an observer with velocity v_2' in the same direction. Using Eqs. (38.1) and (38.2) calculate the apparent frequency heard by the observer. What if $v_1' = v_2'$?

CHAPTER 39

THE NATURE OF LIGHT

39.1. What Is Light? From earliest recorded times up through the period of Newton's and Huygens' great discoveries, *i.e.*, to about 1700, practically all thinking about the physical world was confined to the fields of mechanics and light. Prior to 1678, when Huygens studied the possibilities of a wave theory, it was commonly believed that light consists of corpuscles shooting out in straight lines from a luminous source. The ancient Greek philosophers knew such facts as the regular reflections of light at a smooth surface where the angle of reflection equals the angle of incidence; they knew qualitatively about the refraction of light at a surface separating two transparent media, such as air and water, and the apparent straight-line, or rectilinear, propagation of light. All these facts could be explained quite well by a corpuscular theory.

The transmission of energy by a beam of light, such as sunlight, was also, of course, known to the ancients. As already noted in Chap. 37. energy may be carried from one point to another in only two ways—either by matter moving between the two points or by a wave disturbance traveling through the intervening medium. A stream of rapidly moving, invisible corpuscles could conceivably carry the energy. As to the speed of propagation of light, philosophers debated up to the time of its first measurement in the second half of the seventeenth century whether the speed is large but finite or whether light is propagated instantaneously. The latter view was held, for example, by the great Kepler (1571–1630) The observed large speed of light (186,000 miles/sec) speaks neither for nor against the corpuscular theory.

During the seventeenth and eighteenth centuries, while the foundations of modern science were being laid, both the wave theory and the corpuscular theory found support among the best scientists. Newton championed the corpuscular theory, chiefly because it was the only way in which he could explain the fact that light travels in straight lines. His contemporary, Huygens, about 1678, showed that the laws of reflection and refraction could be explained on the basis of a wave theory. His ideas did not receive general acceptance, however, because it was not observed that light bends around obstacles as sound waves do. We now know, as we have already shown in our discussion of Huygens' principle in Sec. 37.11, that to detect readily this phenomenon of diffraction the wavelength and the width of the obstacle or opening must be comparable in size, but in Huygens' time the smallness of the wavelengths of light was not known. And the great prestige of Newton and of Laplace (1749–1827), who also supported the corpuscular theory, helped to sustain belief in the correctness of this idea, in the absence of any crucial experiment favoring the wave theory.

Both Hooke (1635–1703) and Grimaldi (1618–1663) almost discovered the wave nature of light, for they did observe slight deviations from the straight-line path when light passes through small openings. The satisfactory explanation of rectilinear propagation on the wave theory, however, did not come until the beginning of the nineteenth century. Thomas Young (1773–1829) in 1801 and Fresnel (1788–1827) in 1814 demonstrated the phenomenon of interference of light and showed that their experiments could be explained only if light is a wave motion. We shall discuss the interference of light in some detail in Chap. 46. Young's double-slit experiment enabled him to measure the wavelengths of visible light, and Fresnel showed conclusively that interference effects between these very short waves could account for rectilinear propagation.

This work of Fresnel was so convincing that the wave nature of light came to be fairly generally accepted. Its acceptance became universal after Foucault (1819–1868) about 1850 proved experimentally (Sec. 39.8) that the velocity of light in a liquid such as water is *less* than the velocity in air, a result contradictory to that demanded by the corpuscular theory. For to explain refraction on this theory it must be assumed that the corpuscles, as they approach obliquely the liquid surface, experience an acceleration in a direction normal to the surface, and if the liquid offers no resistance to the motion of the corpuscles (transparent liquid) this increased velocity should be maintained.

With the fact that light is a wave disturbance firmly established, it seemed necessary at that time (mid-nineteenth century) to believe in the existence of an all-pervading medium, the ether, which transmits these waves. This, however, raised new difficulties. For if these were elastic waves, their velocity should equal the square root of the elasticity of the medium divided by its density (Sec. 37.6). But the very large velocity of light then indicated the elastic modulus to be very large, *i.e.*, the ether should be a very rigid medium, and yet it offers no resistance to the motion of bodies like the planets, which travel through it with undiminished speed. This inconsistency concerning the physical nature of the ether was well known; it was not resolved until the beginning of the present century.

39.2. The Electromagnetic Theory of Light. Another difficulty with the concept of an elastic-solid ether was that a study of the polarization of light (Chap. 48) had shown that light is strictly a transverse wave motion. Now all elastic solids are capable of transmitting longitudinal as well as transverse waves, but no longitudinal waves had been found. Then Maxwell in 1864 brought out his famous electromagnetic theory, which required the vibrations in light waves to be transverse and which indicated a fundamental connection between light and electricity. Maxwell's analysis showed that an oscillating electric circuit should radiate electromagnetic waves in which the vectors representing the electric- and magnetic-field strengths are perpendicular to each other and also in general to the direction of propagation of the waves.

Electromagnetic energy can exist in space, even in the absence of currents or metallic conductors. Just as the vibration of a string produces sound waves which propagate energy away from the string, so an oscillating charge produces *electromagnetic waves* which take energy from the circuit in which the charge flows. In an electromagnetic wave the electric field strength varies with position in space and with time according to a wave equation, just as sound waves vary. The vector **E** is given by

$$\frac{\partial^2 \mathbf{E}}{\partial x^2} = \frac{1}{v^2}\frac{\partial^2 \mathbf{E}}{\partial t^2} \tag{39.1}$$

where v is the velocity of the waves, provided that **E** is independent of y and z. The electric field in radiation is accompanied by a magnetic field determined also by a wave equation. It is customary to employ the vector **H** in radiation problems, and we have

$$\frac{\partial^2 \mathbf{H}}{\partial x^2} = \frac{1}{v^2}\frac{\partial^2 \mathbf{H}}{\partial t^2} \tag{39.2}$$

The wave equations for **E** and **H** can be derived from the four fundamental electromagnetic equations stated in Sec. 32.11.

The velocity v of the waves is related to the properties of the medium. In practical units

$$v = \frac{1}{\sqrt{\epsilon\mu}} \tag{39.3}$$

where ϵ is the permittivity and μ the permeability of the medium. In free space the velocity is denoted by c and

$$c = \frac{1}{\sqrt{\epsilon_0\mu_0}} \tag{39.4}$$

The ratio of the velocity in free space to that in the medium is called the *index of refraction* n of the medium, and

$$n = \frac{c}{v} = \sqrt{\frac{\epsilon}{\epsilon_0}\frac{\mu}{\mu_0}} = \sqrt{K\mu_r} \tag{39.5}$$

where K is the specific inductive capacity of the medium. *Light waves are electromagnetic waves.*

The simplest electromagnetic waves, *i.e.*, the simplest solutions of Eqs. (39.1) and (39.2), are *plane transverse waves*. The vectors **E** and **H** lie in a plane transverse to the direction of propagation of the wave and at right angles to each other, as indicated in Fig. 39.1, where direction of propagation is taken as the z axis. The fields **E** and **H** vary sinusoidally with time t and position z in phase with each

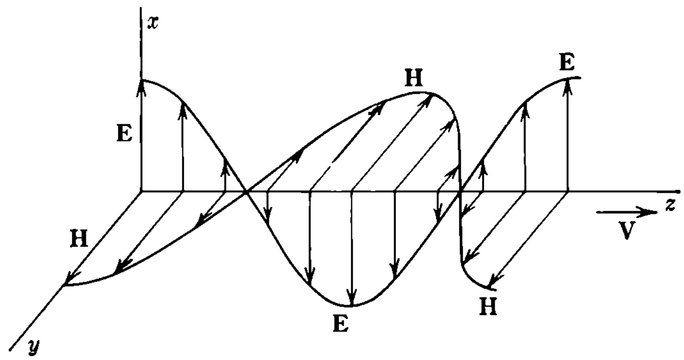


FIG. 39.1. The fields **E** and **H** in a plane transverse electromagnetic wave.

other but are independent of x and y. Therefore the fields are given by the equations

$$\left.\begin{aligned} E_x &= E_0 \cos 2\pi f\left(t - \frac{z}{v}\right) \\ H_y &= H_0 \cos 2\pi f\left(t - \frac{z}{v}\right) \end{aligned}\right\} \quad (39.6)$$

where f is the frequency of the wave. It will be recalled that f is related to the *wavelength* λ and the velocity by

$$v = f\lambda \quad (39.7)$$

We usually think of the electric vector as determining the direction of vibration of the wave. If **E** remains in the same plane, the wave is said to be *polarized;* but according to an old tradition the "plane of polarization" is taken to be the plane of the *magnetic* vector.

The energy density W of the wave is the sum of the energy densities of the electric and magnetic fields. Hence from Eqs. (26.34) and (33.19)

$$W = \tfrac{1}{2}\epsilon E^2 + \tfrac{1}{2}\mu H^2 \quad (39.8)$$

where E and H are the effective values of the fields.

The magnitudes of the fields in a wave are such that the densities of electric and magnetic energy are equal. Hence

$$\tfrac{1}{2}\epsilon E^2 = \tfrac{1}{2}\mu H^2 \quad (39.9)$$

or

$$\sqrt{\epsilon}E = \sqrt{\mu}H \quad (39.10)$$

and

$$W = \epsilon E^2 = \mu H^2 \quad (39.11)$$

The energy S propagated through a unit area per second is related to the energy density W by

$$S = Wv \tag{39.12}$$

or in terms of the field strengths, if Eq. (39.3) is used for v,

$$S = \sqrt{\frac{\epsilon}{\mu}}\, E^2 = \sqrt{\frac{\mu}{\epsilon}}\, H^2 \tag{39.13}$$

The quantity $\sqrt{\mu/\epsilon}$ has the dimensions of an impedance and hence is called the *intrinsic impedance* of the medium. In free space,

$$\sqrt{\frac{\mu_0}{\epsilon_0}} = 377 \text{ ohms}$$

In 1887 Heinrich Hertz (1857–1894) detected these electromagnetic waves radiated from the rapidly accelerated charges in an oscillatory spark discharge. By measuring the wavelength of the waves in a "standing-wave" experiment, and knowing the frequency of the electrical discharge, Hertz determined that the velocity of these waves, now called "radio waves," had exactly the predicted value. It is of interest to note that Hertz's brilliant experiments came 23 years after the publication by Maxwell of the guiding theory. The time interval between theoretical prediction and experimental confirmation is much shorter than that today!

39.3. The Quantum Theory. The conclusion that light consists of electromagnetic waves of smaller wavelength than those produced by an electric circuit is inescapable. Nevertheless, from the beginning of the present century, there has been an impressive accumulation of evidence that when light interacts with matter it behaves as if its energy is contained in packets of value hf, where h is Planck's universal constant of action and f is the frequency of the light. In the photoelectric effect (Chap. 36), for example, there is complete confirmation of the idea that the incident light delivers energy in units of hf. In the Compton effect (Chap. 50), to cite another example, X rays are shown to behave like particles of energy hf in collisions with electrons, although in other experiments X rays behave like electromagnetic waves of short wavelength (1 to 10×10^{-8} cm). The energy packet hf is now referred to as a *photon*.

Our present view is that light has a dual character.[1] In its propagation it consists of electromagnetic waves, but when it interacts with matter in emission, absorption, and scattering processes we must consider it as composed of photons. In optics we are mostly concerned with the study

[1]This duality is only apparent and results from our attempts to explain optical processes in terms of mechanical models, as is shown in modern quantum theory.

of the propagation of light, however, and therefore we shall refer constantly to its wave nature. Also, for the description of most optical phenomena it is not necessary to emphasize the electromagnetic character of the waves. Only when we come to Chaps. 49 and 50 shall we employ the modern corpuscular theory—the so-called "quantum theory."

39.4. Wave Motion. Huygens' Principle.[1] Much of the treatment of wave motion in general in Chap. 37 applies to light waves. The definitions of *wave fronts, rays,* and *trains of waves* hold for light waves, and we shall use these terms continuously throughout our study of light. The description of the course of a train of light waves through an optical system is often simpler if rays are used, a ray being merely a line drawn perpendicularly to the wave front to indicate the direction in which the wave is traveling. Near a small light source the wave fronts are spherical surfaces, and the rays are their radii. When a spherical wave front is sufficiently large, a section of it may be considered as a plane wave and the rays are then nearly parallel straight lines, normal to this plane. In optics spherical waves may be changed into plane waves by means of lenses or mirrors.

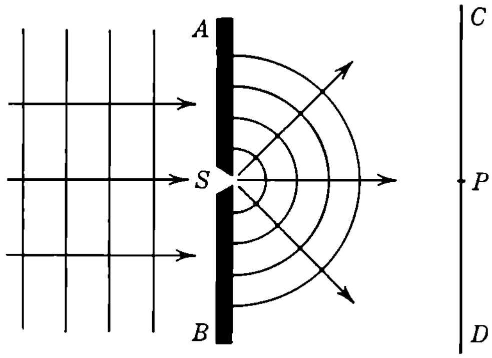


FIG. 39.2. Diffraction of waves at a very narrow opening.

At all points on a wave front the electric intensity and the magnetic intensity oscillate periodically. Since the electric intensity and the magnetic intensity are always at right angles to each other, it might be asked which of these represents the displacement in our previous discussion of wave motion. Actually it could be either one, but in most ways the electric component plays the dominant part, for it is the electric intensity that blackens a photographic plate, causes fluorescent effects, and presumably affects the retina of the eye.

That the ordinary laws of reflection and refraction of waves follow from application of Huygens' principle has already been shown in Chap. 37. The demonstration that the principle of superposition (constructive and destructive interference of two sets of waves) applies to light was first made by Thomas Young in 1801. Diffraction effects will be considered carefully in Chap. 47; but to emphasize the wave nature of light and the correctness of Huygens' principle as applied to it, let us discuss briefly the passage of light through an opening of width small compared with the wavelength. In Fig. 39.2, plane waves from the left fall on the opaque

[1] Review Chap. 37.

plate AB, which contains the very narrow slit S extending at right angles to the plane of the diagram. By Huygens' principle one expects S to be a new source of waves so that to the right of AB the disturbance should fan out as indicated. Experimentally it is found that, if S is considerably wider than the wave length, only a bright line of light at P is observed. As the slit S is gradually narrowed, however, this illumination at P broadens over the entire screen. To be sure, the greatest intensity will occur in the forward direction at P, but with the slit narrower than the wavelength of the light some illumination will be found even at an angle of 90° with the line SP.

39.5. Rectilinear Propagation. For most practical purposes, light does travel in straight lines. It is a matter of common experience that sharp shadows of objects are cast by a small source of light, and we use

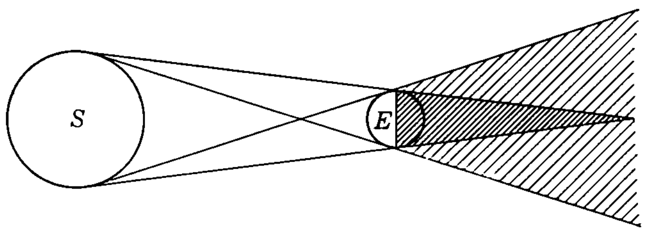


FIG. 39.3. The shadow of the earth illuminated by the sun.

this fact of *rectilinear propagation* whenever we place things in a straight line by "sighting." The line of propagation is a *ray*.

Since all sources of light are of finite size, there is always a region of partial shadow surrounding the region of complete shadow. An interesting example of this is the shadow cast by the earth. The sun being larger than the earth, the region of complete shadow, called the *umbra*, is a cone (Fig. 39.3). Surrounding this is a region, known as the *penumbra*, which is illuminated by light from a part of the sun's surface. Going outward from the umbra, an increasing portion of the surface of the sun contributes light in this region of partial shadow. When the moon is completely in the main shadow cone, there is a total lunar eclipse, while if only a portion of the moon is in the umbra it is partly eclipsed. In the latter case the remainder of the moon's surface which lies in the penumbra takes on a copper-red hue.

The phenomenon of *parallax* arises because of the rectilinear propagation of light. Parallax is the apparent displacement of one object with respect to another when the observer moves. This is illustrated in Fig. 39.4, in which, when the observer's eye moves from E_1 to E_2, the object B appears to move from left to right with respect to object A by an angular displacement $\theta_1 + \theta_2$. The apparent relative motion of near and far objects as viewed from a moving railway train is a matter of common experience. If A is the image from the objective lens of a telescope and B denotes the cross hairs, an absence of parallax between A and B when the eye moves sidewise indicates that A and B are in the same plane. Parallax is also of importance in certain astronomical observations.

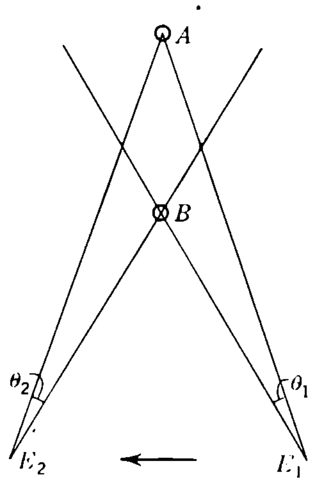


FIG. 39.4. Parallax; the apparent motion of B with respect to A when the eye moves from E_1 to E_2.

A further example of the rectilinear propagation of light is the formation of an image by a pinhole (Fig. 39.5). As the pinhole becomes smaller, the cone of rays, or *pencil*, proceeding from any point of the source and defined by the edges of the hole becomes narrower and the inverted image on the screen becomes sharper. Perfectly good photographs may be made in this way with a small pinhole substituted for the lens in a camera, but of course the exposure time must be prolonged (several minutes in full sunlight) because of the small amount of light passing through the tiny aperture. Although it is true that the sharpness of the image increases as the pinhole decreases in size down to a certain size, diffraction effects will actually make the image more diffuse again if the aperture is decreased further.

39.6. The Velocity of Light. The velocity with which light is propagated in empty space is one of the fundamental constants of the physical world. It has the same value for all frequencies and is the same not only for visible light but for all electromagnetic radiation from the longest radio waves to the shortest X rays. Because it is so very large (186,000 miles/sec), its accurate measurement is difficult. Galileo was the first to attempt it by a method correct in principle but doomed to failure. He caused two observers on hilltops about a mile apart to flash light signals to each other, each noting the time required for the acknowledging signal to come back from the other. This would serve for a fair determination of the velocity of sound in air, but because of the short distance and the reaction time of the observers it could not serve for the measurement of the huge velocity of light.

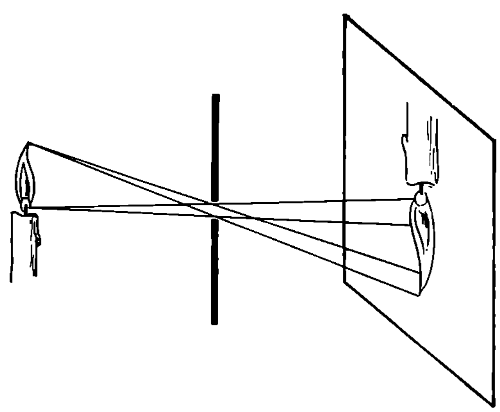


FIG. 39.5. A fairly sharp inverted image is formed by a pinhole.

A few years later (1676) Olaf Römer, a Danish astronomer then working in Paris, made the first successful determination of this velocity by an astronomical measurement. He was engaged in a careful study of the periods of revolution of the satellites of the planet Jupiter. There are 11 of these satellites, all of which revolve in orbits nearly in the same plane as that in which the earth and Jupiter move. Therefore the satellites are eclipsed by Jupiter once during each revolution about the planet. Römer made a long series of observations on the times of the eclipses of the inner satellite, measuring always the times of emergence from the shadow, and determined accurately the average period. This average period of revolution for the inner satellite is 42 hr 28 min 16 sec.

Römer first noted the exact time of an eclipse when the earth was at E_1 and Jupiter at J_1 (Fig. 39.6). Then, using the average period for this satellite, he computed the time for a later eclipse when the earth was, say, at E_2. But this calculation did not agree with the facts: the eclipse was observed to come more than 10 min later than the calculated time. Since the period of revolution of Jupiter about the sun is nearly 12 years, during this interval Jupiter will have moved only to J_2. Römer decided that delays such as this in the observed eclipse times are due to the time required by light to travel the increase in the distance between Jupiter and the earth. His conclusion was that it takes about 22 min for light to go a distance equal to the diameter of the earth's orbit. In Römer's day this distance was thought to be 172,000,000 miles, and hence the indicated value of the velocity of light was 130,000 miles/sec, or 227,000 km/sec.

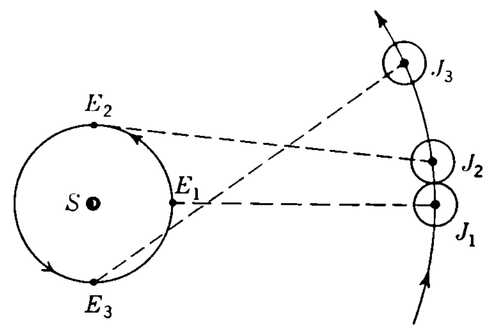


Fig. 39.6. Römer's observations on Jupiter's moons from which the velocity of light could be determined.

Many of Römer's contemporaries refused to believe that the velocity of light could be so large, but in 1727 the English astronomer Bradley determined this velocity conclusively by an entirely independent astronomical method. He discovered that the apparent direction of the light reaching the earth from a star is altered by the motion of the earth in its orbit. This effect, known as *aberration*, causes all stars observed in a direction perpendicular to the plane of the earth's orbit apparently to move in circles of angular diameter about 41 sec. To understand this effect an analogy will be helpful. Suppose a raindrop falls down the axis of a vertical tube (Fig. 39.7*a*), the velocity of the drop being V. If the tube is at the same time moved at right angles to its length with a velocity v, the raindrop will have a horizontal component of motion relative to the tube and will hit its side. The path of the drop *relative to the tube* is BC. Obviously, if such raindrops are to move down the axis of the tube, the latter must be tilted from the vertical through an angle α given by

$$\tan \alpha = \frac{v}{V}$$

Now let the tube be a telescope and the raindrops be light coming with velocity V from a star in a direction perpendicular to the velocity v of the earth in its orbital motion about the sun. The telescope must be tilted forward through the angle α if the ray of light from the star is to travel down the axis of the tube. While the earth makes one revolution about the sun, then, such a star will appear to move in a small circular path of angular diameter 2α (about 41 sec). The velocity v of the earth in its orbit is about 18.5 miles/sec. On combining the best recent values of these two quantities, V is calculated to be 186,200 miles/sec.

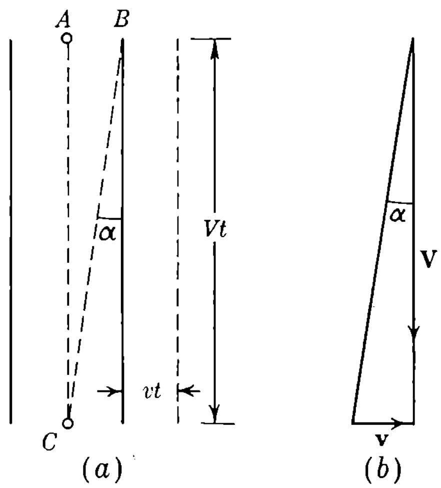


FIG. 39.7. Aberration of light.

39.7. Velocity of Light Measured by Direct Methods. The first successful direct measurement of the velocity of light was made by A. H. L. Fizeau (1819–1896) in Paris in 1849. He used a rotating toothed wheel to interrupt the light, which then traveled over a two-way light path 10.72 miles in length. When the wheel was rotated at a certain speed each returning light pulse was eclipsed by the next cog following the open space on the wheel. From the known speed of the wheel and the length of the path the light traveled, the velocity of light could be calculated (Prob. 3).

An important modification of this method was made by J. B. L. Foucault (1819–1868) in 1850, the principal change being to substitute a rotating mirror for the cogwheel. Foucault's method was improved and brought to a high state of perfection by A. A. Michelson (1852–1931) in a long series of experiments begun while he was an instructor at the U.S. Naval Academy in 1878 and still in progress at the time of his death, to be completed in 1932 by Pease and Pearson. Michelson introduced a long-focus lens L (Fig. 39.8) between the rotating mirror m and the distant mirror M. In this way the distance D could be greatly increased. If the mirror m is stationary, the returning beam of light is brought to a focus at S. But if, during the time when the light is traveling the relatively long path $2D$, m has turned through an angle θ, the returning light will be focused at S', where the angle SmS' is 2θ. If m makes n

revolutions/second and T is the time required for light to travel the distance $2D$, then $T = \theta/2\pi n$ and the velocity of light c (the usual symbol for this constant of nature) is given by

$$c = \frac{2D}{T} = 4\pi Dn/\theta$$

The angle 2θ may be obtained from the displacement SS' of the image.

In 1926, at the Mt. Wilson Observatory, Michelson considerably increased the accuracy of such measurements by using accurate, many-sided rotating mirrors. The speed of these was adjusted until one face just replaced an adjacent face during the time for the light to travel the distance $2D$. The distance D (Mt. Wilson to Mt. San Antonio) is 22 miles, and it was measured by triangulation to the remarkable accuracy of about ⅛ in. For an octagonal mirror the required speed for the next face to replace the first reflecting face was about 528 revolutions/ sec. The average of a large number of observations yielded the value 299,796 km/sec for the velocity of light.

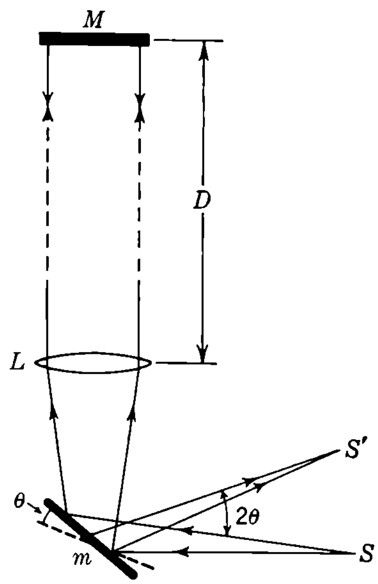


FIG. 39.8. Rotating mirror method of measuring the velocity of light.

There have been a number of modern experiments to measure the velocity of light, using electrical methods. They all give values of the velocity in excellent agreement with that from the last experiments started by Michelson. The latest and most interesting possibility is the timing of reflected pulses of "microwaves" returned from distant objects, using the radar techniques perfected during the Second World War. In 1946 the U.S. Signal Corps succeeded in detecting a radar signal reflected from the moon. Since the mean distance to the moon is some 238,860 miles, the trip there and back takes the pulse of microwaves, which are electromagnetic waves identical with light but of longer wavelength, over 2.5 sec. This time interval may be measured to an accuracy of probably 10^{-6} sec, and with the distance to the moon well known for stated points in its orbit this should constitute another method of determining the velocity of light in empty space. Irregularities on the surface of the moon would be expected to lessen the accuracy of this particular experiment.

From a careful analysis of all the recent measurements Birge has suggested for the best value of the velocity of light

$$c = 299{,}776 \pm 4 \text{ km/sec}$$

39.8. Velocity of Light in Matter. When a pencil of light crosses obliquely from air into water, it is bent, or *refracted*, at the interface, the angle of refraction r being less than the angle of incidence i. Furthermore the shorter wavelength blue light is refracted more strongly than the longer wavelength red light (Fig. 39.9). Newton explained this by saying that the corpuscles of light are attracted toward the surface of the denser medium and that therefore they should have a greater velocity in this medium. To explain the direction of this refraction on the wave theory, using Huygens' principle (Sec. 37.13), we require that the velocity in the liquid be *less* than that in air. It is thus possible to decide between the two theories by determining experimentally the speed of light in water.

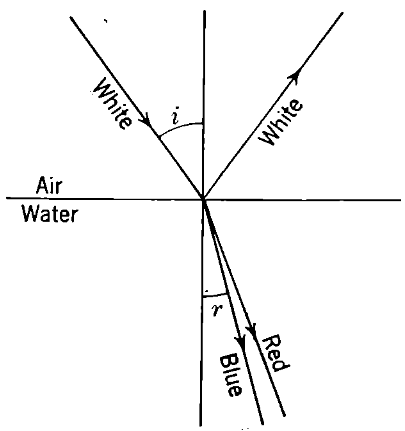


FIG. 39.9. Refraction.

Inserting a long tube of water into the light path, Foucault observed that light does indeed travel more slowly in water. Michelson made more accurate measurements of this kind in 1885. Using white light, he found the velocity in air to be 1.330 times greater than that in water, while the velocity in carbon disulfide was found to be less than that in air in the ratio 1:1.758. He also noticed that with carbon disulfide in the path the final image of the source was spread out into a short spectrum, which indicated that blue light was slowed more than the red light in traveling through the medium. These findings were all in complete agreement with the wave theory of light.

The ratio of the velocity of light in vacuum to that in a medium is called the index of refraction of the medium and is denoted by n. This quantity may be determined by application of Snell's law (Sec. 37.13);

$$n = \frac{c}{v} = \frac{\sin i}{\sin r}$$

In practice n is usually reckoned as the ratio of the velocity in air to that in the medium, since for air under standard conditions n is almost unity ($n = 1.0002926$ for yellow sodium light). For gases, n is always near unity. For transparent liquids and solids, n ranges from 1.333 for water to 2.42 for diamond. The subject of refraction is treated more fully in Chaps. 40 and 41.

39.9. The Origin and Lengths of Electromagnetic Waves. Most light sources are either bodies at high temperatures, such as a lamp filament, an electric arc, or the sun, or they are electric discharges in gases at reduced pressure, such as neon signs and fluorescent lamps. In all cases the light originates in the vibrations of charged particles, chiefly electrons, of which matter is composed. As already mentioned in Chap. 22, all frequencies are present in the radiation emitted by an incandesent solid. This will be discussed in more detail in Chap. 48. In an electric discharge in a gas the atoms and molecules radiate the energy that they have acquired either by collisions with electrons or by absorption of light from other similar atoms and molecules. The emission and absorption of light by matter is always a quantum phenomenon, a light quantum or photon having energy proportional to its frequency.

The quantum theory of the energy states of heavier, more complex atoms resembles that for the hydrogen atom.[1] When one of the outer valence electrons of an atom is knocked into a larger orbit, the return of such an electron to its stable orbit closer to the nucleus is accompanied by radiation, the energy radiated being the difference in the energies in the two states. This radiation is usually in the visible range. Such "spectra" will be treated in Chap. 49.

As for all wave motions, the velocity of propagation, v, is the product of the wavelength λ and the frequency f. The wavelengths of visible light range from 4×10^{-5} cm for the shortest wavelength violet light to 7×10^{-5} cm for the longest red wavelengths. These are often expressed in microns [1 micron (μ) = 10^{-6} m = 10^{-3} mm], millimicrons [1 millimicron ($m\mu$) = 10^{-6} mm], or Angstrom units [1 Angstrom unit (A) = 10^{-7} mm]. The extreme values of λ for visible light are, then,

$$4 \times 10^{-5}\text{ cm} = 0.4\ \mu = 400\text{ m}\mu = 4{,}000\text{A (shortest violet } \lambda)$$
$$7 \times 10^{-5}\text{ cm} = 0.7\ \mu = 700\text{ m}\mu = 7{,}000\text{A (longest red } \lambda)$$

The orange-yellow light from a sodium flame or the modern sodium-vapor lamp has a wavelength

$$\lambda = 0.00005893\text{ cm} = 0.5893\ \mu = 589.3\text{ m}\mu = 5{,}893\text{A}$$

Spectroscopists always use Angstrom units; they are able to measure wavelengths to the third decimal place in Angstrom units, or to 10^{-11} cm, which is indeed high precision!

Since $v = \lambda f$, and the velocity of propagation in a transparent medium such as water is less than that in empty space by a factor $1/n$, the question arises whether it is the wavelength or the frequency of the light in the medium that is reduced in this same ratio. Direct measurement in a double-slit interference experiment in water (Chap. 46) shows that the

[1] Cf Sec. 11.9.

wavelengths are just three-fourths ($= 1/n$ for water) those found in air. A little reflection would lead one to expect this result; for if a train of light waves passes through a transparent material, the number of waves leaving the substance per second must equal the number entering per second.

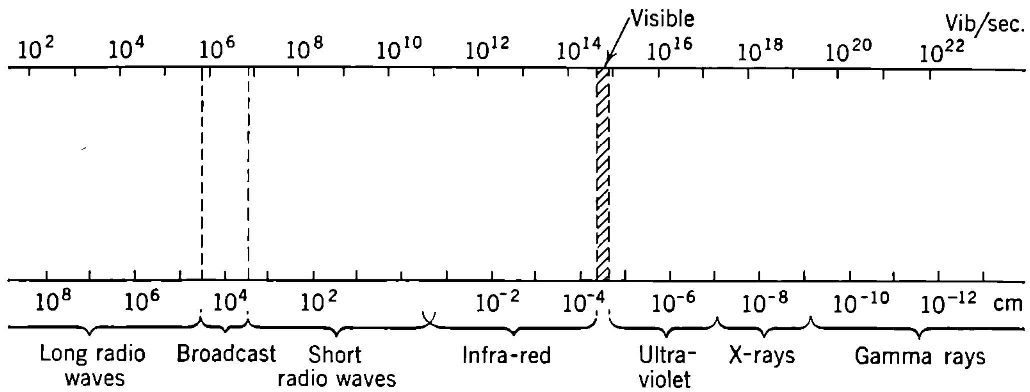


FIG. 39.10. Chart of the entire electromagnetic spectrum.

On the long-wavelength side the electromagnetic radiation spectrum is unlimited; for since $\lambda = v/f$, the radiation from a charge that oscillates slowly, with a period of, say, 2 sec, will have a wavelength

$$\lambda = (3 \times 10^{10} \text{ cm sec}^{-1})/(0.5 \text{ sec}^{-1}) = 6 \times 10^{10} \text{ cm} = 6 \times 10^5 \text{ km}.$$

The ordinary 60-cycle a-c circuit produces a wavelength

$$\lambda = \frac{3 \times 10^{10} \text{ cm/sec}}{60 \text{ sec}^{-1}} = 5 \times 10^8 \text{ cm} = 5{,}000 \text{ km}$$

Radio waves of the very highest frequency, so-called "microwaves" or "radar waves," have now been produced with wavelengths less than 1 cm, and as long ago as 1920 Nichols and Tear experimented successfully with electromagnetic waves as short as 1.8 mm, using sparks between tiny pieces of metal foil. Hence a gap between radar waves and infrared waves no longer exists. In fact, characteristic absorption frequencies of certain molecules have been detected by using microwaves in the centimeter range, and a molecule that absorbs such frequencies also can emit them (Chap. 48). Similar overlapping occurs between all the other regions of the electromagnetic spectrum. A chart of the entire electromagnetic spectrum is given in Fig. 39.10. To be noted are the short range of the visible wavelengths and the fact that the complete chart does not end at the extremes shown. The only limit to the frequency of gamma radiation is apparently the energy of the bombarding particles. A particle accelerator now being planned will produce gamma rays of frequency up to 10^{24} vibrations/sec.

PROBLEMS

1. Assuming the earth to move in a circular orbit of radius equal to our mean distance from the sun, 92,900,000 miles, compute the time for light to travel a distance equal to the diameter of the earth's orbit.

★**2.** Calculate the maximum aberration angle due to the rotation of the earth on its axis, the telescope being placed on the earth's equator.

3. Fizeau's cogwheel had 720 cogs, the cogs and spaces between being of equal width. His two-way light path was 10.72 miles. Using the present value of the velocity of light, calculate the minimum angular velocity of the wheel to produce eclipses of the light pulses returning from the distant mirror.

4. In an experiment on the velocity of light such as Michelson's on Mt. Wilson, if the distant mirror is 35.0 km away, what must be the speed of rotation of an octagonal mirror in order that the following mirror face be exactly in position to reflect the returning pulse back to the source?

5. Calculate the elapsed time for a microwave beam to travel to the moon and back when the moon is 238,860 miles distant, taking our best value for c. If this is to be used as an experiment to determine c and the probable error in the distance is ± 5 miles and that in the time is $\pm 10^{-6}$ sec, calculate the percentage error in the velocity c.

6. Calculate the frequency of light of wavelength 6,000 A; of X rays of wavelength 1A.

7. The index of refraction n for water is 1.33 and for a certain glass $n = 1.60$. Calculate the wavelength of the orange sodium light in these two substances.

★**8.** Using Eq. (39.4) and the values 377 ohms for $\sqrt{\mu_0/\epsilon_0}$ and $4\pi \times 10^{-7}$ henry/m for μ_0, calculate the velocity of light c.

9. Planck's constant h is 6.62×10^{-27} erg sec. Calculate the energy in ergs and in electron-volts of a photon of violet light of wavelength 4,000 A.

10. Assuming that $c = 3 \times 10^{10}$ cm/sec and that for air at NTP $n = 1.0003$, calculate the number of waves per cm (called the wave number) in vacuum and in air for orange sodium light.

CHAPTER 40
REFLECTION AND REFRACTION AT PLANE SURFACES

40.1. Geometrical Optics. In Chap. 37 we derived with the aid of Huygens' principle the laws for the reflection and refraction of plane wave fronts at the surface between two media. Since light is a wave motion, these laws should apply to the reflection and refraction of light. We considered in these derivations only the case of plane waves, but it can be shown that these laws hold in general for waves of any shape. In this chapter, for example, we shall use Huygens' principle in the analysis of the reflection and refraction of spherical wave fronts at plane boundary surfaces. Provided diffraction effects may be ignored, however, it is convenient, in considering the changes in direction of light at surfaces that in optical instruments are usually curved, to employ only the normals to the waves, or the *rays*. That portion of the study of light which deals with the course of rays as they undergo successive reflections and refractions is known as *geometrical optics*. The remainder of the subject, including topics such as energy transmission by light, interference, diffraction, and polarization, can be discussed only in terms of waves. These topics are grouped under the heading *physical optics*.

40.2. Plane Waves Reflected at a Plane Surface. We shall consider only *regular*, as distinguished from *diffuse*, reflection. The latter type of reflection occurs at a rough surface; but since the reflection from each tiny element of the surface follows the laws of regular reflection, we need not discuss the diffuse variety further. The case of a plane wave incident obliquely on a smooth plane reflecting surface has been treated in Sec. 37.12. Using Huygens' principle, it was shown that the reflected wave is also plane and that *the angle of reflection is equal to the angle of incidence*. In Fig. 37.13 the incident and reflected rays are also seen to make equal angles with the normal to the reflecting surface. In this diagram the planes of the incident and reflected wave fronts and the plane of the mirror surface are all perpendicular to the plane of the paper. Therefore the incident and reflected *rays* and the normal to the surface, being each perpendicular to one of these planes, *all lie in the same plane*, the plane of the diagram. From the equality of the angles of incidence and reflection it follows that, if a mirror is rotated through a given angle about an axis perpendicular to this plane, the reflected ray will be rotated through twice that angle.

For this *regular* reflection of light the surface must be so smooth that irregularities are small as compared with the wavelength of the light. Polished metal surfaces form the best mirrors, of course, but a considerable amount of regular reflection occurs, also, when light passes from air into substances such as glass or water. In these cases the fraction of the incident light regularly reflected increases with the angle of incidence. If a beam of light strikes a glass or water surface at grazing incidence (i almost 90°), the reflected light is bright, the intensity of the light refracted into the transparent medium being correspondingly weak.

40.3. Plane Waves Refracted at a Plane Surface. When a wave front passes obliquely from one medium into another in which it is propagated with a different velocity, its direction will change. This has already been shown with the aid of Huygens' principle in Sec. 37.13. Since for light the velocity v_2 in the lower, denser medium is less than the velocity v_1 in the upper, less dense medium, we redraw Fig. 37.16 to indicate that the angle of refraction r is now less than the angle of incidence i (Fig. 40.1). Let the index of refraction of medium 1 be n_1, that of medium 2 be n_2. Then, by definition,

$$n_1 = \frac{c}{v_1} \qquad \text{and} \qquad n_2 = \frac{c}{v_2}$$

where c is the velocity of light in empty space. Therefore

$$\frac{v_1}{v_2} = \frac{n_2}{n_1}$$

and thus Snell's law [Eq. (37.36)], may be written

$$\frac{\sin i}{\sin r} = \frac{n_2}{n_1}$$

or

$$n_1 \sin i = n_2 \sin r \tag{40.1}$$

If medium 1 is air, then n_1 for most purposes may be set equal to unity and the ratio $\sin i/\sin r$ is a constant n_2 for a given medium 2, say glass. Then, since $\sin i > \sin r$, the rays will be bent toward the normal upon passing from air into glass, as indicated in Fig. 40.1. If the light is traveling in the reverse direction, the rays will be bent away from the normal when entering the air. The incident ray, the refracted ray, and the normal to the surface at that point all lie in the same plane.

As an example of the application of Eq. (40.1), suppose that medium 1 is water of index 1.33, that medium 2 is glass of index 1.50, and that $i = 30°$. Then

$$1.33 \times 0.5 = 1.50 \sin r$$

$$\sin r = 0.443 \qquad \text{or} \qquad r = 26.3°$$

If the light is incident on the glass-water interface from the glass side with $i = 26.3°$, then $1.50 \times 0.443 = 1.33 \times \sin r$, or $\sin r = 0.5$, and $r = 30°$. In other words, the light is exactly reversed in its path.

If a pencil of light passes obliquely through a slab of transparent material with parallel sides, there will be some lateral displacement of the

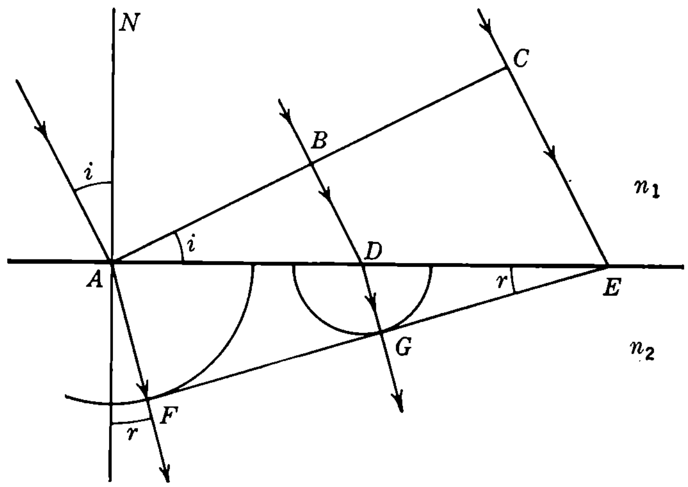


FIG. 40.1. Refraction of a plane light wave at a plane surface.

rays but no change in their direction. This is illustrated in Fig. 40.2. The deviation of the rays by refraction at the first surface is exactly compensated by a deviation in the opposite sense at the second surface. An expression for the lateral displacement LM in terms of the thickness d of the slab and the angles i and r may readily be calculated (cf. Prob. 7).

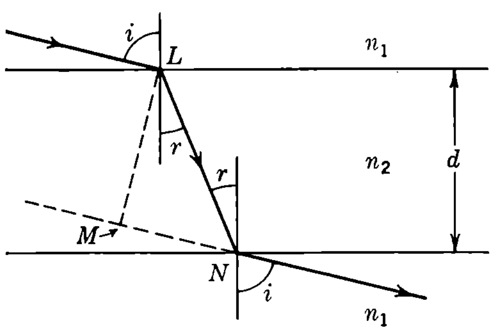


FIG. 40.2. Ray passing through refracting medium with parallel sides.

40.4. Total Reflection. In the example just cited there is some light reflected at the boundary, the light going in either direction and at any angle of incidence. Of particular interest is the case of light traveling in a denser medium and striking a plane refracting surface. On taking water and air as the two substances the division of the light between the refracted and reflected beams may be rendered visible by placing fluorescin in the water in a rectangular glass tank. Upon injecting a beam of *plane-parallel light* (train of plane waves, hence parallel rays) up at an angle through one of the glass sides it is evident that as the light strikes the water-air surface more and more obliquely an increasing fraction of the light is reflected back into the water. Finally, as the refracted waves move parallel to the surface ($\angle r$ in the air equals 90°), the light becomes 100 per cent reflected. For all angles of incidence in the water greater than this *critical angle*, the light is totally reflected back into the water.

This situation is illustrated in Fig. 40.3, in which we draw just the rays. For light going in the direction l, nearly 50 per cent is reflected back into the denser medium, the emainder being refracted in the direction l'. Even for normal incidence ($\angle i = 0$) about 4 per cent of the light is reflected at the surface back into the water. Approximately this amount of light energy is lost by reflection at normal incidence at every refracting surface in any optical instrument. A ray such as m, for which the refracted ray has $\angle r = 90°$, is said to be incident at the critical angle C, since total reflection occurs for any ray such as o incident more obliquely. For $\angle i = C$ it follows from Eq. (40.1), on taking $n = 1$ for air, that

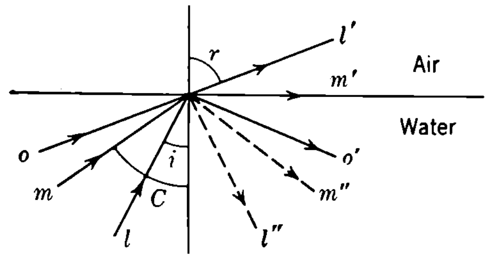


FIG. 40.3. Total reflection occurs when $\angle i > C$.

$$\sin C = \frac{1}{\text{index of dense medium}}$$

Water has an index of refraction equal to 1.33, and hence $C = 48°36'$. A diver in the water looking upward would thus see the whole of the space above the water concentrated into the cone of angle $2C = 97°$ at his eye. Various kinds of glass have indices of refraction such that C lies between about 37 and 43°. Diamond has the uniquely small value 24°26′ for C. Advantage can be taken of this in cutting regular facets so as to cause a large proportion of the light to be totally reflected by the diamond, the multiple reflections inside the diamond producing the characteristic brilliance.

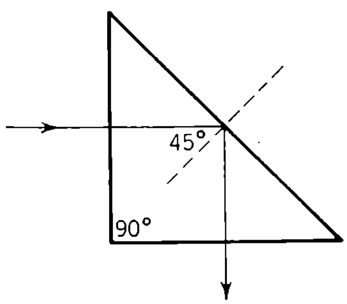


FIG. 40.4. Total-reflecting prism.

Since the critical angle of glass is always less than 45°, a simple right-angled prism (Fig. 40.4) serves by total reflection to turn a beam of light through exactly 90°. No refraction takes place at the entrance and exit faces because of the normal incidence. This device is used, rather than a mirror, in many optical instruments, such as binoculars, because of its greater efficiency, the only loss of light being the small (approximately 4 per cent) loss by reflection at the two right-angled surfaces.

Several more complicated forms of glass prisms for changing the direction of light rays, reversing images, etc., all making use of total reflection, have been perfected. The principle of the "erecting prism" is shown in Fig. 40.5. A converging beam of rays such as AB, which would have come to a focus at I', is refracted and totally reflected so as to give a reversed image at I. If I' would normally have been an inverted image from some optical instrument, this forms a convenient way of rendering the image erect.

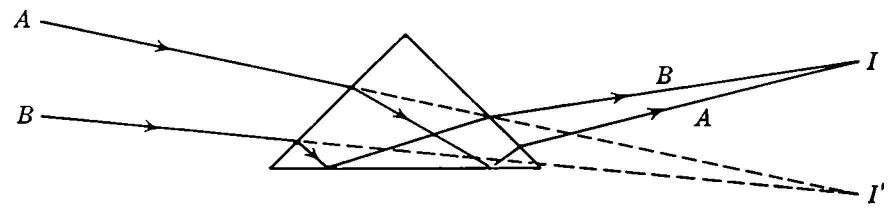


FIG. 40.5. Principle of "erecting prism."

If the less dense medium is not air but a liquid of index n_1, the more refractive medium being of index n_2, then, from Eq. (40.1),

$$n_1 \sin 90° = n_2 \sin C, \text{ or}$$

$$\sin C = \frac{n_1}{n_2} \qquad (40.2)$$

Thus if n_2 is known, the index n_1 of the other material may be determined by measuring the angle of total reflection in the denser medium at the surface of contact of the two.

***40.5. Spherical Waves Reflected at a Plane Surface.** Suppose that spherical waves originate at the point S (Fig. 40.6) and fall on the plane reflecting surface LM. Let ABC be one of the incident wave fronts, with B just becoming the center of a wavelet. Then successively other points on the surface become centers of disturbance. The envelope of some of these wavelets is represented in the figure by DEF for the instant when the advancing wave front would have arrived at DGF had there been no mirror. The wavelet radii satisfy the relations $BE = BG$, $JK = JK'$, etc. Therefore the reflected wave front DEF and the arc DGF must be similar and have equal radii of curvature. The wave front DEF seems to come, then, from the center S', which is called the *virtual image* of S. Evidently S' is as far behind the mirror surface as S is in front of it.

If the eye is at E, the pencil of light entering the pupil and focused on the retina will be a cone from the virtual image at S'. Actually the rays of this pencil come from S, as shown in the figure. For any ray such as SHI the angles SHN and NHI made with the normal NH to the mirror at the point of incidence are equal.

40.6. Images. If the light source is an extended body such as AB in Fig. 40.7, the image of each point is as far behind the mirror as the point is in front. Therefore the virtual image formed by a plane mirror is of the same size as the object, and the two are symmetrically placed on either side of the mirror. An eye at E receives pencils of light as indicated.

If the observer stands between the object AB and the mirror and views

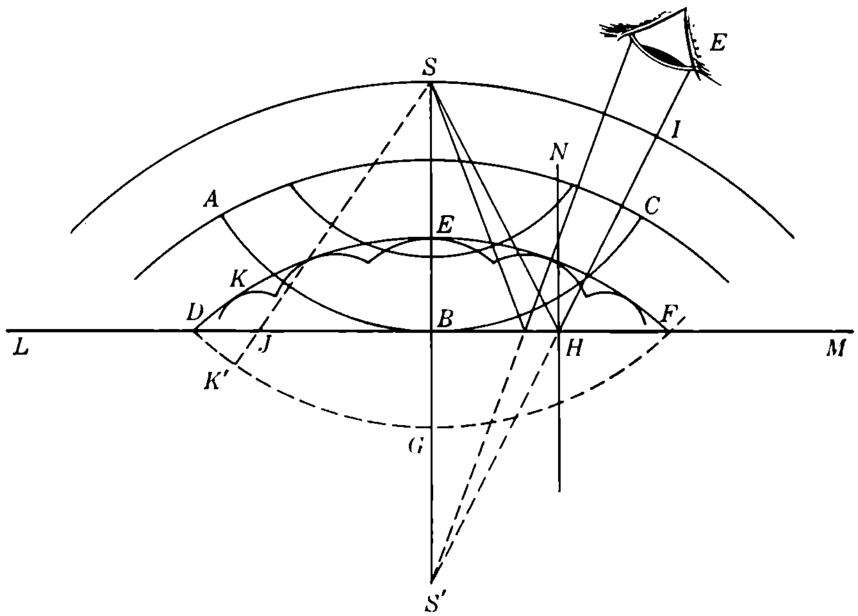


FIG. 40.6. Reflection of spherical waves from a plane surface. S' is a virtual focus.

the arrow directly, he sees the head on his left. Turning around and viewing the virtual image, he observes the head of the arrow to be on his right. This right-left inversion is called *perversion.* A plane mirror produces, then, a *perverted* but *erect* image. This nature of a mirror image is known to everyone who has tried to read the reflected image of a printed page.

Two plane mirrors at right angles to each other form a simple but instructive example of multiple images. The construction in Fig. 40.8 follows all the laws of reflection that we have discussed up to this point. It is seen that the pencil of rays from O is reflected back parallel to its original direction like a billiard ball bouncing in the corner of the table. In addition to the images I_1 and I_2 produced by a single reflection in each mirror an eye at E sees a third image I_3 by a double reflection. The image I_2 might be called a "virtual object" for this third image, and I_3

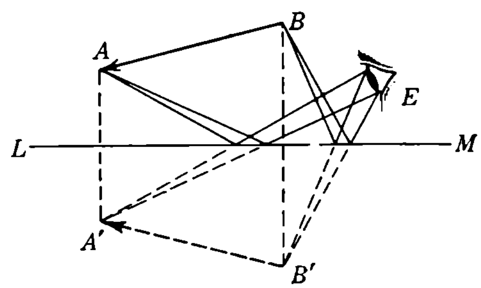


FIG. 40.7. Virtual image produced by a plane mirror.

lies as far behind the mirror AB as I_2 is in front of it. It is easily shown that O, I_1, I_2, and I_3 all lie on a circle about A as a center. If the two mirrors are inclined at 60°, five images are observable, two by single reflections, two by double reflections, and the fifth by a triple reflection.

When two mirrors are parallel and facing each other, a large number of images is observed because of multiple reflections, the limitation on the number being set by some loss of intensity of the light at each reflection. If a pencil of rays is directed into a three-sided corner made of mirrors, the reflected pencil is always parallel to the direction of the incident rays. This is the principle of the *corner reflector* used in radar.

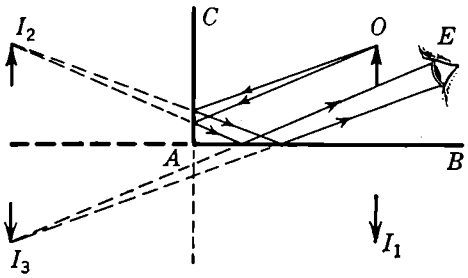


FIG. 40.8. Multiple images in two mirrors at right angles.

★40.7 Spherical Waves Refracted at a Plane Surface. Let spherical waves from the source S (Fig. 40.9) pass through a plane surface into a less dense medium. The distance of S below the surface is d, the velocity in the denser medium is v_2, while the greater velocity in the upper medium is v_1. As the wave front ABC passes through the surface, points successively farther from the vertex B become new sources of disturbance and the envelope of some of these

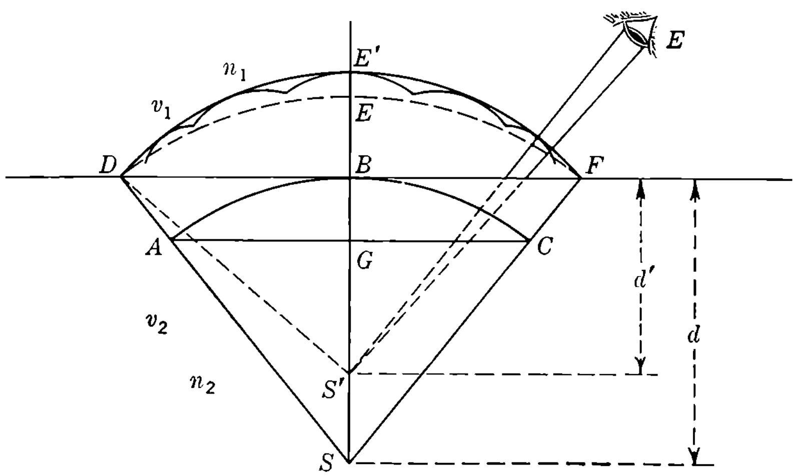


FIG. 40.9. Refraction of spherical wave at a plane surface.

wavelets an instant later will be, say, $DE'F$ rather than DEF, which would have been the wave front at this instant had there been no change in velocity. That is, while the disturbance at B travels the distance $BE' = v_1t$, that at A on the same wave front travels the shorter distance $AD = v_2t$ in the lower medium. To an eye at E the wave fronts seem to diverge from a center at S', which is therefore the virtual image of S. Let d' be the depth of S' below the surface. Before developing the relation between d and d' we must digress to establish an expression for the curvature of an arc.

The formula we need may be derived by reference to Fig. 40.10. The distance x measured along the bisecting radius of the arc AC between the arc and the chord AC is called the *sagitta*[1] of the arc. Now

$$r^2 = y^2 + (r - x)^2 = y^2 + r^2 - 2rx + x^2$$

or

$$2rx - x^2 = y^2$$

If x is small compared with r or y (or $\cos\theta$ may be considered equal to unity), we may neglect the x^2 term. Therefore within the limits of error in the measurements we may often write

$$x = \frac{y^2}{2r} \qquad (40.3)$$

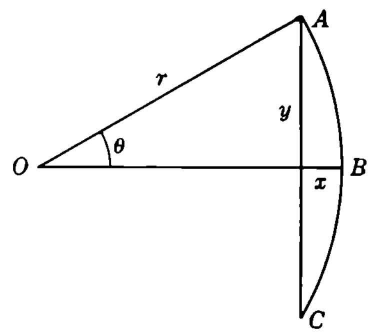


FIG. 40.10. Sagitta x of arc AC equals $y^2/2r$ if θ is small.

We shall have occasion to use this relation in discussing spherical mirrors and lenses in the following two chapters.

To return to our problem of the change in curvature of the wave fronts upon refraction (Fig. 40.9), if the cone of light from S has a small apex angle, $AD = BG$ is the sagitta of the wave in the denser medium, BE' is the sagitta of the wave in the less dense medium, and $d = AS$ and $d' = DS'$ are, respectively, their radii of curvature. Therefore

$$DA = y^2/2d = v_2 t$$

$$BE' = y^2/2d' = v_1 t$$

and thus

$$\frac{d}{d'} = \frac{v_1}{v_2} = \frac{n_2}{n_1} \qquad (40.4)$$

If an eye looking vertically downward views an object below the surface in the refracting medium, the portion of the wave front actually used is very small indeed. The pupil of the eye is on the average about 3 mm in diameter, and hence DF in Fig. 40.9 is less than 3 mm. This is justification for the method used in developing Eq. (40.4) for the ratio of the object distance to the image distance. If the lower refracting medium is water ($n_2 = 1.33$) and the other medium is air ($n_1 = 1$), the apparent depth d' of an object below the surface is three-fourths of its actual depth.

When the cone of rays from the source S is wide, the wave fronts in the upper medium are no longer spherical. Such a deviation from spherical shape is known as *spherical aberration.* Whenever this departure from spherical shape exists—and it is a common fault adversely affecting good image formation in optical instruments—it is more convenient to use rays in our discussions. The large aberration in the present example is illustrated in Fig. 40.11. The external rays

[1]Latin, arrow.

projected back into the refracting medium are all tangent to the *caustic curve* $SS'Y$, and the virtual image always appears at some point along this curve. If the eye is at E, the image of S at S' has large lateral as well as upward displacement.

For the case of a slab of transparent material of index n and with plane-parallel sides a distance d apart, we may use Eq. (40.4) in the form

$$d - d' = \frac{n-1}{n} d$$

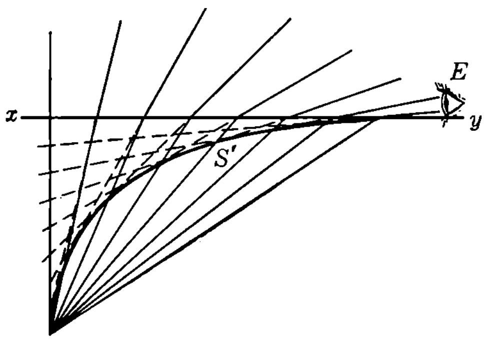


Fig. 40.11. Spherical waves refracted at plane surface xy are no longer spherical. Eye at E sees image of S at S'.

and obtain n from a direct measurement of $d - d'$ with a microscope. The microscope is focused on a mark on the table, the slab is then interposed, and the exact amount the microscope must be raised to bring the virtual image of the mark into focus is determined.

40.8. Refraction of Light through a Prism. A prism is a portion of a refracting medium bounded by two plane surfaces that are not parallel. This forms one of the most useful pieces of optical equipment. If a source S is viewed through a prism, as in Fig. 40.12, the curvature of the waves striking the eye is such that they seem to come from S', the intersection of the pair of rays shown. With almost all light sources there are color effects in this image, but for the moment we shall assume that the light is *monochromatic, i.e.*, but one color or wavelength is present. The angle A is the *refracting angle*, and the line along which the two refracting surfaces meet is called the *edge* of the prism. We shall consider all the rays to be perpendicular to this edge.

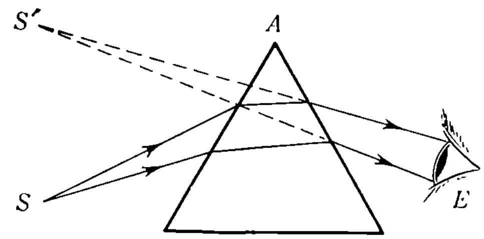


Fig. 40.12. Source of light as viewed through a prism.

The total deviation D (Fig. 40.13) of a ray in passing through a prism of known angle A may be obtained as follows: We have

$$D = d_1 + d_2$$

$$d_1 = i_1 - r_1 \qquad d_2 = i_2 - r_2$$

$$E + A = 180° = E + r_1 + r_2$$

and hence $\quad A = r_1 + r_2$

Therefore $\quad D = i_1 + i_2 - (r_1 + r_2) = i_1 + i_2 - A$

We consider that the prism material has an index of refraction n and that it is mounted in air of index unity. Now if the ray passes through the prism in the opposite direction, being incident on the right face at an angle i_2, the same deviation D results. The angle D varies continuously as the angle of incidence is changed from i_1 to i_2. Therefore, it may be seen that when D is plotted against i, D must go through either a maximum or a minimum when $i_1 = i_2$. *Experiment shows that then D is a minimum,* with the incident and emergent rays symmetrical with respect to the angle A of the prism.

For this angle D of *minimum deviation*, from the last equation,

$$i_1 = i_2 = \frac{A + D}{2} \qquad \text{and} \qquad r_1 = r_2 = \frac{A}{2}$$

Therefore

$$n = \frac{\sin i}{\sin r} = \frac{\sin \frac{1}{2}(A + D)}{\sin \frac{1}{2}A} \tag{40.5}$$

Thus the index of refraction of any transparent substance in the form of a prism may be obtained by measurement of the refracting angle of the prism and the angle of minimum deviation for the color or wavelength of the light being used. Since these angles may be measured with considerable precision with a spectrometer (Sec. 43.10), this forms a most accurate method of determining n.

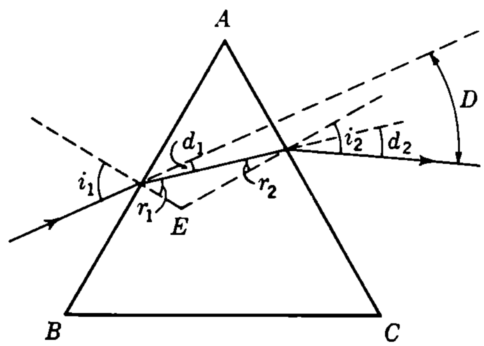


FIG. 40.13. Deviation of a ray by a prism.

40.9. Dispersion. The index of refraction of a substance depends upon the wavelength or color of the light. Therefore a given prism will give a different angle D of minimum deviation for the light of each wavelength emitted by the source. If the source S in Fig. 40.12 is a narrow, illuminated slit parallel to the edge of the prism, the eye will observe a number of virtual images of the slit in a row at S', one image for each color emitted by the source. These diverging rays from the prism may be concentrated into real line images by means of a lens as illustrated in Fig. 40.14. These images constitute a *line spectrum*, and the separation into the various colors is called *dispersion*. If all wavelengths are present, as in sunlight or the light from an incandescent tungsten filament, then these images completely overlap into a continuous spectrum. In practice another lens is always inserted between the slit S and the prism to render the light plane-parallel before it enters the prism (cf. Sec. 43.10).

Newton in 1666, using sunlight, was the first to observe a prismatic spectrum. Wollaston in 1802 and then Fraunhofer in 1815 observed with a very narrow slit that the otherwise continuous solar spectrum contains numerous dark lines parallel to the slit. These lines, called *Fraunhofer lines*, are caused by characteristic absorption frequencies of elements in vapor form in the outer, cooler layers of the sun's atmosphere. We shall discuss their production in Chaps. 43 and 49. The reason for mentioning them here is that in quantitative measurements of the dispersion of materials the index of refraction is always given for some of the prominent Fraunhofer wavelengths. Precisely these same wavelengths can be produced in emission if these elements are introduced into laboratory light sources. The wavelengths usually employed for dispersion data, Fraunhofer's letter designations, and the chemical elements responsible are given in Table 40.1.

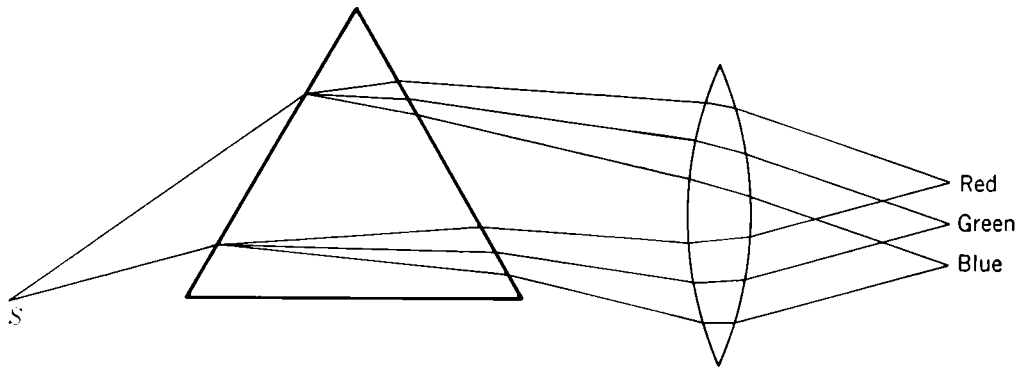


FIG. 40.14. Production of a line spectrum of a narrow slit S illuminated by light of several colors.

As indicated in Fig. 40.14, blue light is deviated through a larger angle than is red light. Thus the shorter, blue wavelengths undergo a greater decrease in velocity in the prism material, and hence for them the index of refraction has

Table 40.1. *Prominent Fraunhofer Spectral Lines in the Visible Spectrum.*

Designation	Color	Element	Wavelength, Angstrom units
A	Red	Oxygen	7,594
C	Red	Hydrogen	6,563
D	Yellow	Sodium	5,893
F	Blue	Hydrogen	4,861
G	Violet	Calcium	4,308
H	Ultraviolet	Calcium	3,968

larger values. Figure 40.15 is a plot of the variation of index with wavelength for three kinds of glass as well as for quartz and fluorite.[1] For all these optical materials the index increases with decreasing wavelength, but it is to be noted that the slope of each of these curves is different. Such curves are called "dis-

[1]It is not always the case that substances of higher density ρ have larger indices of refraction n. Note the following comparisons:

persion curves," and those plotted here are all cases of *normal dispersion*. As an exact measure of the dispersion of a substance the value of $dn/d\lambda$, or the slope of the dispersion curve at a given λ, is used.

If the refracting angle of the prism is small, the angle of minimum deviation D will also be small; hence the sines of the angles in Eq. (40.5) may be replaced by the angles themselves.

Therefore
$$n = \frac{A + D}{A}$$
or
$$D = (n - 1)A \tag{40.6}$$

If D_F and D_C designate the deviations for light of the wavelengths of the F and C lines, respectively, and D_D is the deviation for the sodium D line of intermediate

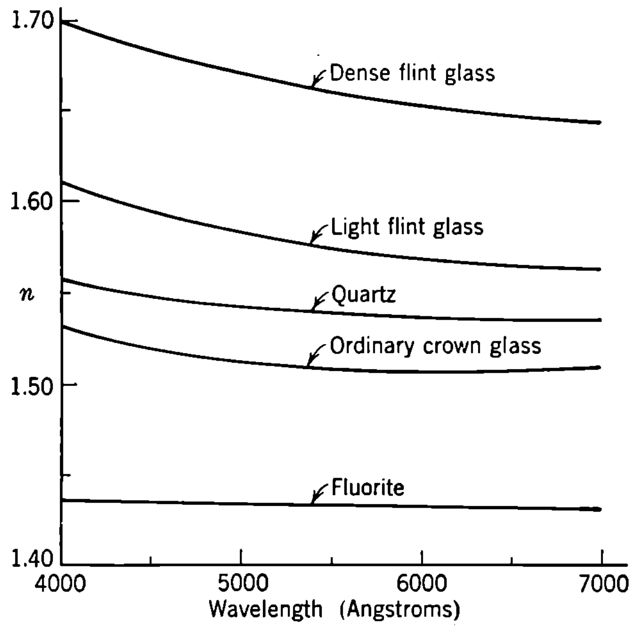


FIG. 40.15. Variation of index of refraction with wavelength.

wavelengths, $D_F - D_C$ is the *angular dispersion* of these particular red and blue colors, while D_D is a mean deviation. The *dispersive power* d of a substance is defined by the relation

$$d = \frac{D_F - D_C}{D_D}$$

	ρ	n
Ethyl alcohol	0.8	1.36
Water	1.0	1.33
Fluorite (CaF_2)	3.18	1.43
Rock salt (NaCl)	2.18	1.54

It has become customary, however, to refer to a substance of large index of refraction (*i.e.*, an *optically* dense substance) simply as a dense substance.

For a prism of small angle this becomes

$$d = \frac{(n_F - 1)A - (n_C - 1)A}{(n_D - 1)A} = \frac{n_F - n_C}{n_D - 1} \tag{40.7}$$

In Table 40.2 we list the values of the *refractive power* and the dispersive power for a number of transparent materials. It is evident that there are very considerable differences between the dispersive powers of various substances.

Table 40.2. Refractive Indices n and Dispersive Powers d for Some Substances

	n_D	n_F	n_C	$n_F - n_C$	d
Water	1.3330	......	......	0.0060	0.0180
Crown glass	1.5171	1.5232	1.5146	0.0087	0.0168
Light flint glass	1.5804	1.5903	1.5764	0.0139	0.0240
Dense flint glass	1.6555	1.6691	1.6501	0.0190	0.0290
Fused quartz	1.4584	1.4644	1.4564	0.0080	0.0175
Carbon disulfide	1.6303	......	......	0.0345	0.0547
Air (0°C, 76 cm)	1.000243	......	......	0.000003	0.0121

★40.10. Achromatic and Direct-vision Prisms. If two prisms having the same refracting angle, one of crown glass and the other of dense flint glass, are used in the same apparatus to project a spectrum on a screen it will be found that (1) the mean deviation is greater with the flint glass because of its higher index n_D and (2) the total length of the spectrum is considerably greater for the flint-glass prism. Suppose both prisms have an angle of 10°. We compute the angular dispersion for each from the relation $D_F - D_C = d(n_D - 1)A$, using the data in Table 40.2.

For the crown glass,

$$D_F - D_C = 0.0168(0.517) \times 10° = 0.087°$$

For the dense flint glass, similarly,

$$D_F - D_C = 0.0290(0.6555) \times 10° = 0.190°$$

We see, then, that the length of the spectrum is over twice as great for the flint-glass prism as for the crown-glass prism.

It is thus possible to choose a combination of two different kinds of prism so that there will be very little dispersion but an appreciable net deviation of the light. This is called an *achromatic combination* (greek *a*, not; *chroma*, color) of two prisms. The spectra from two different prisms are not necessarily just copies of one another on different scales. Therefore the choice of wavelengths does make a slight difference in the calculated refracting angles. For visual purposes achromatic combinations are usually calculated for the C and F wavelengths.

Worked Examples. a. Using the constants in Table 40.2, find the angle of a dense flint-glass prism that produces the same length spectrum between the C and F wavelengths as a 10° crown-glass prism.

For the crown glass, $D_F - D_C = 0.087°$, as already computed. Therefore, for the flint glass, $D_F - D_C = 0.0290(0.6555)A = 0.0190A$.

Hence $$0.0190A = 0.087° \quad \text{or} \quad A = 4.58°$$

If these two prisms are placed together as in Fig. 40.16, the dispersion of one prism will counteract the dispersion of the other. That is, the red and the blue rays will proceed in the same direction, which, because of the much smaller

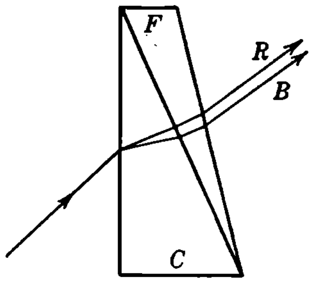


Fig. 40.16. Achromatic prism.

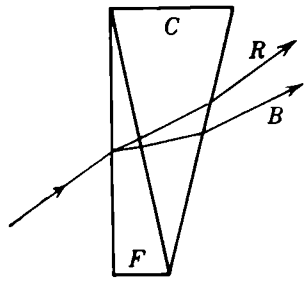


Fig. 40.17.—Direct-vision prism.

refractive angle of the more dispersive flint prism, will be a deviation in the same sense as that due to the crown prism.

To calculate this net deviation,

For C light, $(1.5146 - 1)10 - (1.6501 - 1)4.58 = 2.17°$
For F light, $(1.5233 - 1)10 - (1.6691 - 1)4.58 = 2.17°$
For D light, $(1.5171 - 1)10 - (1.6555 - 1)4.58 = 2.17°$

It is also possible to calculate for these same two glasses a combination of two prisms that will produce dispersion but with no net deviation of a mean wavelength, say that of the D line. Such a combination is called a *direct-vision* prism (Fig. 40.17).

b. Design such a direct-vision prism, given that the flint-prism part should have a refractive angle of 10°. Then, for the mean deviation D_D,

$$(n_D - 1)A \text{ for the crown} = (n_D - 1)10 \text{ for the flint}$$

Therefore $$A = \frac{0.6555 \times 10}{0.5171} = 12.7°$$

The two prisms must be combined with the edge of one next to the base of the other (Fig. 40.17), just as in the achromatic prism.

To compute the net dispersion, we take

$$(n_F - n_C)10° \text{ for the flint} - (n_F - n_C)12.7° \text{ for the crown}$$

Therefore, the angular dispersion is

$$0.190° - 0.0087 \times 12.7° = 0.080°$$

which is just about the same as that produced by a single 10° prism of crown glass.

PROBLEMS

1. A small mirror revolves about a vertical axis once per second. A beam of light is reflected from this mirror to a plane scale 10 ft away at the nearest point. What is the speed of the spot of light on the scale at this point? How does this speed vary with the angle between the reflected beam and the normal to the scale?

2. What is the shortest plane mirror in which a man can see a full-length image of himself?

3. Why does a plane mirror not produce inversion as well as perversion?

4. Two parallel plane mirrors are 2 m apart. Calculate the distances from the mirror surfaces of the first three images in each of an object 50 cm from one of them.

5. Two plane mirrors are inclined at 60° to each other. Plot the position of each of the images of an object placed at a nonsymmetrical position between them. Sketch the pencil of rays that makes three reflections before striking the eye of the observer.

6. A layer of ether of index 1.36 rests on water of index 1.33. A ray of light in the ether is incident on the surface of separation between the two liquids at an angle of incidence of 30°. Calculate the angle of refraction in the water and the angle of deviation.

7. If the ether layer in Prob. 6 is 3 cm deep and the water layer is 5 cm deep, what is the apparent distance of the bottom of the vessel below the top surface of the ether?

8. A ray of sodium light is incident at an angle of 30° on a slab of glass having parallel sides exactly 3 cm apart. Calculate the lateral displacement of the ray emerging from the other side of the slab if the index of this glass for the D line is 1.60.

9. Calculate the critical angle between water ($n = 1.33$) and carbon disulfide ($n = 1.67$).

10. A hollow prism of 60° refracting angle, made of thin glass plates having parallel sides, is filled with carbon disulfide ($n = 1.63$). Calculate the angle of minimum deviation and the angular deviation between the F and C lines.

11. A glass hemisphere of index 1.65 has a drop of liquid on its horizontal flat surface. For light passing radially in as shown in Fig. 40.18 the critical angle for total reflection is found to be 59.0°. What is the index of the liquid?

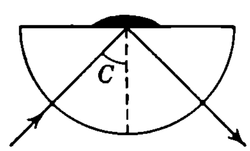


FIG. 40.18.—Problem 11.

12. A microscope is focused on a mark on the top surface of a slab of transparent plastic material exactly ½ in. thick. Then the microscope is focused on a mark on the lower surface of the slab, viewing through the material. From the first to the second setting the microscope is dropped 8.14 mm. Calculate the index of the slab.

★13. A crown-glass prism and a dense flint-glass prism are made with proper angles so as to form an achromatic combination for C and F light, giving a net deviation of 2°. Compute the refracting angles of the two prisms.

CHAPTER 41

REFLECTION AND REFRACTION AT SPHERICAL SURFACES

41.1. Ray Method. The optical parts of instruments usually have either plane or spherical surfaces. A spherical wave front is reflected as a spherical wave from a plane surface, but the curvature of the wave is changed by reflection from a spherical surface and is usually made non-spherical. Only if a mirror has dimensions small compared with its radius of curvature are the reflected wave fronts approximately spherical and therefore the resulting images fairly sharp. Similarly, plane or spherical wave fronts are never strictly spherical after refraction at a spherical surface.

This complication makes it difficult to trace the course of the light passing through optical instruments by analyzing the shapes of the wave surfaces. It is therefore more convenient to use the ray method. In our treatment of image formation we need concern ourselves, then, only with geometrical relations, using the following four experimental facts:

1. Light travels in straight lines in a homogeneous medium.
2. Two independent beams of light may intersect each other and continue on as independent beams.
3. Upon reflection the angle of incidence equals the angle of reflection.
4. Snell's law of refraction.

41.2. Reflection by a Concave Mirror. A point source S is at a distance p from a concave mirror LM (Fig. 41.1) having a radius R of curvature centered at C. A ray SL is reflected at L so that SL and LS' make equal angles with the normal CL. The ray SV incident normally on the vertex V (the center of symmetry) of the mirror is reflected back on itself. These two reflected rays intersect on the axis SV at the image S', a distance q from the mirror. We assume that the angular aperture of the mirror is small, so that the angles ϕ, ϕ', and ϕ'' are small and *all points* on the mirror may be taken to be at the *same* distances p and q from S and S', respectively. Such rays making but a very small angle with the *principal axis* are called *paraxial rays*. By principal axis of an

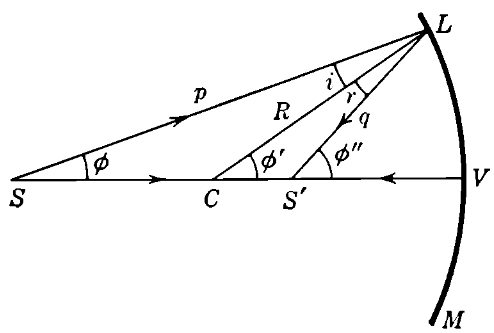


FIG. 41.1. Reflection from concave mirror.

optical surface is meant its axis of symmetry, SV in this case. The angles in Fig. 41.1 are exaggerated for the sake of clarity. From the law of sines,

$$\frac{SC}{\sin i} = \frac{p - R}{\sin i} = \frac{p}{\sin \phi'}$$

Also,

$$\frac{CS'}{\sin r} = \frac{q}{\sin \phi'}$$

Therefore, since $\angle i = \angle r$, division of the first of these equations by the second gives

$$\frac{SC}{CS'} = \frac{p}{q}$$

or

$$\frac{p - R}{R - q} = \frac{p}{q}$$

Hence $qR + pR = 2pq$, and, dividing through by pqR,

$$\boxed{\frac{1}{p} + \frac{1}{q} = \frac{2}{R}} \tag{41.1}$$

This equation holds, subject to the paraxial-ray limitation, for any point L and therefore for all rays from the object that are reflected from the mirror. Consequently there is a "focusing" action; all rays from a point source will be focused to a common image point.

The correctness of this formula may be verified by experimenting with any concave mirror "stopped down" to small aperture. If the source is at S', the image will occur at S. Any such pair of points is therefore referred to as a pair of *conjugate foci* and the distances p and q as *conjugate focal distances.* If the source is at a great distance from the mirror, p is practically infinite, the corresponding value of q is called the *principal focal length* f, and S' is the *principal focus* of the mirror. Equation (41.1) then becomes

$$\frac{1}{f} + \frac{1}{\infty} = \frac{2}{R} \qquad \text{or} \qquad \boxed{f = \frac{R}{2}} \tag{41.2}$$

showing that the principal focus of a concave mirror is midway between its vertex and its center of curvature.

Hence we may write the conjugate focal relation in the form

$$\boxed{\frac{1}{p} + \frac{1}{q} = \frac{2}{R} = \frac{1}{f}} \qquad \text{MIRROR EQUATION} \tag{41.3}$$

If $p > R$, $q < R$; the image is between C and the mirror.

If $p < R$ but $> f$, $q > R$; the image is farther than C from the mirror.

If $p = f$, $q = \infty$; the reflected rays are parallel.

If $p < f$, q is negative; the reflected rays appear to come from a point S' *behind* the mirror (Fig. 41.2). This is called a *virtual focus.*

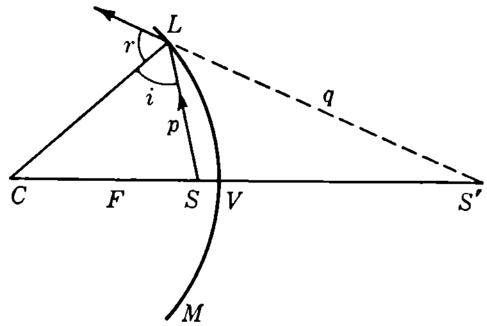


FIG. 41.2. Virtual image with concave mirror.

In using the equations for conjugate focal lengths, we must adopt the following convention for signs: *The quantities p, q, and R are each positive if they refer to a distance measured in front of the mirror but negative if the distance is measured behind the mirror.* This convention enables us to use our formulas with positive as well as negative signs for p and q.

If $f = \infty$, we have a plane mirror.

41.3. Convex Mirror. The formula for the conjugate focal relations for a convex mirror may be derived in similar fashion (Fig. 41.3). We assume as before that all rays are paraxial, and hence $SV = p$ and $VS' = q$. Since CL bisects the exterior angle at L in the triangle SLS',

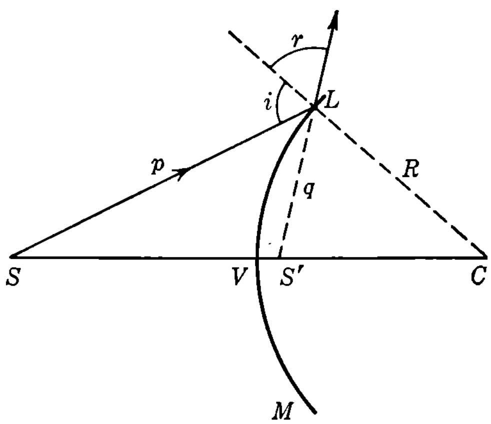


FIG. 41.3. Reflection from convex mirror. The rays have been drawn far from paraxial for the sake of clarity. The figure therefore, while a useful illustration, gives little evidence for the correctness of the statements in the text.

$$\frac{SC}{S'C} = \frac{SL}{S'L}$$

$$\frac{p + R}{R - q} = \frac{p}{q}$$

$$Rq - pR = -2pq$$

and, dividing through by pqR,

$$\frac{1}{p} - \frac{1}{q} = -\frac{2}{R} = -\frac{1}{f} \qquad (41.4)$$

It is clear that with the above convention of signs, Eq. (41.4) is identical with Eq. (41.3). For the convex mirror Eq. (41.3) predicts a negative q since f is negative and p is positive for a real object. That is, *only virtual images are formed of real sources by a convex mirror.* The incident light could be convergent onto a point a distance $-p$ behind the mirror, however, so that q might become positive for such a *virtual object.*

41.4. Images Formed by Spherical Mirrors. The mirror formulas hold for object and image points slightly off the axis. Any two rays drawn from a point of the source will, after reflection, intersect at the corresponding point of the image. This fact is used in the *graphical construction* of images. In practice any two of the following four rays are used for that purpose since their course is most easily drawn.

1. The ray directed through or toward the center of curvature and reflected back over the same path.
2. The ray directed at the vertex of the mirror and reflected with $\angle i = \angle r$.
3. The ray through the principal focus and reflected parallel to the axis.
4. The ray parallel to the axis and reflected through the principal focus.

Figure 41.4 shows how the first two rays are used in the graphical construction of an image.

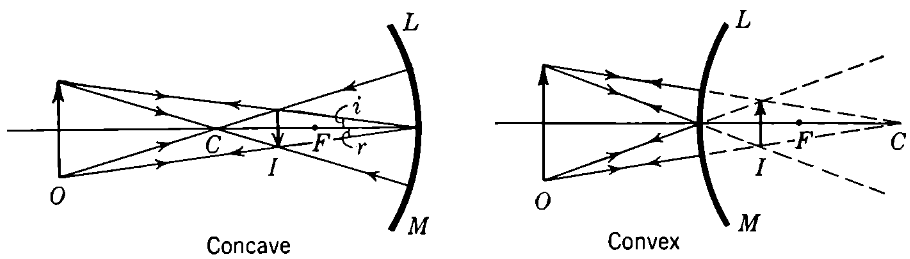


FIG. 41.4. Image formation by spherical mirrors.

The real image formed by the concave mirror is inverted, whereas the virtual image formed by the convex mirror is erect. As to the relative sizes of the object and image, it is clear from the constructions of Fig. 41.4 that, since the two subtend equal angles at the mirror, their sizes are to each other as their respective distances from the mirror. Therefore

$$\frac{\text{Size of } I}{\text{Size of } O} = \text{magnification } M = \frac{q}{p} \qquad (41.5)$$

The magnification may be either greater or less than 1; for the two cases illustrated in Fig. 41.4 the images are smaller than the object.

Worked Example. An object 3 cm in front of a spherical mirror produces a virtual, erect inage three times larger than the object. Find the focal length of the mirror. The mirror must be concave, for the virtual image formed by a convex mirror is always of reduced size (Fig. 41.4). Therefore q is -9 cm, and substituting in Eq. (41.3) we have

$$\frac{1}{3} - \frac{1}{9} = \frac{1}{f} \qquad \text{or} \qquad f = +\,4.5 \text{ cm}$$

***41.5. Spherical Aberration with Mirrors of Large Aperture.** The only case where the reflected wave from a spherical mirror is itself perfectly spherical is the special one of the source being at the center of curvature of the mirror. In general those incident rays parallel to the axis and close to the vertex V of the mirror are reflected through F, but rays farther from the axis are reflected so as to cross the axis closer to the mirror, as illustrated in Fig. 41.5. The reflected rays are tangent to a *caustic curve* having a sharp *cusp* at F. This lack of a sharp focus, indicating that the reflected wave is not spherical, is called *spherical aberration*.

If light is incident obliquely on a concave mirror, the reflected rays do not form a sharp focus even if the aperture is small. The reflected rays are concentrated into a pair of broadened focal lines, the light in this focal region forming a sort of lemniscate, or "figure 8." To explain this aberration, akin to *astigmatism*

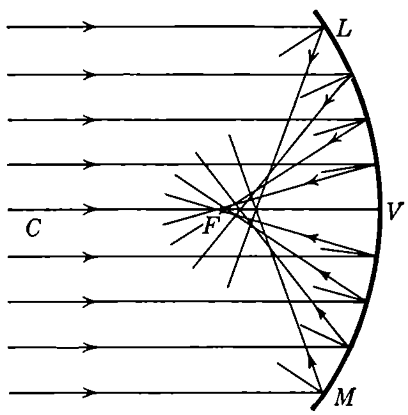


FIG. 41.5. Spherical aberration with mirror of large aperture.

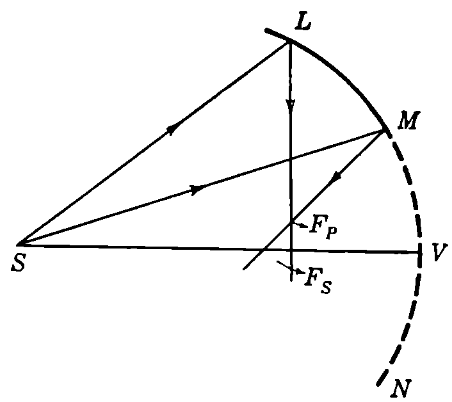


FIG. 41.6. Light incident obliquely on a mirror is concentrated into focal lines.

in lenses (Sec. 42.9), consider as in Fig. 41.6 that the mirror is a portion of a large mirror LN on the axis of which is the point source S. It is clear that the rays from L and M will intersect in the manner shown. The diagram is, of course, a sectional view. If the figure is revolved slightly about the line SV, it is seen that the intersection F_P is actually a line perpendicular to the paper. This is called the *primary focal line*. The *secondary focal line* F_S is a patch of light slightly elongated in the direction SV.

If the mirror is a paraboloid of revolution, incoming rays parallel to the axis are all concentrated at the focus, as a consequence of a geometrical property of the parabola. There is thus a complete lack of spherical aberration, no matter how large the aperture of the parabolic mirror. If the source of light is placed at the focus, the reflected rays all issue parallel to the axis. Automobile headlight and searchlight mirrors are therefore always paraboloidal, as are the large mirrors for astronomical telescopes.

41.6. Refraction at a Spherical Surface. Let us now consider refraction at a single spherical surface separating two media. The object space, of index n, is on the left of the surface (Fig. 41.7), and the image space, of index $n'(n' > n)$ is on the right. *Consider the object distance* *p positive when the object lies to the left of the vertex V of the refracting surface and the image distance w positive when a real image is formed to the right of the vertex. The radius of curvature of the surface will be taken as positive when the surface is convex toward the object,* as in Fig. 41.7, *but negative if it is concave toward the object.* A ray from the axial point O_1 is refracted at the point L of the surface and proceeds to the axial image point I_1. We employ the subscripts 1 because later (Sec. 42.2) we shall combine the equation governing the refraction at this surface with a similar equation

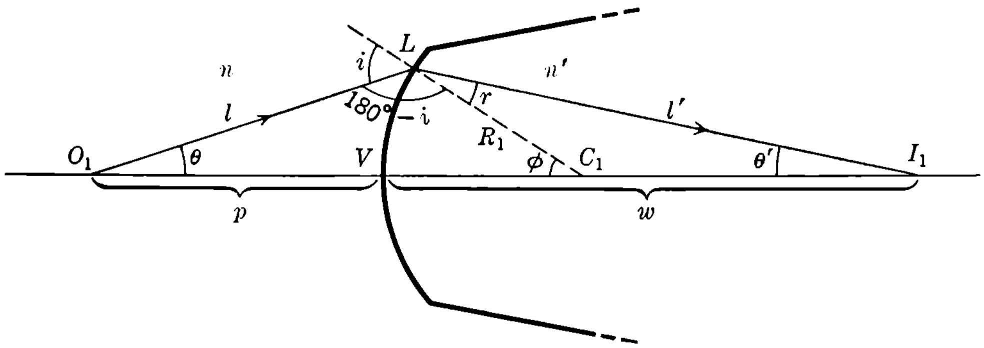


FIG. 41.7. Refraction at a single spherical surface.

for a second surface to obtain the lens formula. We reserve the letter q to denote the image distance after the two refractions.

In the triangle LC_1I_1,

$$\frac{I_1C_1}{\sin r} = \frac{w - R_1}{\sin r} = \frac{l'}{\sin(\pi - \phi)} = \frac{l'}{\sin \phi}$$

and, in the triangle O_1LC_1,

$$\frac{O_1C_1}{\sin(\pi - i)} = \frac{p + R_1}{\sin i} = \frac{l}{\sin \phi}$$

Dividing the first of these relations by the second, we have

$$\frac{w - R_1}{p + R_1}\frac{\sin i}{\sin r} = \frac{l'}{l}$$

According to Snell's law, $(\sin i/\sin r) = n'/n$, and if we assume that the rays are paraxial (θ and θ' are small), there results (since $l' \rightarrow w$ and $l \rightarrow p$)

$$\frac{w - R_1}{p + R_1} = \frac{nw}{n'p}$$

or

$$n'pw - n'pR_1 = nwp + nwR_1$$

Dividing through by pwR_1,

$$\frac{n'}{R_1} - \frac{n'}{w} = \frac{n}{R_1} + \frac{n}{p}$$

$$\boxed{\frac{n}{p} + \frac{n'}{w} = \frac{n' - n}{R_1}} \tag{41.6}$$

This equation for the refraction of paraxial rays at a spherical boundary is one of the most important formulas in geometrical optics. Had we chosen a concave surface for which R_1 is negative, Eq. (41.6) would have resulted, but with a change of sign of the right-hand side.

41.7. Focal Points and Image Formation. When the source O_1 is at infinity, the axial image point I_1 is by definition the principal focus F' in the image space. The focal distance f' of F' from the vertex V may be obtained from Eq. (41.6) by setting $p = \infty$ and $w = f'$. Thus

$$\frac{n}{\infty} + \frac{n'}{f'} = \frac{n' - n}{R_1}$$

or

$$f' = \frac{n'}{n' - n} R_1 \tag{41.7}$$

Similarly, if the source is at the principal focus F in the object space, the rays after refraction are parallel to the axis and the image I_1 is at infinity. Setting $p = f$ and $w = \infty$ in Eq. (41.6),

$$\frac{n}{f} + \frac{n'}{\infty} = \frac{n' - n}{R_1}$$

or

$$f = \frac{n}{n' - n} R_1 \tag{41.8}$$

The ratio of these two focal lengths for a single refracting surface is, then,

$$\frac{f'}{f} = \frac{n'}{n} \tag{41.9}$$

If, for example, the optical system consists of a glass rod of index 1.5 with a convex surface of any curvature at one end, the surrounding medium being air, the ratio of the focal lengths in the glass and in the air would be 1.5/1.

We have already seen that, if the direction of a ray is reversed, the ray retraces its path through the optical system. That is, if the source is placed at I_1, the image will be found at the original source point O_1.

If the locations of the two focal points for a refracting surface are known, the size and position of the image of a given object may be determined by a graphical method similar to that given in Sec. 41.4 for spherical mirrors. The method is illustrated in Fig. 41.8. The three rays drawn from the head of the arrow follow a known course. That proceeding parallel to the axis in the object space is refracted so as to go through the focal point F'. The ray directed along the normal to the surface through C is undeviated. The intersection of these two rays locates the head of the arrow in the image. Of course, the two rays might diverge after being refracted at the surface, in which case their intersection would occur back in the object space and the image would be virtual.

The third ray drawn through the focal point F is refracted so as to proceed parallel to the axis in the image space and must go through the

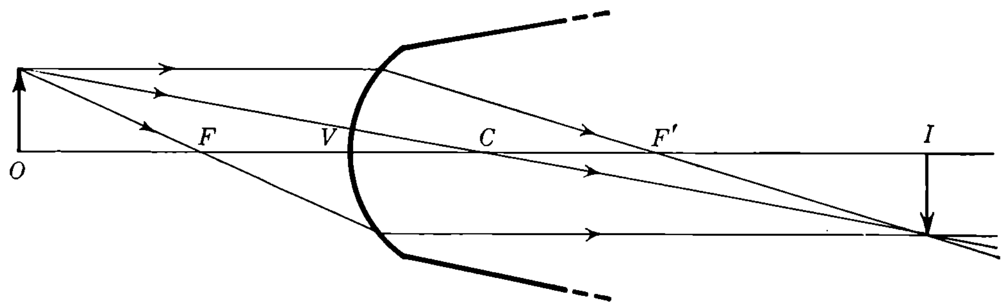


FIG. 41.8. Graphical method of image formation by spherical refracting surface.

same intersection point. In fact, the corresponding point in the image now having been located, the course of all other rays from the head of the arrow O incident on the refracting surface is determined; they are all refracted so as to pass through the head of the arrow I.

When the rays go successively through a number of refracting surfaces, as in a train of lenses of some optical instrument, the image formed by each surface is the object for the next surface in line. If the rays approaching one of the surfaces diverge either from a real image or from a virtual image, the distance from that image to the surface in question is a *positive* object distance. But if, as often is the case with lenses, the rays from the first surface are converging on approaching the second surface, then the point of convergence is farther along in the image space of the second surface and thus the distance measured back from this point to the second surface constitutes for it a *negative* object distance. Such a convergence point is a *virtual object* for the second surface.

★41.8. Worked Examples. *a.* The front convex surface of a crown-glass lens of index 1.5 has a radius of curvature of 30 mm. An axial point source is 90 mm in front of this lens surface. Assuming that the glass extends for some distance behind this convex surface, calculate the image distance w (1) if the lens is in air, (2) if the lens is in water of index 1.33.

1. We may use Eq. (41.6) with $n = 1$, $n' = 1.5$, $p = 90$ mm, $R_1 = 30$ mm.

$$\frac{1}{90} + \frac{1.5}{w} = \frac{1.5 - 1}{30}$$
$$w = +270 \text{ mm}$$

That is, if the glass extended farther than this from the convex surface, there would be a *real* image at this distance from the vertex.

2. With $n = 1.33$,

$$\frac{1.33}{90} + \frac{1.5}{w} = \frac{1.5 - 1.33}{30}$$
$$w = -165 \text{ mm}$$

Since w is negative, the rays in the lens as they leave the first surface are diverging from a *virtual* image in the water 165 mm from the vertex of this surface.

b. Locate the principal focus F' in the glass, assuming that the glass extends indefinitely beyond this convex surface.

If in air, $f' = (1.5 \times 30)/(1.5 - 1)$ mm $= 90$ mm from the vertex.

If in water, $f' = (1.5 \times 30)/(1.5 - 1.33)$ mm $= 265$ mm from the vertex.

***41.9. Other Refraction Phenomena.** There are a number of interesting effects due to atmospheric refraction. When the sun is near the horizon, its rays traverse air of increasing density while passing through to the earth's surface. The denser the air, the greater its refractive power. Therefore these rays are bent downward so that the sun is visible when it is actually below the horizon by an amount about equal to its diameter. Also, the rays from the lower edge of the sun will be bent more than those from the upper edge. This produces an apparent shortening of the vertical diameter, causing the sun (and the moon) to appear elliptical in shape when near the horizon. The horizontal diameter of the sun is the same at the horizon and near the zenith, appearances notwithstanding.

The twinkling of stars is produced by rapid local changes in the density of the atmosphere along the line of sight. This causes small, sudden shifts in the apparent direction and intensity of the incoming rays.

A mirage is another example of changes in the direction of pencils of light passing through air of varying density. It is often observed as an apparent pool of water on the road ahead by a motorist looking forward along a concrete pavement heated by the sun. This warms the adjacent layers of air and reduces the index of refraction. Figure 41.9 is a representation of how an observer at E might see the object OM directly by means of pencils such as OE and ME and an inverted image $O'M'$ by way of the refracted pencils OAE and MBE, which are deviated in this manner in the warmer air layers near the ground. Just the opposite type of mirage, known as "looming," is observed in sighting an object over water. The air layers adjacent to the cold water are the denser, with the result that light traveling from O to the observer's eye E is refracted slightly in the manner shown in Fig. 41.10.

Very similar effects have been observed with radar beams directed along the earth's surface. In the propagation of underwater sound the continuous variation of n from one ocean layer to another gives rise to most curious effects such as trapping of the waves between layers. Investigation of these matters is carried out under the Navy's "Sofar" researches.

The rainbow is caused by sunlight that has been refracted and internally reflected by raindrops. Sometimes two bows are seen, both being arcs of circles with centers on the extension of the line from the sun to the observer. The inner, or primary, bow, which is considerably the brighter of the two, is violet on

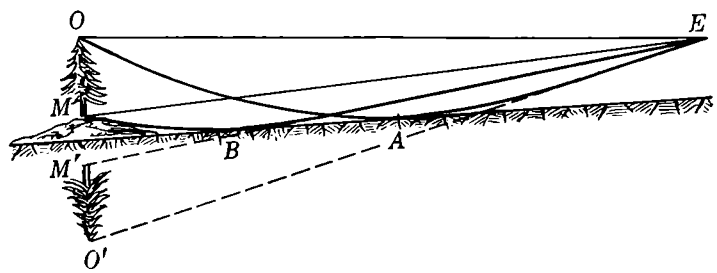


FIG. 41.9. A mirage.

the inside, red on the outside. In the secondary bow the order of the colors is reversed. The rays composing the primary bow are refracted and once internally reflected in the spherical drop in the manner depicted in Fig. 41.11. For all the parallel rays entering the upper half of the drop there is an angle of maximum deviation, which is 137°42′ for red light and 139°37′ for violet. All the rays emerging at nearly this angle are about parallel, and hence the intensities of all such rays total to an appreciable amount at the observer's eye at E. For rays emerging at other angles there is no such accumulation of light from drops at

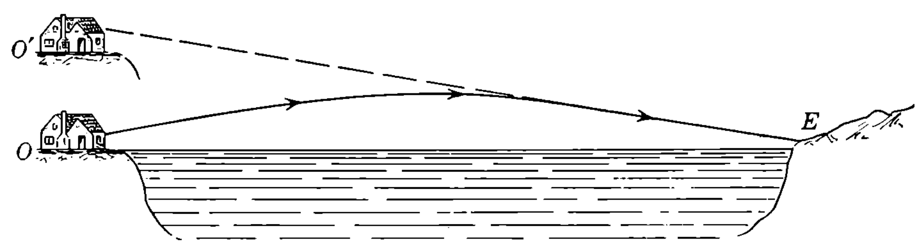


FIG. 41.10. Looming.

various distances. An observer at E looking away from the sun will receive these rays from all raindrops lying in an arc forming with the line from the sun extended an angle of 180° − 137°42′, or 42°18′, for the red light and 180° − 139°37′, or 40°23′, for the violet. Therefore this primary bow is violet in color on the lower side and red on the upper, the other colors falling in between.

The secondary rainbow is produced by rays incident on the lower half of the drop, refracted, twice internally reflected, and then refracted downward as shown in Fig. 41.12. This inverts the order of the colors as compared with the primary bow, the deviation being such that the observer, looking away from the sun, sees the red color in an arc subtending an angle of 50°34′ and the violet an angle of 54° with the line from the sun extended.

For an observer on the ground the sun must be not more than 42° above the horizon if a primary bow is to be visible. From an airplane, however, with the sun overhead and looking down on a cloud, a circular rainbow may be seen. An artificial rainbow may be made in the laboratory by sending a beam of plane-parallel light from an arc lamp in the proper direction into a hollow glass sphere filled with water. Halos sometimes observed at a 23° angle with the line to the sun or moon are in somewhat analogous fashion produced by refractions and

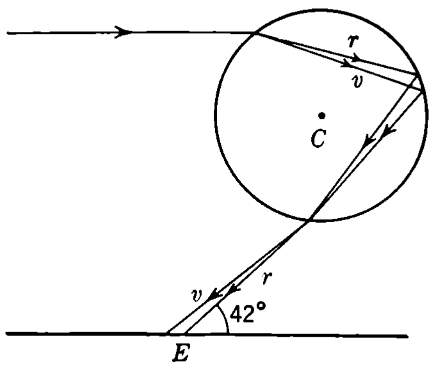


Fig. 41.11. Production of the primary rainbow.

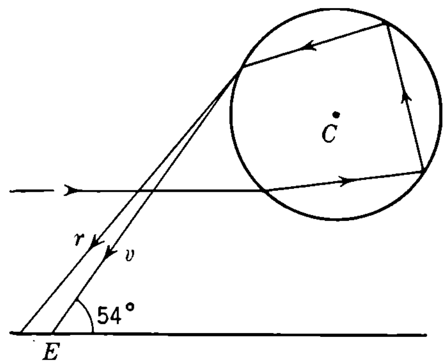


Fig. 41.12. Secondary rainbow production.

internal reflections in thin, parallel hexagonal ice crystals that form the high cirrus clouds.

PROBLEMS

1. How far from a concave mirror must an object be placed to make the image (either real or virtual) four times as large, if the focal length of the mirror is 25 cm?

2. A convex mirror A of focal length 10 cm and a concave mirror B of focal length 15 cm face each other at a distance of 25 cm. An object 5 cm long is placed at right angles to the common principal axis and at a distance of 15 cm from A. Calculate the position, size, and nature of the image formed by rays reflected first at A and then at B.

3. The sun has a diameter of 8.65×10^5 miles and is at a distance of 9.3×10^7 miles. What is the size of its image formed by a spherical concave mirror of 40 ft focal length?

4. A concave mirror has a radius of curvature of 20 cm. An object is placed in such positions that its real image is (*a*) half the size of the object; (*b*) twice its size. Find how far the object is from the mirror in each case.

5. In Prob. 4*a* the object moves through a small distance x toward the mirror. Through what distance and in which direction will the image move?

6. A long glass rod of index 1.5 has one end ground and polished to a convex spherical shape of radius of curvature 10 cm, with its vertex on the axis of the rod. A luminous arrow 2 mm long at right angles to the axis of the rod is placed 30 cm beyond the vertex on the extended axis of the rod. Calculate the position and the size of the image of this arrow formed by paraxial rays passing through the surface.

7. If the end surface of the rod of Prob. 6 is concave and of radius of curvature 10 cm, find the position and size of the image.

8. Solve Probs. 6 and 7 for the case where the source and rod are immersed in water of index 1.33.

9. The rays from the sun enter a spherical bowl of liquid along a diameter and are focused on the far side. Neglecting any refraction by the thin glass shell, what must be the index of refraction of the liquid?

10. Show that for paraxial rays the ratio of the size of the image I to that of the object O in Fig. 41.8, that is, the *magnification* M, is given by $M = nw/n'p$. If the source in Example (a) of Sec. 41.8 has a lateral dimension of 2 mm, find the corresponding length of the image in the glass.

11. A spherical bowl of 10 cm radius and filled with water contains a small object $2\frac{1}{2}$ cm from the side of the bowl. Calculate the position and magnification of the image of the object for paraxial rays from the object directed radially through (a) the near side, (b) the far side of the bowl.

12. A piece of glass of index 1.5 has one face plane, the other convex outward and of radius of curvature 10 cm. The thickness of the glass along the axis is 5 cm. A small object is placed outside the block on the axis 5 cm from the plane face. What is the apparent location of the object when viewed through the block?

CHAPTER 42

LENSES

42.1. Types of Lens. Lenses are pieces of transparent substance, usually bounded by spherical surfaces, for the purpose of image formation. There are two main types of single lens, those thicker in the middle than at the edges, thus retarding the center portions of incident plane waves so that they converge to a real image, and those thinner in the middle than at the edges, with the result that incident plane waves are retarded more at the edges and hence emerge diverging from a virtual image behind the lens. The common forms of single lens are shown in section in Fig. 42.1. Lenses of the first type are known as *converging*, or *positive*, lenses, while those in the second category are called *diverging*, or *negative*, lenses. The positive and negative designations arise from the fact that a converging lens brings plane waves to a real, positive focus in the image space, whereas a diverging lens causes such waves to diverge from a virtual, negative focus back in the object space. These effects on incident plane wave fronts are illustrated in Fig. 42.2.

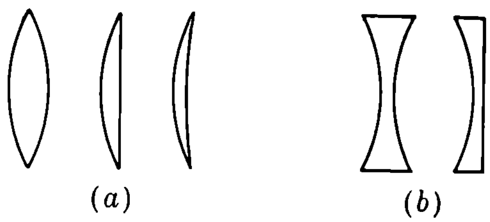


FIG. 42.1. (*a*) Double-convex, plano-convex, and miniscus converging lenses. (*b*) Double-concave and plano-concave diverging lenses.

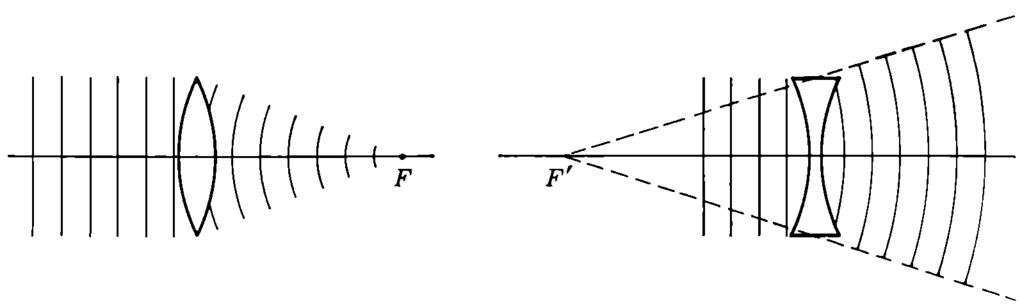


FIG. 42.2. Converging and diverging lens action on plane waves.

42.2. Thin Lens. We wish to find an expression for the focal length of a thin lens. Suppose that we have a meniscus converging lens of index n in air (Fig. 42.3). Equation (41.6) for the first surface becomes

$$\frac{1}{p} + \frac{n}{w} = \frac{n-1}{R_1} \tag{42.1}$$

Now I_1 is the object for the second surface, for which the object distance is, then, $-(w - t)$, *the negative sign being used because I_1 is a virtual*

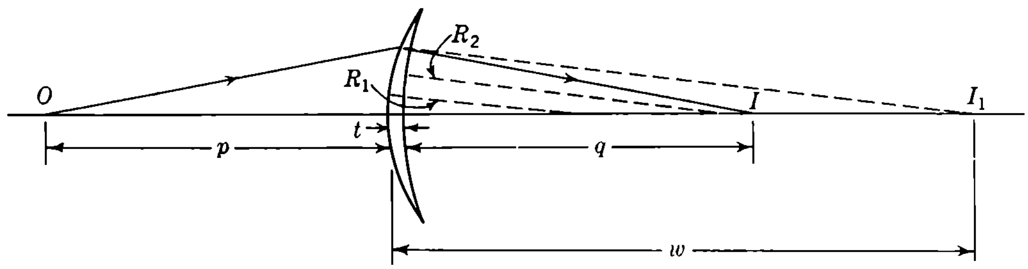


FIG. 42.3. Paraxial ray through miniscus lens.

object. If we may consider the thickness t of the lens to be negligible, Eq. (41.6) gives for the refraction at the second surface

$$\frac{n}{-w} + \frac{1}{q} = \frac{1 - n}{R_2} \tag{42.2}$$

Adding Eqs. (42.1) and (42.2), we obtain

$$\frac{1}{p} + \frac{1}{q} = (n - 1)\left(\frac{1}{R_1} - \frac{1}{R_2}\right) \tag{42.3}$$

If $p = \infty$, q is by definition equal to f, the focal length. Hence the right-hand side of Eq. (42.3) equals $1/f$, and we may write

$$\boxed{\frac{1}{p} + \frac{1}{q} = \frac{1}{f} = (n - 1)\left(\frac{1}{R_1} - \frac{1}{R_2}\right)} \quad \text{LENS EQUATION} \tag{42.4}$$

RULES FOR SIGNS IN THE LENS EQUATION

1. *The object distance p and image distance q are positive if disposed relative to the lens as they are for a converging lens forming a real image of a real object. The length p is negative for a virtual object (waves converging onto the lens), and q is negative for a virtual image (waves diverging from the lens in the image space).*

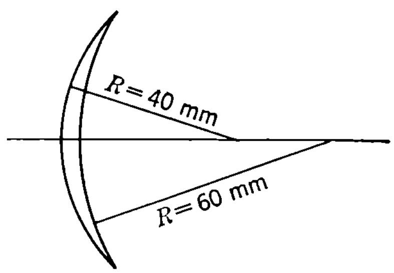


FIG. 42.4.

2. *The focal length is positive for a converging lens, negative for a diverging lens.*

3. *Each R is a positive quantity if it is the radius of curvature of a lens surface convex toward the incident rays, but negative if the lens surface is concave toward the incident rays.*

As an example, let us calculate the focal length of the meniscus lens in Fig. 42.4, given that its index is 1.5. If light is incident from the left,

$$\frac{1}{f} = (1.5 - 1)\left(\frac{1}{40} - \frac{1}{60}\right)$$
$$f = +240 \text{ mm}$$

If light is incident from the right,

$$\frac{1}{f} = (1.5 - 1)\left(\frac{1}{-60} - \frac{1}{-40}\right)$$
$$f = +240 \text{ mm}$$

A lens thus has the same focal length for light incident from either side, provided the medium on the two sides is the same. Application of Eq. (42.4) to a thin concave lens yields, as expected, a negative value for its focal length. We shall use this thin-lens equation further in Sec. 42.12 in the calculation of achromatic combinations.

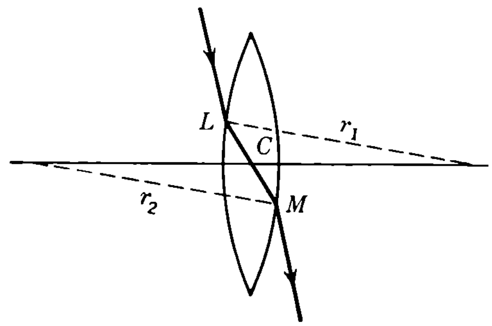


FIG. 42.5. Ray passing through the optical center of a thin lens is unchanged in direction.

To locate graphically images produced by thin lenses we make use of the fact that in any lens there is a point on the principal axis called the *optical center* which has the property that all rays passing through it are undeviated because they traverse a refracting medium with parallel sides, although they will suffer a slight lateral displacement. This is illustrated in Fig. 42.5, where, since r_1 and r_2 are parallel radii of curvature, the surfaces at L and M are necessarily parallel to each other. For a thin lens the optical center and geometrical center may be considered as coinciding. The graphical construction of images for two thin-lens

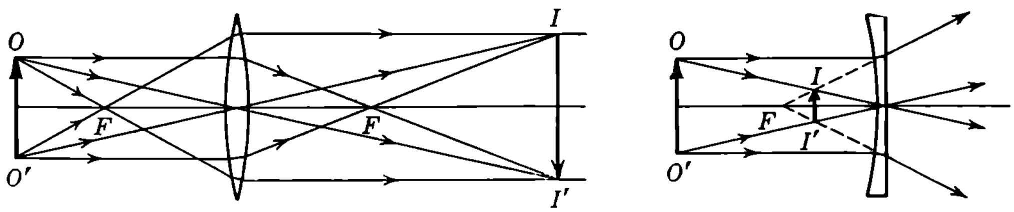


FIG. 42.6. Graphical construction of images with thin lenses.

situations is illustrated in Fig. 42.6. The rays from O and O' are undeviated in going through the center of the lens. Those rays passing through the principal focus F emerge on the other side of the lens, traveling parallel to the principal axis. For the converging lens those rays from O and O' drawn parallel to the principal axis are refracted through the principal focus on the far side of the lens. The intersections of these rays serve to locate the corresponding image points I and I'. The variation in construction for the diverging lens is evident from the drawing. It is also evident that all virtual images, including those with convex lenses when the object is less than the focal distance from the lens, are erect, while all real images are inverted.

Since image and object in Fig. 42.6 subtend equal angles from the lens center, their relative sizes are in the ratio of their respective distances from the lens. Therefore

$$\frac{\text{Size of image}}{\text{Size of subject}} = \text{magnification } M = \frac{q}{p} \tag{42.5}$$

The magnification M is negative if either the object or the image is virtual.

42.3. Thick Lens. If the lens thickness t cannot be neglected, Eq. (42.2) should be written

$$\frac{n}{-(w-t)} + \frac{1}{q} = \frac{1-n}{R_2} \tag{42.6}$$

Application of Eq. (42.1) to obtain w and then solution of Eq. (42.6) for the distance q of the final image from the second surface constitute a valid procedure for a lens of any thickness. Also, the ratio of the sizes of final image and object is just the product of the magnifications effected by the refraction at each surface.

Since for a thick lens the distances of the two focal points from the nearest lens surface in general are not equal, the question arises as to how the focal length of the lens is best specified and measured. Now to every ray originating at a point of the object and lying in the object space there is a corresponding ray in the image space that is a continuation of this ray in the object space. These two are *conjugate rays*. For every lens or optical system there are a pair of planes perpendicular to the principal axis, called *principal planes*, which have the property that any pair of conjugate rays will intersect these planes at equal distances from the axis. The intersections of the principal planes with the axis are the *principal points*, and the focal length is the distance from either focal point to the corresponding principal point. Thus the two focal lengths so defined are equal if the same medium exists on both sides of the lens.

The importance of principal planes may be judged from the statement that, once their position has been determined for a lens or optical system, they may be used in place of all lens surfaces for image calculations. If a simple lens may be considered as a thin lens, then the principal planes and the lens surfaces all coincide and we may measure the object and image distances from either surface.

42.4. Combination of Lenses. Two or more thin lenses are often used together, either in contact or with air gaps between them. Such a combination is called a "compound" lens. The location and nature of the final image produced by a compound lens may be calculated by successive applications of the thin-lens equation to each component lens.

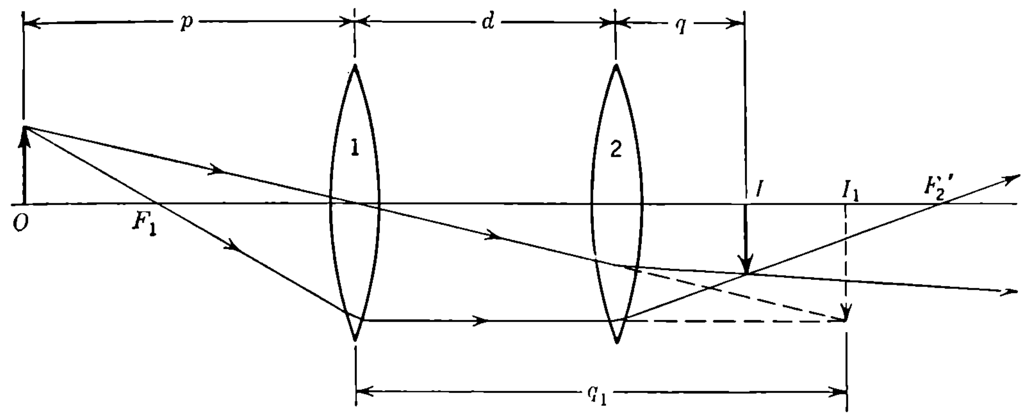


FIG. 42.7. Combination of two thin lenses.

Consider the combination of two thin lenses separated by a distance d, as shown in Fig. 42.7. For lens 1,

$$\frac{1}{p} + \frac{1}{q_1} = \frac{1}{f_1} \tag{42.7}$$

The object distance for the second lens is $-(q_1 - d)$. Hence, for lens 2,

$$\frac{1}{-(q_1 - d)} + \frac{1}{q} = \frac{1}{f_2} \tag{42.8}$$

which gives the distance q from lens 2 to the final image I. The over-all magnification is the product of the magnifications produced by the separate lenses, or

$$M = \frac{q_1}{p} \times \frac{q}{q_1 - d} \tag{42.9}$$

If the first is a converging lens and the second a diverging lens, the combination is called a *telephoto lens* (Prob. 14).

When the two lenses are in contact, $d = 0$ and addition of Eqs. (42.7) and (42.8) gives

$$\frac{1}{p} + \frac{1}{q} = \frac{1}{f_1} + \frac{1}{f_2} = \frac{1}{f} \tag{42.10}$$

The two lenses are therefore equivalent to a single lens of focal length f, and of course either f_1 or f_2 may be negative as well as positive.

42.5. Aberrations in Lenses. Throughout our discussion of lenses up to this point we have assumed that the rays make but a small angle with the axis. If this is not the case, the emergent wave fronts are no longer spherical, even though the lens surfaces are perfectly spherical, and therefore the rays from every point in the source will not be brought to a definite point image. When these rays, making such a large angle θ with the axis that one cannot approximate $\sin \theta$ by θ, are included, one or more of five types of *aberration*, or defect, may appear in the images. These aberrations are *spherical aberration*, the defect that we have already described for large mirrors; *astigmatism; coma; curvature of the image field;* and *distortion* of the image. To these must be added the defect known as *chromatic aberration*, arising from the fact that the index of refraction of the lens material varies with the wavelength or color of the light.

Spherical and chromatic aberration occur even in the case of point objects on the axis, while the other four occur only for point objects off the axis. The detailed consideration of lens aberrations and how to eliminate or minimize them becomes so complicated that it must be left to the experts in this field. It is important to describe these defects, however, and to give a few of the methods commonly used to reduce their magnitude.

One simple way of reducing all aberrations except chromatic aberration in any optical system is to "stop down" the aperture. This procedure eliminates the rays making large angles with the axis but at the same time decreases the light intensity in the image.

***42.6. Spherical Aberration.** When the ratio of aperture to focal length of a lens is relatively large, the refraction of rays from a point source on the axis produces an image as illustrated in Fig. 42.8. Paraxial rays in the image field cross the axis at I, which is called the *Gaussian image point.* The *rim rays*, *i.e.*, those which are refracted at the edge of the lens, however, cross the axis at K. The totality of rays produces as the image of the point source a *caustic curve* with point I as its apex. At J, which is known as the *circle of least confusion*, the diameter of the circular cross section of the entire bundle of rays has its minimum value. It can be reduced in size by stopping down the aperture with a diaphragm.

Spherical aberration may also be reduced by making the two surfaces of the lens of different curvature. As a simple example take the case of parallel rays from a distant source incident on a plano-convex lens of large aperture, first with the plane surface (Fig. 42.9*a*), then with the convex surface toward the light (Fig. 42.9*b*). On thinking of the edge of the lens as approximating a prism, it is clear that in (*b*) the rim ray is incident on the refracting surface at much more nearly the angle which will make its deviation D a minimum and that therefore the ray JK will be directed closer to the focus I than is the case in (*a*). In general, for the greatest reduction in spherical aberration the relative curvatures of the two lens surfaces should be such that the deviation is distributed equally between the two. Detailed calculation shows that the radius of curvature of the surface toward the light should be about one-seventh of that toward the image, but a plano-convex lens oriented as in Fig. 42.9*b* reduces the spherical aberration almost as much.

The first component lens in a microscope objective is commonly a hemisphere with the plane surface toward the object. When the space between the lens and the object is filled with an oil of the same refractive index as the lens, even though the angular aperture may be large, there is very little spherical aberration. The object is placed at a distance R/n from the center of the spherical surface of radius R, n being the index of the glass (Fig. 42.10). Applying Eq. (41.6) to

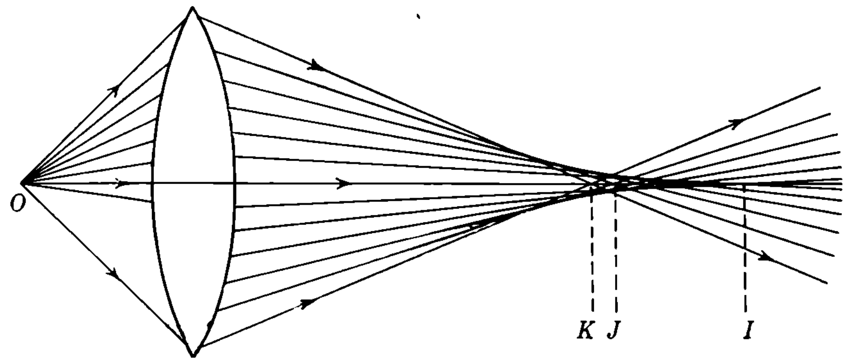


FIG. 42.8. Spherical aberration. I = Gaussian image point. J = circle of least confusion.

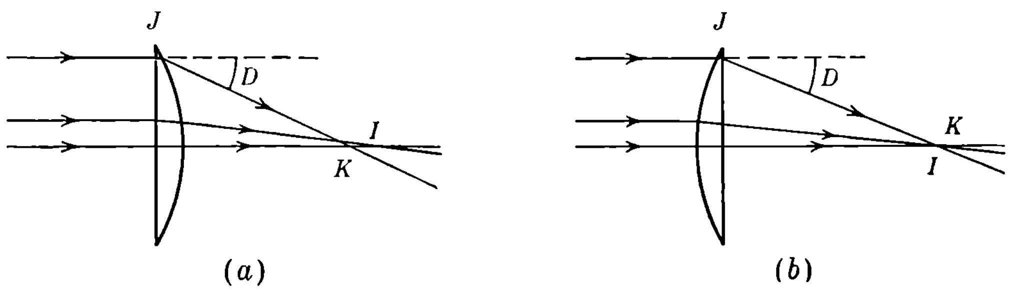


FIG. 42.9. There is less spherical aberration in (*b*) than in (*a*).

the refraction of a ray such as OA and remembering that R is negative for this surface,

$$\frac{n}{R + (R/n)} + \frac{1}{w} = \frac{1 - n}{-R} \tag{42.11}$$

Solution of this equation for w gives

$$w = -(R + Rn) \tag{42.12}$$

showing that the image I is virtual and at a distance Rn below the center of the hemisphere. In the triangle COA,

$$\frac{\sin \theta}{R} = \frac{\sin r}{R/n}$$

or

$$\frac{\sin \theta}{\sin r} = n = \frac{\sin i}{\sin r}$$

and hence $i = \theta$. In the triangle CIA, $\sin i/Rn = \sin \phi/R$

or

$$\sin i/\sin \phi = n$$

and hence $\phi = r$. Therefore all rays, not just paraxial rays, proceed after refraction as if they were coming from I. The two points O and I are called *aplanatic points* of this refracting surface.

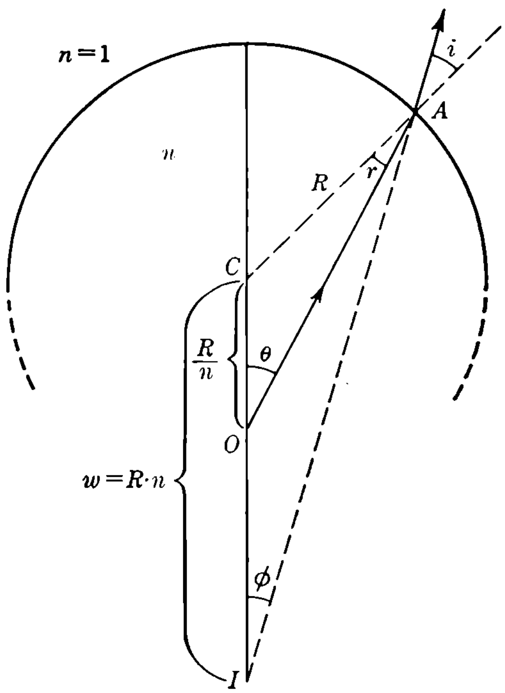


FIG. 42.10. Wide-angle hemispherical lens produces no spherical aberration.

★42.7. Astigmatism. If the point source is not on the axis, the change in curvature of the incident spherical wave front is not symmetrical even though spherical aberration is absent and the resulting image is *astigmatic* (Greek *a*, not; *stigma*, point; " not a point," *i.e.*, the image of a point source is *not a point* but a pair of lines). The cross sections of the wave fronts near the focus are as shown in Fig. 42.11. There are two line foci at right angles to each other, while between them there is a circular confusion of light, which is the nearest approach to a point image. In general the bundle of rays has an elliptical cross section. The line focus nearest to the lens is called the *primary focus*, the further one the *secondary focus*.

The *astigmatic difference*, the difference between the distances of I_p and I_s from the lens, increases rapidly with the obliquity of the incident rays. The difference $I_p - I_s$ is positive for a divergent lens and is somewhat larger but negative for a convergent lens. It is thus possible to combine the two into a compound lens in which the astigmatic differ-

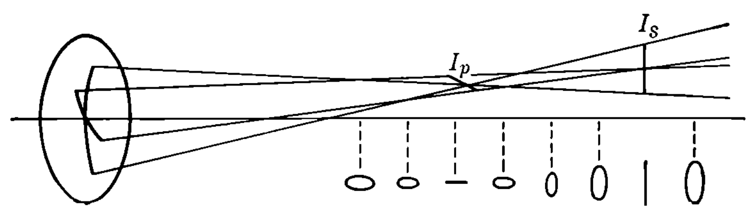


FIG. 42.11. The shapes of bundles of rays near the two astigmatic line images of a point object.

ences compensate for one another to some extent. In the photographic *anastigmat* (Chap. 43) both astigmatism and curvature of field are largely eliminated over a considerable area in the image plane.

★42.8. Coma. The image defect known as *coma* (Greek *komē*, hair) arises from the fact that for object points off the axis there is in general a variation of lateral magnification with the height h of the zone in the lens through which the rays pass. This is illustrated in Fig. 42.12. Furthermore the rays that pass through the lens at a distance h from the axis may meet in the focal plane in a circle of points, the size of the circle depending on h and other factors. Each zone of the lens produces one of these *comatic* circles, and the effect of the overlapping of all these circles is a pear-shaped image.

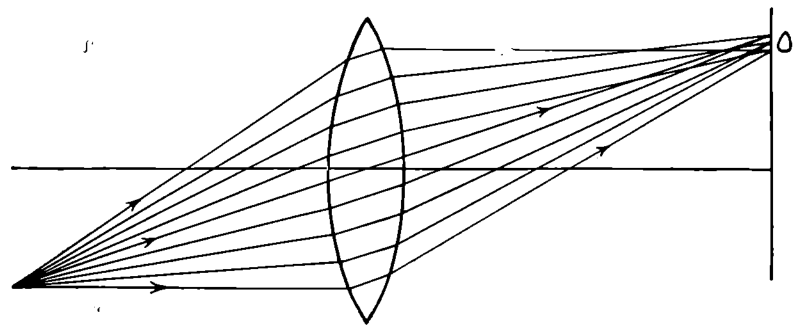


FIG. 42.12. Coma; pear-shaped image of non-axial object points.

To eliminate coma the lateral magnification must be constant for all zones in the lens. A lens may be corrected for both spherical aberration and coma for only a single object position. An example of such an *aplanatic* lens is the hemispherical objective lens described in Sec. 42.6. Coma may also be eliminated by an aperture stop of the correct size properly placed on the axis of the lens.

***42.9. Curvature of Field and Distortion.** If the object is an extended plane, the astigmatic images are curved surfaces. For points in the object far from the

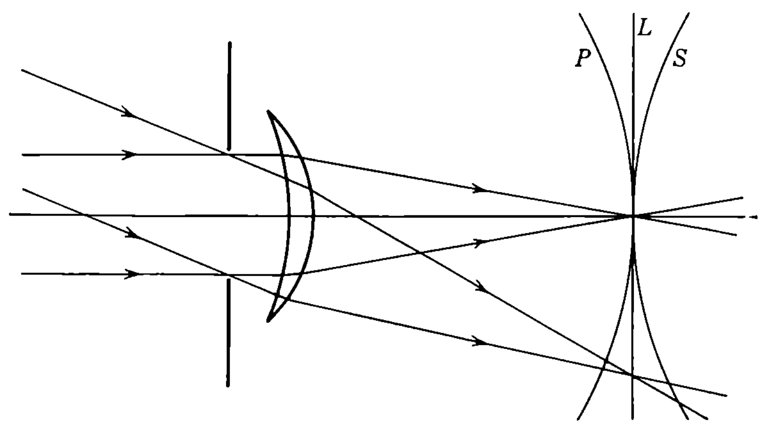


FIG. 42.13. Astigmatic primary and secondary focal planes for meniscus lens. Circles of least condusion lie in plane L.

axis the images are not points but blurred patches, the size of which increases with distance from the axis. With spherical aberration and coma eliminated, these patches are circles of confusion and are the best possible focus. The surface containing this best focus for the entire image is a surface of revolution of a curved line about the axis. Hence we speak of *curvature of field.*

By using special glasses, it is possible to make a pair of thin lenses that form an achromatic doublet (Sec. 42.10) and also produce a flat image field. But astigmatism and curvature of field may be corrected to a considerable extent by the use of a front limiting diaphragm with a meniscus lens (Fig. 42.13). This is done in inexpensive cameras. At all parts of the image field the circles of least confusion, midway between the primary- and secondary-image surfaces, lie very nearly in a plane.

If the magnification is not constant for rays crossing the axis at all angles, there is *distortion* in the image. A properly placed stop aids in correcting distortion, for in a pinhole image (Fig. 39.5) there is constant magnification. Therefore with a single *thin* lens there are also rectilinear projection and freedom from

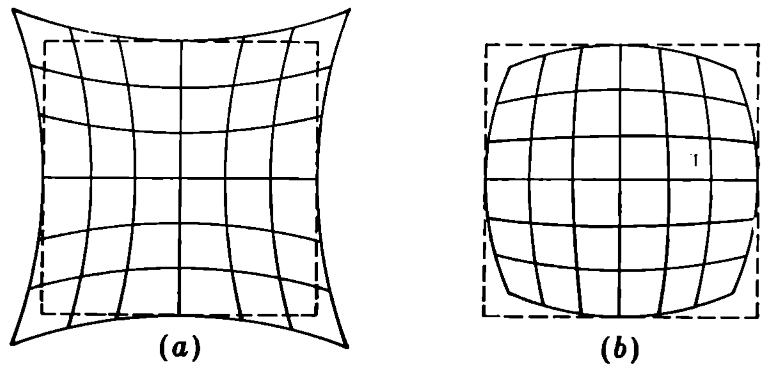


FIG. 42.14. (*a*) "Pincushion" distorted image of a square lattice. (*b*) "Barrel distortion" of same square lattice.

distortion. For an ordinary lens the presence of a stop behind the lens will produce distortion of the pincushion variety, while if the stop is in front of the lens the distortion will be of the barrel type (Fig. 42.14). Therefore, if two lenses are used with a stop midway between them, the two kinds of distortion will tend to counteract each other. A symmetrical doublet-lens system [each component being also an achromatic pair (Sec. 42.10)] with the diaphragm midway between the two components and with properly designed lens surfaces will have this and the other aberrations all fairly well corrected.

***42.10. Chromatic Aberration.** Since the focal length of a lens depends upon the index of refraction, which varies with the wavelength or color of the light, the different colors have different foci. The shorter, blue wavelengths are refracted more than the longer, red wavelengths, as illustrated in Fig. 42.15. This defect, called *chromatic aberration*, cannot be eliminated with a simple lens; but an "achromatic doublet" that has the same focal length for any two colors can be constructed of two thin lenses of different kinds of glass in contact. The construction of an achromatic prism (Sec. 40.10) should be recalled in this connection.

The two colors usually chosen for visual achromatization are those for the C and F lines (Table 40.1), and it is customary to make one of the component thin lenses of crown glass, the other of flint glass. For purposes of illustration we shall use the indices and dispersive powers for the ordinary crown and the dense flint glasses listed in Table 40.2. We shall take the focal length f_D of the combination for the intermediate D line to be 50 mm. The thin-lens equation (Eq. 42.4) may be written

$$\frac{1}{f} = (n - 1)K \tag{42.13}$$

where $K = (1/R_1) - (1/R_2)$. Using primes to denote the crown-glass lens and double primes for the flint-glass lens, equality of focal lengths for C and F light

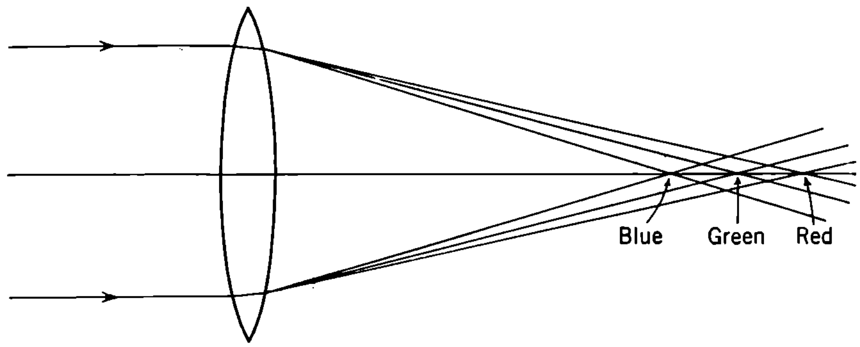


FIG. 42.15. Chromatic aberration.

requires [*cf.* Eq. (42.10)]

$$\frac{1}{f} = \frac{1}{f_C'} + \frac{1}{f_C''} = \frac{1}{f_F'} + \frac{1}{f_F''}$$

or $$(n_C' - 1)K' + (n_C'' - 1)K'' = (n_F' - 1)K' + (n_F'' - 1)K''$$

Therefore $$(n_F' - n_C')K' = (n_C'' - n_F'')K'' \tag{42.14}$$

For the intermediate D-line wavelength,

$$\frac{1}{f_D'} = (n_D' - 1)K' \qquad \text{and} \qquad \frac{1}{f_D''} = (n_D'' - 1)K''$$

or $$\frac{(n_D'' - 1)f_D''}{(n_D' - 1)f_D'} = \frac{K'}{K''} \tag{42.15}$$

Equating the values of K'/K'' from Eqs. (42.14) and (42.15) and rearranging the terms, we have

$$\left(\frac{n_F' - n_C'}{n_D' - 1}\right)\frac{1}{f_D'} = \left(\frac{n_F'' - n_C''}{n_D'' - 1}\right)\frac{1}{f_D''}$$

or $$\frac{d'}{f_D'} = -\frac{d''}{f_D''} \tag{42.16}$$

where d' and d'' are the dispersive powers of the two kinds of glass. Hence f_D' and f_D'' must be of opposite sign, for the dispersive powers are positive. Commonly the convergent lens is made of the crown, the divergent lens of the flint glass. From Table 40.2,

$$d' \text{ (crown)} = 0.016778$$
$$d'' \text{ (dense flint)} = 0.029070$$

Combining Eq. (42.16) with the relation $1/f_D = 1/f_D' + 1/f_D''$,

$$f_D' = \frac{f_D(d'' - d')}{d''} = \frac{50 \times 0.012292}{0.029070} = 21.1 \text{ mm}$$

$$f_D'' = \frac{f_D(d' - d'')}{d'} = \frac{50 \times (-0.012292)}{0.016778} = -36.6 \text{ mm}$$

It is left as an exercise for the student (cf. Prob. 10) to calculate, knowing that $f_C/f_D = (n_D - 1)/(n_C - 1)$, the values of f_C' and f_C'', and hence f_C, to see how closely f_C agrees with the 50 mm value for f_D. The difference between f_F and f_D may be calculated similarly. These differences will be found to be practically negligible.

A common form of achromat is a double-convex crown lens with a divergent lens of flint glass cemented to it. Then $R_1' = -R_2'$ for the convergent lens, and R_1'' for the divergent lens is the same as R_2'. On using Eq. (42.4) with n_D' and n_D'' from Table 40.2, there results $R_1' = 21.82$ mm, $R_2' = -21.82$ mm, $R_1'' = -21.82$ mm, $R_2'' = -241.33$ mm. This achromat is sketched in Fig. 42.16.

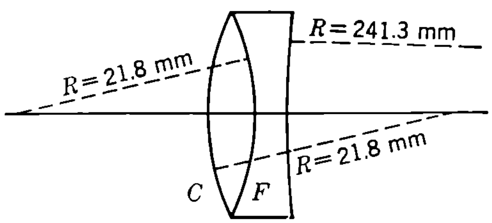


FIG. 42.16. An achromat.

Chromatic aberration may also be greatly reduced by using two lenses of the same kind of glass at a certain distance apart. It may be shown that, if the two lenses are placed at a distance $(f_1 + f_2)/2$ from each other, the combination is nearly achromatic. Many eyepieces (Sec. 43.4) are built in this way.

PROBLEMS

1. If the angular diameter of the sun is 32 min, calculate the focal length of a lens that will produce an image of the sun 4 in. across.

2. A simple lens of index 1.5 has radii of curvature of its surfaces of +50 mm and +75 mm. Calculate the focal length if the lens is "thin."

3. If the lens of Prob. 2 has a thickness of 15 mm, use Eqs. (42.1) and (42.6) to find (*a*) the distance of the focal points from the vertex of each surface; (*b*) the location of the image of a small object on the axis, 150 mm beyond the vertex of the surface, with $R = +50$ mm.

4. Using the formula calculated in Prob. 10 of Chap. 41 for the magnification produced by a refracting surface, compute the relative size of image and object in Prob. 3*b* by taking the product of the magnifications for the two surfaces.

5. Two thin convex lenses of focal lengths 10 cm and 15 cm are 10 cm apart. Calculate the position and size of the image of an object 5 mm long 5 cm in front of the first lens.

6. A thin convex lens of index 1.55 has a focal length of 10 in. in air. What is the focal length in water?

7. Show that for a thin converging lens the minimum distance between an object and its real image is four times the focal length.

8. A hemisphere of glass of index 1.5 has a radius of 5 cm. Locate the principal focus for plane waves incident (*a*) on the flat surface; (*b*) on the spherical surface of the glass.

9. A point source on the axis of a convergent lens is imaged at 30 cm from the lens on its other side. When a second thin lens is placed in contact with the first, the image is formed 40 cm from the combination, on the same side as the first image. Calculate the focal length of the second lens.

***10.** For the achromat calculated in Sec. 42.10, compute the focal length for the C and F lines. Also compute the focal length for the G line in the violet for which $n = 1.52818$ for crown and $n = 1.68079$ for dense flint glass.

***11.** A converging crown lens and a diverging flint lens cemented together, with the common surface of radius of curvature 50 mm, form an achromat (C and F lines) of focal length 100 mm. · Calculate the radius of curvature of each of the other surfaces of the two lenses.

12. A beam of sunlight falls on a thin divergent lens of focal length 30 cm; 60 cm beyond this lens is placed a convergent lens of 45 cm focal length. Where is the image of the sun formed by this combination?

13. A thin double-convex lens of index 1.55 has radii of curvature 20 cm. If one of the surfaces is silvered so that it acts as a mirror, where will the light from a source at a distance of 50 cm from the unsilvered surface be brought to a focus?

14. A thin converging lens of focal length 20 cm has behind it at a distance of 16 cm a thin diverging lens of focal length 5 cm (telephoto lens). Calculate the position of the image of a distant object.

***15.** Show that if the distance between two thin lenses is $(f_1 + f_2)/2$ the chromatic aberration is a minimum.

CHAPTER 43
OPTICAL INSTRUMENTS

43.1. The Eye. Since it is the purpose of most optical instruments to enable us to see better than with the unaided eye, it is proper that we begin our study of these devices with a brief description of the eye itself as an optical system. The parts of the eye are shown in Fig. 43.1. The eye is a nearly spherical shell enclosed by the tough *sclerotic membrane S,* which becomes transparent in front to form the *cornea C.* The *iris I,* the colored part of the eye, contains an aperture, the *pupil.* There are muscular fibers that automatically control the size of the pupil which may vary in diameter from 2 mm when the light is intense to 5 mm for dim light. The "crystalline" lens, together with the ligaments that attach it to the *ciliary muscle M*, divide the eye into an anterior chamber, filled with a clear fluid called the *aqueous humor*, and a posterior chamber, containing a rather jellylike substance called the *vitreous humor V*. Both the aqueous and vitreous humors have an index of about 1.3365, while the crystalline lens has an average index of about 1.437.

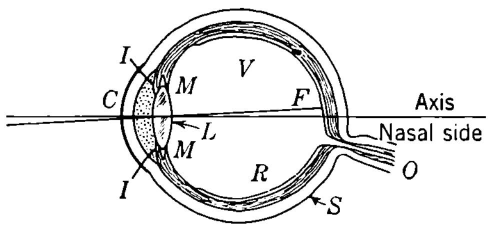


FIG. 43.1. Schematic diagram of the eye.

The lens *L* focuses images on the *retina R*, a covering of the inner surface of the eye containing the nerve fibers, which spread out from the *optic nerve O*. The sensitive organs at the ends of the nerve fibers in the retina are called "rods" and "cones" because of their shapes. At *F* there is a small region about 0.25 mm in diameter, called the *fovea;* this contains only cones and is the portion of the retina where vision is most distinct. The fovea is so small that it subtends an angle of but one degree in the object space. When one looks at an object, the eye always focuses the image on the fovea. The sensitivity of the retina decreases with distance from the fovea. At the place where the optic nerve enters the eye the retina is insensitive to light. This region is called the *blind spot.*

When the normal eye is relaxed, it is focused on objects at infinity. In focusing on near-by objects the ciliary muscles change the shape of the crystalline lens to a more nearly spherical form. The normal eye can focus well on objects as close as 10 in., or 25 cm, without strain. This adaptation of the eye to large changes in object distance is called *accommodation.* The nearest distance at which a small object can be distinctly seen is called the *near point*, while the object distance when the eye is relaxed is known as the *far point*. The latter is infinity for the normal eye. The power of accommodation decreases with age, however, so that the near point for an elderly person may be as far as 200 cm.

43.2. Defects of Vision. The three most common defects of vision are *myopia* (Greek *myō*, close; *opsis*, eye) (nearsightedness), *hyperopia* (Greek *hyper*, over; *opsis*, eye) (farsightedness), and *astigmatism*. In the normal eye the image is formed exactly on the retina. In the myopic eye, however, the eyeball is too long for the lens, and hence the light is brought to a focus in front of the retina. Bringing an object closer to such an eye causes the image to move back toward the retina, possibly exactly on it; and therefore a person with myopic eyes is said to be nearsighted. In the hyperopic eye, on the other hand, the eyeball is too short for the focal length of the lens, and thus the light is brought to a focus behind the retina. A person whose eyes have this defect sees far objects more clearly, because for these the image distance is shorter, and hence he is farsighted. These defects are illustrated in Fig. 43.2.

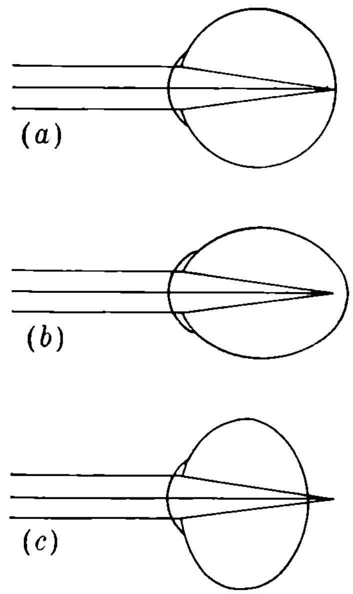


FIG. 43.2. (*a*) Normal eye, (*b*) myopic eye, (*c*) hyperopic eye.

To bring the effective near point for a hyperopic eye to the *distance of distinct vision*, which is taken to be 25 cm for the normal eye, a spectacle lens must be used that will form a virtual image of the object placed at 25 cm at or beyond the real near point. For example, suppose that such an eye has a near point of 100 cm. To calculate the focal length f of the correcting spectacle lens, take $p = +25$ cm, $q = -100$ cm. Therefore

$$\frac{1}{f} = \frac{1}{25} - \frac{1}{100} \qquad \text{or} \qquad f = +33\frac{1}{3}\text{ cm}$$

That is, a converging spectacle lens of this focal length should produce good vision at the normal reading distance.

The myopic, or nearsighted, eye has a far point nearer than infinity. Suppose that it is at 150 cm. What spectacle lens is required so that very distant objects may be seen clearly? Then $p = \infty$; and since a virtual image must be formed at 150 cm, $q = -150$ cm. Hence

$$\frac{1}{f} = \frac{1}{\infty} - \frac{1}{150} \qquad \text{or} \qquad f = -150\text{ cm}$$

Therefore the spectacles should have a diverging lens of this focal length.

In optometry $1/f$ is used as a measure of the *power* of a lens. The unit of lens power, called a *diopter*, is that of a lens with a focal length of one meter. Thus the power of the converging lens with $f = +33\frac{1}{3}$ cm in our example is $+3$ diopters, while the power of the negative lens for our myopic-eye example is $-\frac{2}{3}$ diopter. Clearly, the power of a lens is the sum of the powers of its two refracting surfaces (Prob. 4, Chap. 42) and, from Eq. (42.10), the power of two thin lenses in contact is the sum of their separate powers.

Astigmatism in the eye is caused by lack of sphericity of its refracting surfaces, chiefly the cornea. Then rays of light from lines coming, for example, from the spokes of a wheel-shaped figure are not all focused in the same plane at the retina. This astigmatism is not to be confused with that described in Sec. 42.7, the latter arising solely because of the obliquity of rays incident on a spherical refracting surface. This defect of vision is overcome by using spectacle lenses with slightly cylindrical curvature so oriented as to give, together with the eye surfaces, a total focal length the same for rays in all planes through the optic axis.

43.3. Simple Microscope. If an object is placed slightly nearer to a thin converging lens than the principal focal point, an eye brought close to the lens sees an erect, virtual, enlarged image. A lens used in this way is a simple microscope, or magnifier. If the object is placed at the distance of most distinct vision (25 cm) (Fig. 43.3), it subtends a certain angle, say 2β, at the lens or eye. When the object is at some point closer than the focal point to the lens so that a virtual image is formed at 25 cm, then the object and its image subtend a larger angle 2α. Now the magnifying power of any optical instrument depends upon this gain in angle subtended by the image at the eye. Thus the lens simply allows the object to be brought close to the eye, effectively increasing the power of accommodation of the eye and increasing the size of the image on the retina. The magnifying power M in this case is

$$M = \frac{\alpha}{\beta} = \frac{25}{p}$$

if the angles are small. From the thin-lens equation

$$\frac{1}{p} - \frac{1}{25} = \frac{1}{f}$$

Hence

$$\frac{25}{p} - 1 = \frac{25}{f}$$

or

$$M = 1 + \frac{25}{f} \tag{43.1}$$

The eye can see clearly an object at any distance from the near point, 25 cm, to infinity, and the normal adjustment of the eye is for infinity. If the object is placed at F in Fig. 43.3, the image is at infinity and $p = f$. Then the magnifying power is just

$$M = \frac{25}{f} \tag{43.2}$$

or slightly less than if the virtual image is at 25 cm distance.

A converging lens of focal length 5 cm, for example, has according to Eq. (43.1) a magnifying power $M = 1 + 25/5 = 6$ or, by Eq. (43.2),

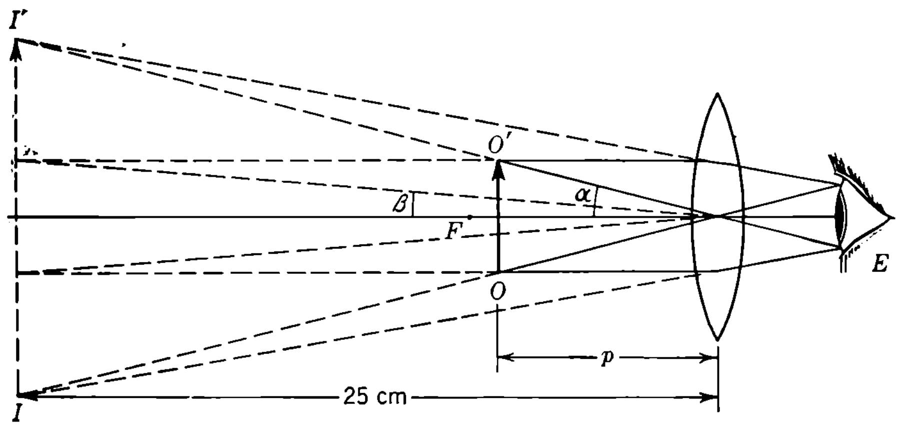


FIG. 43.3. Simple microscope.

$M = 25/5 = 5$. A magnification larger than this with a simple magnifier is not very useful because of the large aberrations for a simple lens of short focus. Therefore magnifiers with powers greater than about 5 should be compound, designed to minimize the aberrations. The Ramsden eyepiece described in Sec. 43.4 is an example of a compound magnifier. A simple magnifier may, of course, be an achromat.

***43.4. Eyepieces.** Any magnifier will serve as the eyepiece, or *ocular*, in a telescope or compound microscope, but with a single lens much of the light from the real image formed by the objective would be lost. Therefore a second lens, called the "field lens," is added to gather in toward the axis more of the light from the objective. The two lenses are of the same kind of glass, and the combination is made nearly achromatic by making their distance apart equal to $(f_1 + f_2)/2$.

The two common types of ocular are the Huygens and the Ramsden. The *Huygens eyepiece* is illustrated in Fig. 43.4. The focus F of the objective serves as a virtual object for the field lens, which forms an erect image at I. This image is at the focus of the eye lens, so that the final image is formed at infinity. A scale or cross hairs may be placed at I, but the scale must not be very large if the eye lens is uncorrected for aberrations. The curvature of the convex surface is such as to reduce spherical aberration to a minimum, but curvature of field may be considerable.

In the *Ramsden eyepiece* (Fig. 43.5) the focal lengths of the two lenses are equal and the convex surfaces face each other. If these two lenses were placed a distance $(f_1 + f_2)/2$ apart, the focal plane of the eye lens would be at the surface of the field lens, with the consequence that dust particles on this surface would be in focus. Therefore at some sacrifice of achromatism the distance between

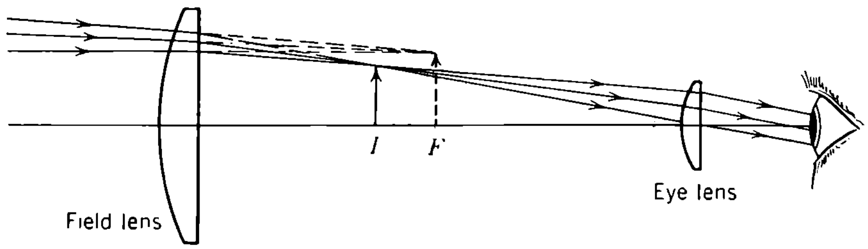


FIG. 43.4. The Huygens eyepiece.

lenses is made $2f/3$. The image I formed by the objective precedes the field lens so that a scale or cross hair may be used successfully at that point. In fact the Ramsden eyepiece is always used in telescopes, such as those on surveyor's instruments, where a cross hair is essential.

★43.5. Compound Microscope. To obtain greater magnification than is possible with a simple magnifier a combination of lenses known as a compound microscope is used. The optical parts of a microscope consist of an objective and an ocular, as sketched in Fig. 43.6, where for simplicity we indicate the lenses to be single. In reality the objective is a complicated train of lenses, and the ocular may be similar to one of those just described. The object O is placed just

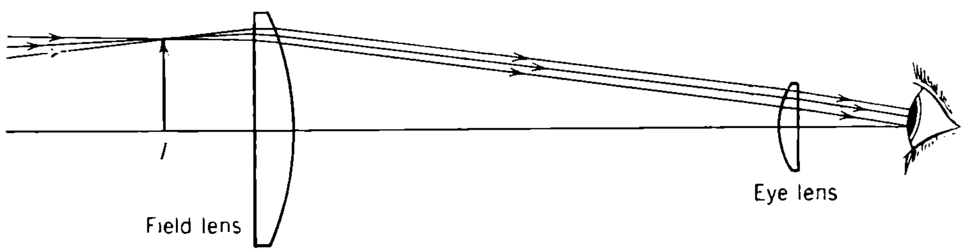


FIG. 43.5. The Ramsden eyepiece.

beyond the focal point of the objective, so that an enlarged real image is formed at I. This image is viewed through the ocular, which produces a virtual image I', further enlarged, at a distance somewhere between 25 cm from the eye and infinity.

The magnifying power of a microscope is the product of the magnification M_1 of the objective and the magnifying power M_2 of the ocular. For the objective,

$$M_1 = \frac{q_1}{p_1} \cong \frac{q_1}{f_1}$$

where q_1 is the distance of the image I from the objective, whose focal length is f_1. On taking the ocular to be a simple magnifier forming the image I' approximately at infinity, $M_2 = 25/f_2$. Therefore the magnifying power M of the microscope is

$$M = M_1 M_2 = \frac{25q_1}{f_1 f_2} \cong \frac{25L}{f_1 f_2} \tag{43.3}$$

since q_1 is approximately the length L of the microscope tube. All lengths in Eq. (43.3) are measured in centimeters.

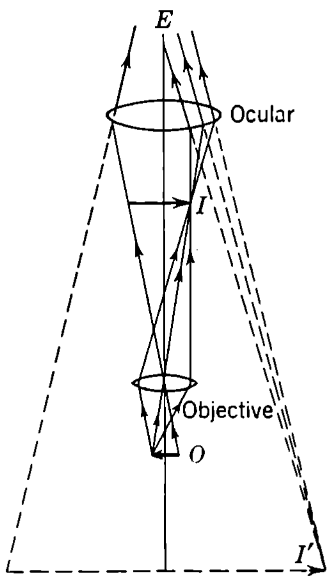


FIG. 43.6. Compound microscope.

A modern microscope objective is in reality a train of lenses, with the front lens usually of the hemispherical form described in Sec. 42.6. A possible arrangement of components is shown in Fig. 43.7. The combination of lenses is designed to have large angular aperture, a minimum of spherical and chromatic aberration, and a flat focal surface. It is usually necessary to illuminate the object by means of a so-called *condenser* located beneath the microscope stage. This condenser may be a concave mirror reflecting a convergent beam of light into the objective; or, with short-focus objectives, a corrected condensing lens is used.

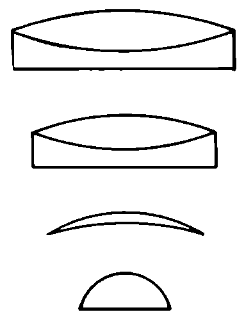


FIG. 43.7. Microscope objective lens.

43.6. Refracting Telescopes. The telescope is similar to the compound microscope in that they both consist essentially of an objective lens and an ocular. Since the object of which a telescope forms an image is very distant, however, there are great differences in the construction of the objective and other features of the two instruments. In the astronomical telescope (Fig. 43.8) the objective forms at I a real image of the distant object. Since the object is at a great distance, I is practically at the focal distance f_1 from the objective lens. Furthermore, the virtual image I' is usually formed at infinity, so that I is at the focal point of the ocular. The distance between the two lenses is, then, $f_1 + f_2$, where f_2 is the focal length of the ocular.

To obtain an expression for the magnifying power M of the telescope we note that the object and its image I subtend the same angle β at the objective. The object subtends practically this same angle β if viewed directly by the eye. When viewed through the ocular, the image I and its image I' subtend the larger angle α, the eye being at E, just behind the ocular. The magnifying power is therefore equal to the ratio α/β, and the retinal image is larger by this factor than it would be if the object were observed directly. Since these angles are small, we may replace the angles by the tangents; and hence the ratio α/β equals f_1/f_2, or

$$M = \frac{f_1}{f_2} \tag{43.4}$$

The modern astronomical telescope is used principally for photographic purposes, and the larger the diameter of the objective the greater the light-gathering power and the greater the resolving power (Chap. 47). In visual work, however, the diameter of the objective sets a limit to the useful magnifying power, for the objective is the so-called *entrance pupil* of the system. The conjugate *exit pupil*

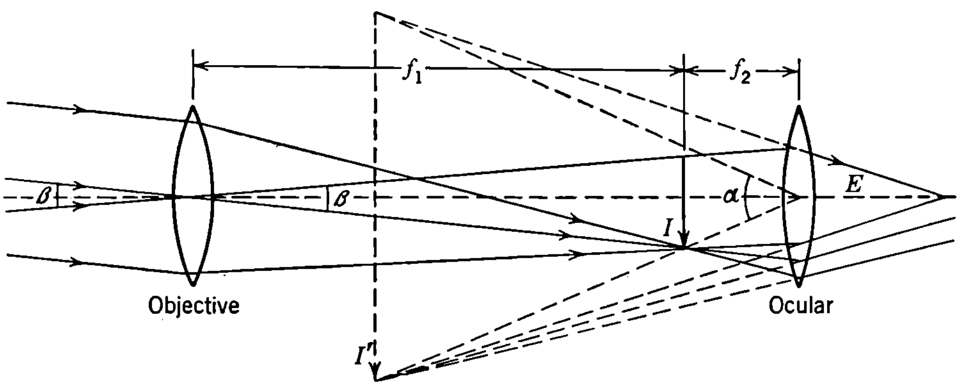


FIG. 43.8. Astronomical telescope.

is the image of the objective formed by the ocular, and at this point the transmitted light has the smallest cross-sectional area. Since the eye is placed at this point, the diameter of the exit pupil should be no greater than the pupillary diameter of the eye if all the light is to be utilized. If the objective lens is the object at a distance $p = f_1 + f_2$ from the ocular, the corresponding image distance q from the ocular to the exit pupil is obtained from $1/q = (1/f_2) - [1/(f_1 + f_2)]$, or $q = f_2(f_1 + f_2)/f_1$. If D_O is the diameter of the objective and D_i that of its image,

$$\frac{D_O}{D_i} = \frac{p}{q} = \frac{f_1 + f_2}{(f_1 + f_2)f_2/f_1} = \frac{f_1}{f_2} = M \tag{43.5}$$

Thus the ratio of the diameter of the objective to the diameter of the exit pupil also is a measure of the magnifying power of a telescope. To find the diameter of the exit pupil for any telescope it should be pointed to the sky and a paper screen used to locate the position where the beam emerging from the ocular has the smallest size. The diameter of the disk of light at this point is D_i.

The pupillary diameter of the eye may be taken as 2 mm. For a telescope with a 150-mm objective, $M = {}^{150}\!/_2 = 75$ if the exit pupil has exactly this 2-mm value. If this objective has a focal length of 1,800 mm, the focal length of the eyepiece should be 1,800/75 = 24 mm to give this normal magnifying power. Although the objective admits 75^2 times as much light as the unaided eye and all of this light enters the eye, there is no increase in the apparent brightness of the object, for the retinal image has now 75^2 times the area.

When a telescope is used for astronomical purposes, the fact that the image is inverted is no disadvantage; but for terrestrial work it is desirable to have an erect image. This is accomplished by inserting between objective and ocular another lens (or, better, a pair of achromats as in the "rectilinear" camera lens to diminish the aberrations) to reinvert the image. This makes the tube of the terrestrial telescope longer by four times the focal length of the erecting lens.

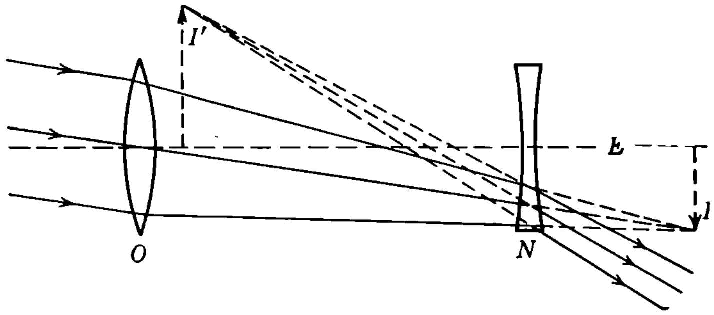


FIG. 43.9. Galilean telescope (opera glass).

In the Galilean telescope, used in opera and field glasses, an erect, virtual image is obtained by replacing the eyepiece with a diverging lens (Fig. 43.9). The converging rays from the objective, which would form the inverted image I if there were no additional lens, are made diverging by the negative lens N. The image I forms a virtual object for lens N, and an eye at E sees an erect, virtual image I'. Since the object is distant and approximately parallel rays emerge from lens N (that is, I' in Fig. 43.9 should be practically at infinity), the magni-

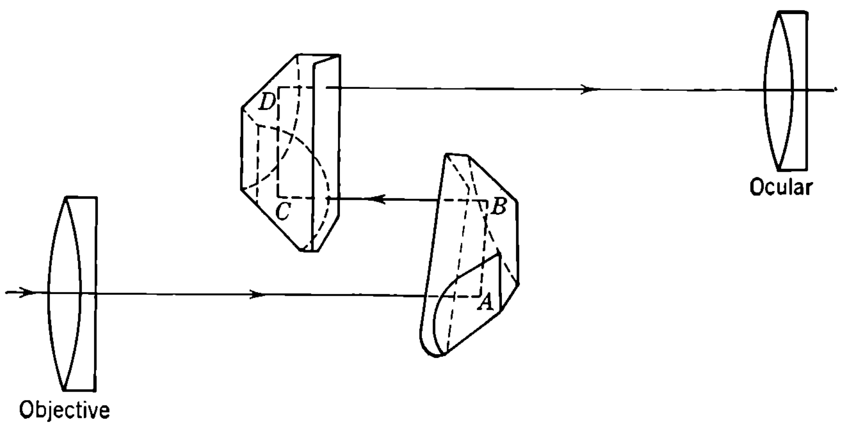


FIG. 43.10. Prism binocular.

fying power is also f_1/f_2 and the tube has a short length approximately equal to $f_1 - f_2$. In this telescope the exit pupil, or the image of the objective lens formed by the ocular, is virtual, and the field of view is restricted.

Prism binoculars also produce an erect image, while at the same time avoiding the excessive length of the terrestrial telescope by placing a pair of right-angled total-reflecting prisms between the objective and eyepiece in the manner indicated in Fig. 43.10. The image formed by the objective serves as a virtual object for the prisms. By the first two internal reflections at A and B, not only is the light reversed in direction, but also the two sides of the image are interchanged so that it is no longer perverted. In the second prism, which is placed at right angles to the first one, the reflections at C and D direct the light toward the ocular and make the image erect. Hence the inverted (and perverted) image produced by the objective is completely restored by the reflections in the two right-angled (so-called "Porro") prisms to an erect position. The ocular forms a further enlarged but erect image. And because of the several reflections an objective of fairly long focal length may be used, with resulting gain in magnifying power, but with the instrument still rather short in length.

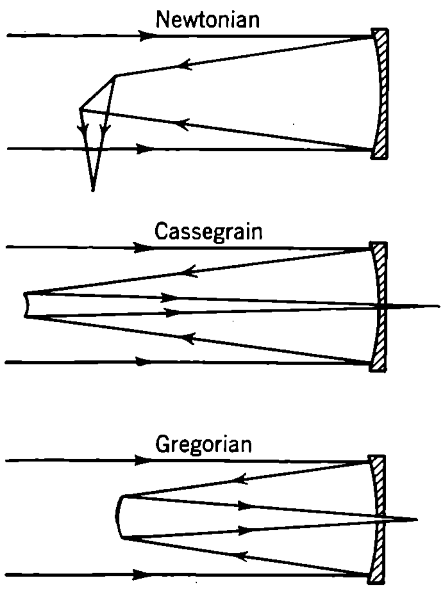


FIG. 43.11. Reflecting telescopes.

***43.7. The Reflecting Telescope.** Because of the difficulties of making a very large objective lens, which must be an achromat of high-quality homogeneous optical glass, the large modern astronomical telescopes are all of the reflecting type. The largest refractor is the 40-in. telescope of the Yerkes Observatory, but the largest reflector at the Mt. Palomar Observatory has a diameter of 200 in. Because of the large area of its reflector, this telescope has both very large light-gathering and resolving powers (Chap. 47). The mirror block is a casting of a borosilicate glass having a small expansion coefficient, but it need not be of optical quality since the light does not pass through it. The mirror has a parabolic surface and is coated with aluminum by evaporation in a high vacuum. Such an aluminum surface has greater durability (because of a protecting surface layer of transparent aluminum oxide) than the silver formerly used, but the reflectance of aluminum is somewhat less than that of silver for visible light.

The image formed by a paraboloidal mirror is free from spherical aberration on the axis, and chromatic aberration is always absent in mirrors. Therefore the focal length may be shorter relative to the aperture than in a refracting telescope. The 82-in. reflector of the McDonald Observatory has a focal length of 26 ft, or a ratio of the mirror's diameter to its focal length of 1/3.8. This ratio of the diameter of a mirror or lens to its focal length is called the *relative aperture* of the mirror or lens and is a measure of its light-gathering power. If of two mirrors of the same diameter one has twice the focal length of the other, and if both are used to produce images of the same object, the image formed by the longer-focal-length mirror will cover four times the area and therefore be only one-fourth as intense as the image from the mirror with the shorter focal length. Therefore the photographic exposure time is only one-fourth as long for the shorter-focal-length mirror, which has double the relative aperture of the other.

The f-number of an optical instrument is defined as the ratio of its focal length to its aperture diameter, 3.8 for the McDonald reflector. In photographic practice this is written $f/3.8$ because this ratio equals the diameter of the effective aperture or "stop."

It is necessary to change the path of the reflected rays to a point where one can observe the image without blocking the incident light. Figure 43.11 indicates three of the methods by which this is done. In the Newtonian method a small plane mirror or right-angled prism reflects the image off to one side. In the Cassegrain type a small convex paraboloidal mirror sends the reflected rays back through a small hole in the large mirror, while in the Gregorian telescope a small concave mirror performs similarly. Several modifications of these last two forms have been perfected to reduce spherical aberration, astigmatism, and coma over a fairly large field. The giant 200-in. telescope on Mt. Palomar is of the Cassegrain form.

***43.8. The Camera.** In this instrument the lens and the photographic plate or film are enclosed in a lighttight box, usually with adjustable lens-film distance. The demands on the lens are exacting and often contradictory. The image must be free from spherical and chromatic aberration, distortion, and curvature of field, while at the same time the aperture angle must be large and the lens must have great light-gathering power. There must be good "depth of focus," *i.e.*, all objects in a considerable range of distances must be brought to a focus in the plane of the film. All points in the object space except those in one plane are not brought to point images at the film, of course, but fortunately the "circles of confusion" which are formed in place of point images are not so large but that a good picture results of all objects within a certain range of distances. The smaller the lens aperture, the greater the "depth of focus."

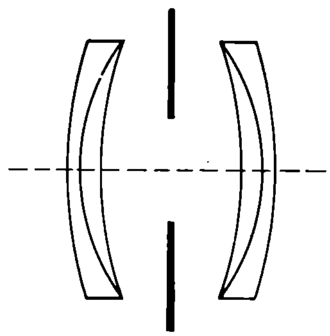


FIG. 43.12. Rectilinear doublet lens minimizes most image defects.

If the lens consists of a symmetrical doublet (Fig. 43.12), each component of which is an achromat, with the iris diaphragm stop at the optical center of the combination, distortions produced in each half correct each other (Sec. 42.9). Chromatic aberration is obviously taken care of, and by allowing the diaphragm opening to be no larger than that demanded for sufficient light intensity the other defects are minimized. Modern high-speed anastigmats such as the famous Zeiss "Tessar" lens have about eight surfaces, very carefully ground so as to correct for all aberrations.

The f-number referred to in Sec. 43.7 measures the speed or light-gathering power of a lens. The smaller the f-number, the "faster" the lens; in fact the required exposure time varies as the square of the f-number. Lenses have been constructed to operate at a stop of $f/1.5$, which means that the focal length is but 1.5 times the diameter of the lens. An inexpensive fixed-focus camera operates at about $f/11$.

***43.9. The Projection Lantern.** The essential features of a projection lantern are shown in Fig. 43.13. The condensing lens C, consisting of a pair of large, simple, uncorrected plano-convex lenses, gathers in the light from the intense source S and focuses it through the lantern slide AB onto the projection lens L. If it were not for the condensing lens much of the light from S passing through the lantern slide would not fall on the lens L, and therefore that is the field to be imaged on the screen would be limited to only the central part of the slide AB.

***43.10. Prism Spectrometers.** The *spectrometer* is the most important optical instrument for the study of light itself, for with it the wavelengths of the light

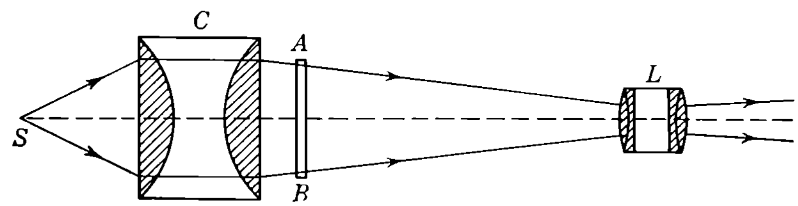


FIG. 43.13. Projection lantern.

emitted from any source may be measured, the effects of interference (Chap. 46), diffraction (Chap. 47), and polarization (Chap. 48) may be studied, and indices of refraction may be determined. If a spectrometer is equipped with an eyepiece for visual observation, it is often referred to as a *spectroscope*, while if a camera replaces the eyepiece it is called a *spectrograph*.

Figure 43.14 shows the essential features of a spectrometer. A collimator tube C carries a narrow slit S, which is illuminated by the light from the source, and an achromatic lens L_1 to render this light a parallel beam as it strikes the prism. The prism disperses the light into the lens L_2 of the telescope T, forming images of the slit, one for each wavelength present in the light, in the plane of the cross hairs just in front of the eyepiece E. The collimator, telescope, and table on which the prism rests all rotate about the central vertical axis of the instrument. The settings of the telescope on the various spectral-line images may be read on the circular scale by means of the verniers V_1 and V_2. Many spectrographs do not have a circular scale.

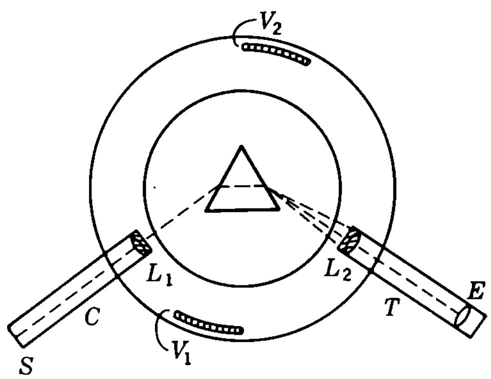


FIG. 43.14. Prism spectrometer.

A cone of rays incident on a prism does not remain a cone after transmission. For all angles of deviation, astigmatism distorts the images. Only when the incident light is parallel is this distortion absent for all colors, and then each color emerges from the prism in a parallel beam. It is therefore necessary for the lens L_1 to be an achromat; and if it is desired to have all colors in focus at the same time at the eyepiece, L_2 must also be achromatic.

To produce a "pure" spectrum the slit S must be very narrow (slit widths as narrow as 0.01 mm are employed) so that the colored images of the slit will not overlap. The "length" of the spectrum depends upon the angular dispersion effected by the prism and upon the focal length of the telescope T. Since the spectrum "lines" are really diffraction patterns (Chap. 47), they will be sharper the larger the apertures of the lenses and prism. Of course, if the source, such as the sun or an incandescent-lamp filament, emits a continuous distribution of frequencies, no matter how powerful the spectrometer the spectrum remains continuous; but the greater the dispersion and the apertures, the easier the detection of weak absorption lines such as the Fraunhofer lines in the solar spectrum. Some characteristic spectra are shown in Fig. 45.1.

For rapid, direct reading of the wavelengths of spectral lines a constant-deviation spectrometer (Fig. 43.15) is often used. The prism is equivalent to two 30° prisms and one 60° prism, cemented together as shown by the dotted lines. For any angle of incidence the light of the wavelength that is at minimum deviation emerges at right angles to the direction of the incident beam, with one total reflection in the prism. The prism may be rotated by a micrometer and screw accurately calibrated in Angstrom units, thus bringing any desired wavelength to the exit slit. This instrument is therefore sometimes called a *monochromator*. The prism and lenses may be made of quartz for use in the ultraviolet region of the spectrum. Also, it may be converted into a spectrograph by attaching a photographic-plate holder at the exit end.

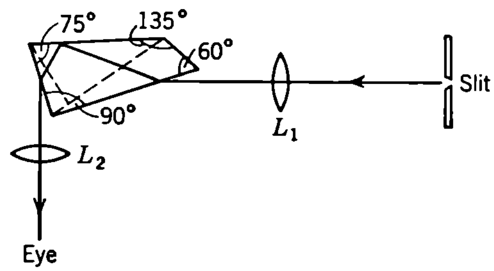


FIG. 43.15. Constant-deviation spectrometer.

A *direct-vision spectroscope* is a hand instrument, useful for rapid inspection of the character of the spectrum of any light. By making use of the differences in the dispersion produced by two different kinds of glass, dispersion without deviation may be produced as described in Sec. 40.10. To obtain sufficient dispersion, about five prisms in a train are used (Fig. 43.16). The entire optical system—slit, collimating lens, prism train, and telescope—is approximately in a straight line, and hence the instrument may be easily pointed at a source. A scale may be included in the focal plane of the eyepiece for semiquantitative wavelength comparisons.

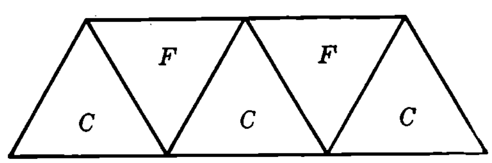


FIG. 43.16. Direct-vision prism train.

As already mentioned, for the production of spectra in the *ultraviolet* an all-quartz optical system must be used. All glasses absorb strongly ultraviolet wavelengths shorter than about 3,300A. Also, since the eye is not sensitive to light of wavelength shorter than 4,000A, the detection of the spectral lines must be photographic. For *infrared* spectrometers, prisms of rock salt have transparency farther into the infrared than those made of most other refracting materials. Concave mirrors may replace lenses; and since both photography and vision fail in the infrared (photographic emulsions sensitive in the near infrared to wave lengths only up to 12,000A are available), a thermopile (Chap. 28) may be used as the detector behind the exit slit.

PROBLEMS

1. The near point for a hyperopic eye is 2 m. What is the focal length of the lens that will enable this eye to see clearly an object 40 cm from the eye? State the power of the lens in diopters.

2. The far point of a myopic eye is at 75 cm. What lens is required for distant vision? Give the power of the lens.

3. If a hyperopic eye needs a spectacle lens having a power of 2 diopters in order to see distinctly an object 25 cm from the eye, what is the shortest distance at which the object could be clearly seen without spectacles?

4. Most of the refraction of light entering the eye takes place at the cornea. Assuming the eye to be filled with a homogeneous medium of index 1.337, calculate (*a*) the radius of curvature of the cornea to place the focal point at the retina 25 mm from the vertex of the cornea; (*b*) the length of the retinal image of an object 25 cm long, at a distance of 1 m from the eye.

5. A simple magnifier of focal length 6 cm is used to view an object by a person having the normal distance of distinct vision, 25 cm. If he holds the glass close to his eye, what is the best position of the object?

6. A compound microscope has an objective lens, of focal length 5 mm, that forms an image 20 cm from the lens. If the magnifying power of the ocular is 10×, what is the total magnifying power?

***7.** The objective of the Yerkes refracting telescope has a diameter of 40 in. and a focal length of 65 ft. What is the normal magnifying power to give an exit pupil of 2 mm? What should be the focal length of the eyepiece to give this magnifying power?

8. What should be the size of the image of the moon produced by the Yerkes telescope objective? The mean distance of the moon is 238,860 miles, and its diameter is 2,163 miles.

***9.** The objective for opera glasses has a focal length of 25 cm. If the magnifying power of the glasses is 4×, what is the focal length of the eyepieces? What is the approximate length of the glasses?

***10.** A projection lantern produces a magnification of 50 diameters on a screen 50 ft from the focusing lens. Considering this as a thin lens, calculate the distance of the lantern slide from the lens and the focal length of the lens.

11. Calculate the relative exposure times of a camera lens when its diaphragm is set at $f/4.5$ and $f/16$.

12. Show that, if a magnifying power either greater or less than the normal value calculated in Prob. 7 is used, there is no gain in the total light in the retinal image.

CHAPTER 44
PHOTOMETRY

***44.1. Luminous Flux.** Photometry is that branch of optics concerned with the measurement of the intensity of light sources. Here "source" is used as a general term including self-luminous bodies and surfaces transmitting or reflecting light falling on them. Most primary light sources, such as the sun and the incandescent lamp, are luminous because of their high temperature. Although these sources are not strictly "black bodies" as discussed in Secs. 22.5 and 22.6, still the distribution in wavelength of the radiant energy that they emit resembles the black-body curves shown in Fig. 48.8. Such thermal radiation is discussed in detail in Chap. 48. The total *radiant flux* coming from a source is defined as the radiant energy passing per second through any surface enclosing the source. Radiant flux is measured in watts. In light measurements we are interested only in the small fraction of the radiant energy having wavelengths lying in the visible region of the spectrum. This is known as *luminous flux*. An exact definition of luminous flux will be given below. In photometry the matters of chief interest are the luminous flux from the source, the luminous intensity of the source, and the illumination, or "illuminance," of a surface. All these terms need definition and explanation.

Because of the fact that the sensitivity of the eye varies with wavelength, the specification of radiant flux in watts alone does not measure the effective "brightness" of the light. The wavelength of the light and the variation of the sensitivity of the eye must be known, also. It is therefore considered necessary in practice to have an arbitrary standard source of luminous flux with which other sources may be compared. This standard is specified in Sec. 44.2. Figure 44.1 is a plot showing the relative magnitude of the visual sensation produced by equal amounts of radiant energy of different wavelengths. This curve is called the *relative luminosity* curve. The maximum ordinate comes at a wavelength of 5,550A, a yellow-green color, and is arbitrarily set equal to unity. This curve is an average of the response of a large number of individuals under good lighting conditions. The brightness sensitivity of the eye falls off so rapidly with wavelength shift from the optimum value that at 6,000A it is only 0.6 and at 5,000A only 0.4 of the value at 5,550A. The International Commission on Illumination has adopted this as a standard representative of a normal eye.

Luminous flux F is defined in terms of this curve giving the *relative luminosity* for every wavelength interval $d\lambda$. Let dF denote the luminous flux of the source in the range λ to $\lambda + d\lambda$; let L_λ be the relative luminosity at the wavelength λ. Then

$$dF = L_\lambda f(\lambda) d\lambda$$

where $f(\lambda)d\lambda$ is the radiant flux in this wavelength range. The total luminous flux F is therefore

$$F = \int_0^\infty L_\lambda f(\lambda) d\lambda \tag{44.1}$$

and may be said to represent the total visual sensation. The function $f(\lambda)$ may be known [it is given by the famous Planck equation (Sec. 48.7) for a black-body radiator]. The integration limits are zero and infinity because the luminosity curve in Fig. 44.1 goes to zero asymptotically at each end; but since an analytical expression cannot be written for this curve, the integration must be carried out graphically. Since this process is cumbersome, the unit of luminous flux is more commonly defined in terms of the flux from a standard source, as explained in the following section.

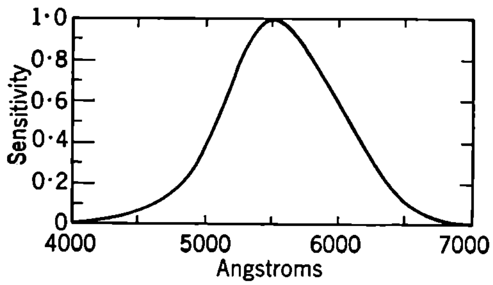


FIG. 44.1. Sensitivity curve ("relative luminosity") for the normal eye.

★44.2. Luminous Intensity of a Point Source. If the dimensions of a source of light are sufficiently small, it may be taken to be a point source. From such a source at S in Fig. 44.2 light radiates out in all directions, and the flux dF passes outward through a cone of solid angle $d\omega$. The luminous intensity I of the source is defined as

$$I = \frac{dF}{d\omega} \tag{44.2}$$

Hence the *luminous intensity*[1] *of a point source is the luminous flux emitted per unit solid angle.* The common unit of I is the *candle.* The unit of luminous flux F, the *lumen,* is the total luminous flux from a point source of *one candle.* Since there are 4π steradians in the total solid angle about a point, the source emits a luminous flux of $4\pi I$ lumens if it has a luminous intensity of I candles.

The standard candle was formerly defined as the luminous intensity in a horizontal direction of the flame of a sperm candle of specified dimensions, burning the wax at the rate of 120 grain/hr. Since the flame of a candle is a rather unsatisfactory source, the primary photometric standards of the United States are a number of specially made carbon-filament lamps operating at 4 watt/candle and calibrated at the National Bureau of Standards by comparison with the standard candle. Other lamps whose intensities have been determined by comparison with these lamps form secondary standards.

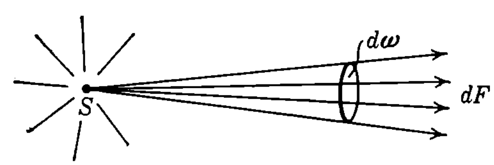


FIG. 44.2. Luminous intensity I of source is $I = dF/d\omega$.

There is now a *new international candle,* established by the International Committee on Weights and Measures and recently introduced by the National Bureau of Standards. This new candle is one-sixtieth of the luminous intensity

[1]Note that the word *intensity* as used here has a meaning different from its usage in other parts of physics where it has the connotation "per unit area."

of a square centimeter of a black-body radiator operated at the temperature of freezing platinum. The new unit is about 1.9 per cent smaller than the former international candle. The photometric ratings of most tungsten-filament lamps will be practically unaffected by the change. Following the recommendation of the Bureau of Standards, we shall refer to the new unit of intensity as the "new" candle and the "new" lumen/steradian.

The term "candlepower" was formerly used in rating the luminous intensity of light sources. Now electric lamps are rated in terms of their power consumption. This is a convenient rating as long as all electric lamps are of the incandescent-filament variety; for although all the input energy is not converted into radiant energy, still the luminous output of a given kind of lamp is roughly proportional to the power consumed. But, in view of the increasing use of fluorescent (Sec. 36.6) and other new types of lamp, comparisons of intensity by power ratings are meaningless. A 20-watt incandescent lamp gives 240 lumens, whereas a fluorescent lamp of the same power rating produces 800 lumens. Obviously it would be much more informative to rate lamps according to their luminous-flux output. The greater efficiency of fluorescent lamps comes from the conversion of the predominant, invisible ultraviolet radiations emitted by the ionized and "excited" gas atoms and molecules into the less energetic visible wavelengths by the fluorescent substances coated on the inner walls of the discharge tubes.

The *luminous efficiency* of any light source is defined as the ratio of the total luminous flux emitted to the radiant flux emitted, expressed in lumens per watt. But note that the watts here measure the total radiation emitted, not the power input to the lamp. Experiment shows that, if some monochromatic light of wavelength of 5,550A has a luminous-flux value of 1 lumen, the radiant flux is $1/650 = 0.00154$ watt. But because the relative sensitivity of the eye at 6,000A is only 0.6, if the radiant flux in light of this latter wavelength is 1 watt, the luminous flux is only 0.6×650 lumens, or 390 lumens.

***44.3. Illuminance.** The measure of the illumination on any surface is called *illuminance*. If the luminous flux incident on a small area dA of the surface is dF, the illuminance E at this spot of the surface is defined by

$$E = \frac{dF}{dA} \tag{44.3}$$

Illuminance is, then, the luminous flux incident on unit area of the surface. If there is uniform illuminance over a finite area A produced by a total luminous flux F, Eq. (44.3) may be written

$$E = \frac{F}{A} \tag{44.4}$$

The flux F may come from any number of sources and from various directions. Illuminance is expressed in lumens/ft^2 (*foot candles*) or lumens/m^2 (*luxes*).

If the illuminance E of the element of surface dA is produced by a point source (Fig. 44.3), there is a simple relation between E and the intensity I of the source in the direction of dA. Let θ be the angle between the normal to the surface element and the distance r between S and dA. The solid angle $d\omega$ subtended by dA at the source is $dA \cos \theta / r^2$, and in this solid angle the flux dF is

$$dF = I d\omega = \frac{I\, dA \cos \theta}{r^2} \tag{44.5}$$

Hence

$$E = \frac{dF}{dA} = \frac{I \cos \theta}{r^2} \tag{44.6}$$

Thus for light from a point source the illuminance of a surface varies inversely as the square of the distance from the source and directly as the intensity of the source and as the cosine of the angle between the normal to the surface and the line connecting the surface element and the source. When the surface is everywhere perpendicular to the light rays from the point source, $\cos \theta = 1$ and Eq. (44.6) becomes

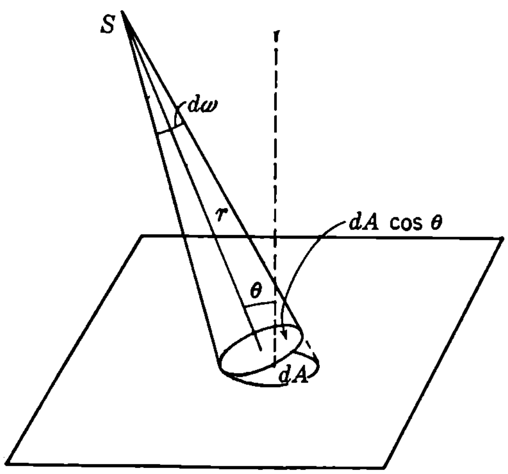


FIG. 44.3. Illuminance produced by a point source.

$$E = \frac{I}{r^2} \tag{44.7}$$

If a uniform point source of luminous intensity 1 candle is at the center of a sphere 1 ft in radius, the illuminance of this surface is 1 lumen/ft², or 1 foot candle. If the surrounding spherical surface is of radius 1 m, its illuminance is 1 lumen/m², or 1 lux. Recently our standards for good illumination have been steadily rising. We now specify 500 luxes or more as desirable for sustained close work. That this is not excessive is indicated by the fact that dull daylight supplies about 1,000 luxes, while in direct sunlight the maximum illuminance is about 100,000 luxes, or 9,600 foot candles.

***44.4. Comparison of Luminous Intensities.** The comparison of the intensities of two light sources is carried out with a *photometer*. This is an instrument used to determine when two identical adjacent surfaces, each illuminated by one of the two sources being compared, are of the same brightness. The two sources to be compared are mounted at the ends of a long, graduated bar (Fig. 44.4), and a movable screen having a mat-white surface on both sides is placed between the two lamps. This screen is moved back and forth until the observer judges that both sides are equally illuminated. Then, from Eq. (44.7),

$$\frac{I_1}{x_1^2} = \frac{I_2}{x_2^2}$$

or

$$\frac{I_1}{I_2} = \frac{x_1^2}{x_2^2} \tag{44.8}$$

where I_1 and I_2 are the intensities of the two sources in the direction of the screen and x_1 and x_2 their respective distances from the screen. If one source is a standard lamp of known luminous intensity at specified voltage and current, the intensity of the other lamp may be found from this measurement.

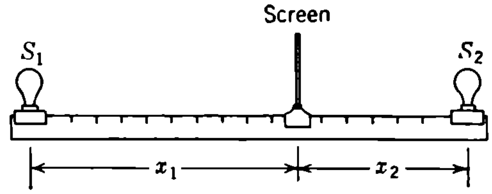


FIG. 44.4. A bar photometer.

There must be some optical device in any visual-match type of photometer to enable the operator to view simultaneously the two illuminated surfaces. This may be done by the simple procedure of mounting two mirrors at such inclinations that juxtaposed images of the two surfaces are provided; the screen may be merely a sheet of paper with a grease spot at its center (*Bunsen grease-spot photometer*). A better method is to observe the screen with a *Lummer-Brodhun photometer* head, which is illustrated in Fig. 44.5. Its elements consist of the screen S, a pair of right-angled prisms P_1 and P_2, an observing telescope T, and a Lummer-Brodhun "cube" C. The latter consists of two right-angled prisms cemented together, one of its prisms having a design etched on its surface to give a field of view, as shown in Fig. 44.5*a*. The paths of the light rays are indicated in the figure. Where the cube is not etched, rays from P_2 pass through to the telescope but such rays are totally reflected at the etched portions. The rays from P_1 striking the etched parts are totally reflected into the telescope, but those rays from P_1 incident on the unetched parts pass through the cube and out to the right. The shaded portions of the field of view (Fig. 44.5*a*) receive light, then, only from one side of the screen S, while the remainder of the field of view is illuminated only by light from the other side of the screen. The procedure is to move the photometer head back and forth on the track between the two light sources until the pattern in the field of view disappears, indicating equality of illumination of the two surfaces of the screen. This method of viewing the two surfaces of the screen simultaneously permits of rapid and precise judgment of this position.

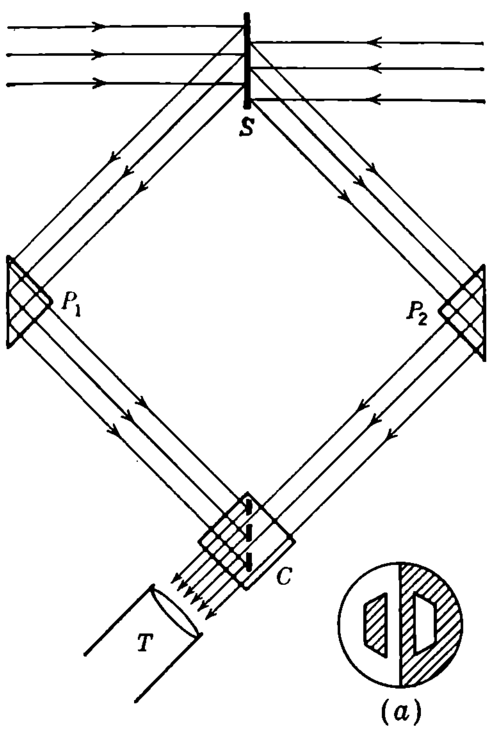


FIG. 44.5. The Lummer-Brodhun photometer.

The *shadow photometer* consists merely of a rod R and a screen (Fig. 44.6). The sources S_1 and S_2 are adjusted in position and distance until the two shadows on the screen are touching and have equal illuminance E. Then, if r_1 and r_2 are the respective distances of the two sources from the screen, it follows from Eq. (44.7) that

$$\frac{I_1}{I_2} = \frac{r_1^2}{r_2^2} \qquad (44.9)$$

is the ratio of the intensities of the two sources.

A photometer should be operated in a darkened room with black walls so that no illumination of the screen surfaces arises from extraneous sources. Also, the distances x_1 and x_2 from the two sources to the screen should be large compared with the source dimensions, for it is assumed that the sources are effectively points.

Such visual-match types of photometer have been rendered somewhat obsolete, however, by photovoltaic cells such as those described in Sec. 36.7. These are used in the modern *foot-candle meter* and in the photographic-exposure meter. The sensitive surface of the cell when exposed to light of appreciable intensity generates an electric current large enough to actuate a microammeter. Since the current produced is a function of the light intensity, the meter may be calibrated to read directly in footcandles or luxes. Filters are now available that give the meter practically the same wavelength-sensitivity characteristics as are possessed by the normal eye.

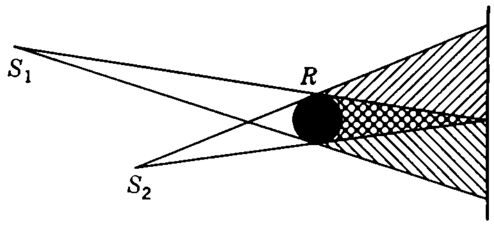


FIG. 44.6. Shadow photometer.

★44.5. Intensity Comparisons Involving Color Differences. In comparing the intensities of two lamps with a visual-match photometer, difficulties are encountered if the light from the two sources is not of the same color. The eye cannot accurately judge the equality of illumination of two surfaces unless their colors match. Lamps of different colors can be compared, however, with a *flicker photometer*, with which the observer views alternately first one side of the screen and then the other. Upon increasing the frequency of alternation a value will be found at which the flicker due to color difference disappears but the flicker due to illumination differences persists. Then the screen is moved until this flicker also disappears, and the intensities are compared in the usual way.

In a *spectrophotometer* the light from the two lamps is first dispersed into spectra, and the comparison is then made step by step, by matching always light of the same color. One of the beams could have passed through a specimen of colored transparent material, while the other beam from the same lamp by-passes the specimen. Comparison of the intensities of the two beams, color by color, enables one to draw a spectral transmission curve giving the fraction of light of each wavelength transmitted by the material.

★44.6. Extended Sources. Luminance and Total Flux. Most sources of light may not be considered as points unless the distance to them happens to be very large. Consider, for example, an incandescent sheet of metal close at hand, the reflecting walls of the room, or the diffusely reflecting page you are reading. Let A be a small element of area of the surface, F be the total flux it emits throughout the entire solid angle of 2π steradians on the one side of the surface, $d\omega$ be the solid angle formed by a pencil of rays (Fig. 44.7), and dF be the portion of the total flux F included in this solid angle $d\omega$. If the pencil of rays is along the normal n, the intensity I_n of this surface element in that direction is $I_n = dF/d\omega$, while in the pencil at an angle θ with the normal the intensity I_θ is $dF_\theta/d\omega$. The *luminance* (formerly called "brightness") of this surface element in any direction is defined as the ratio of its intensity in that direction to the projection of the area A on a plane perpendicular to that direction. Thus, denoting the luminance in the direction θ by B_θ,

$$B_\theta = \frac{I_\theta}{A \cos \theta} \qquad (44.10)$$

or *luminance is the intensity per unit projected area of emitting surface.*

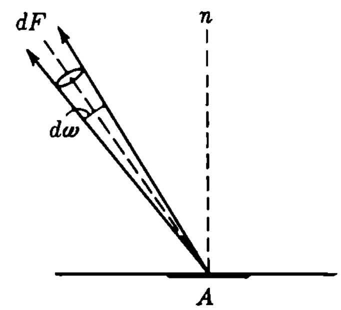


Fig. 44.7. The intensity of the surface element in any direction is $dF/d\omega$.

Lambert found that the luminance of a diffusely radiating or reflecting surface is practically independent of the angle from which it is observed. If that is the case,

$$I_\theta = I_n \cos \theta \qquad (44.11)$$

for then $B_\theta = (I_n \cos \theta)/(A \cos \theta) = I_n/A = B$, a constant independent of θ. Equation (44.11) is known as *Lambert's cosine law.* Hence an incandescent sphere observed from a distance considerably larger than its radius appears to be a uniformly illuminated disk. The sun, being a luminous sphere, should be an example of this effect; but since it is surrounded by an absorbing atmosphere through which light from the edges passes more obliquely and hence is more strongly absorbed, photographs of the sun show the edges of the disk to be less bright than the center.

The "new" candle mentioned in Sec. 44.2 is actually a unit of luminance rather than of luminous intensity. That is, the luminance of a black body at the temperature of freezing platinum is by agreement declared to be 60 new candles/cm².

***44.7. Worked Examples.** *a.* A standard lamp of intensity 30 candles and an unknown lamp are at opposite ends of a photometer bar 200 cm long. They produce equal illuminance on the photometer screen when the latter is 120 cm from the standard lamp. What is the intensity of the unknown lamp?

The illuminance of the screen is $(30 \text{ lumens/m}^2)/(1.2)^2 = I/(0.8)^2$ lumens/m².

$$I = \left(\frac{0.8}{1.2}\right)^2 \times 30 \text{ candles} = 13\frac{1}{3} \text{ candles}$$

b. A lamp whose intensity is 100 candles is placed 5 m from a screen that reflects 75 per cent of the light incident upon it. If the screen is a diffuse reflector, what is its luminance?

The illuminance E of the screen is $100/5^2$ lumens/m² = 4 lumens/m².

Its luminance B is 0.75×4 candles/m² = 3 candles/m² = 3π meter lamberts. (1 *lambert* = $1/\pi$ candles/cm², 1 *foot lambert* = $1/\pi$ candles/ft², 1 meter lambert = $1/\pi$ candles/m².)

*44.8. Summary

Since the terminology used in this chapter tends to be burdensome, we close with a résumé of definitions.

Radiant flux is power emitted by a source and is measured in watts.

Luminous flux is a measure of the visual sensation. It is defined by Eq. (44.1) and measured in lumens.

Luminous intensity is luminous flux per unit solid angle [Eq. (44.2)] and is measured in candles (lumens/steradian).

Illuminance is luminous flux per unit area incident on a surface [Eq. (44.3)] and is measured in lumens/ft^2 (foot candles) or lumens/m^2 (luxes).

Luminance is the luminous intensity per unit area emitted by an extended source. It is measured in candles/ft^2 or candles/m^2.

PROBLEMS

1. Calculate the illuminance on the pavement at a point directly beneath a street lamp of intensity 1,000 candles if the lamp is at a height of 20 ft. What is the illuminance at a point on the pavement 30 ft from this spot?

2. A midget high-pressure mercury-vapor lamp is reported to have an intensity of 46 million candles. At what distance from this lamp would the illuminance be equal to that of bright sunlight (9,600 foot candles)?

3. A floor is illuminated by four lamps of intensity 200 candles each, spaced at the corners of a rectangle 20 by 30 ft and hanging at a height of 15 ft above the floor. Compute the direct illuminance on the floor at a point equidistant from each of the lamps.

4. Two lamps of 75 and 100 candles intensity, respectively, are 10 ft apart. A screen is placed between the lamps with its plane at right angles to the line joining them. Find the position of the screen where its two sides have the same illuminance. Also find a possible location for the screen, not between the two lamps, where the illuminance of one side of the screen equals that of either side before.

5. A photometer bar 3 m long has a standard lamp of intensity 30 candles at one end, while at the other end is the circular, flat window of an "end-on" gaseous-discharge tube. The window, of 2 cm diameter, with its surface normal to the bar, has a uniform luminance from the flux issuing from the tube such that the photometer is balanced with the screen 1 m from the window. Calculate this luminance in candles/cm^2 and in lamberts.

6. A 100-watt lamp delivers 1,200 lumens when operated at its rated power. At what height must this lamp be suspended above a desk to produce an illuminance on the desk equal to that of dull daylight (100 foot candles)?

***7.** Prove that for an infinitely long line source of light the illuminance on a parallel surface at a distance x is $2I/x$, where I is the intensity per unit length.

8. A 200-candle lamp is suspended 10 ft above a sheet of white blotting paper, which reflects 50 per cent of the light incident on it. Calculate the illuminance on the blotter, as well as its luminance.

9. To give the desirable 500 luxes of illuminance on a workbench, how high above it must a 200-watt lamp be placed, if the over-all efficiency of the lamp is 12 lumens/watt?

10. If the full moon produces an illuminance of 0.02 lumen/ft^2, at what distance will a 500-candle lamp give the same illuminance?

CHAPTER 45

SPECTRA AND COLOR

The study of the light emitted from a self-luminous body or that reflected from a surface is important from two points of view. On the one hand, we can obtain information about the mechanisms by which light is produced or absorbed; on the other, the analysis and classification of the subjective sensations of the color of the light have many technological applications.

45.1. Types of Spectra. The light from any source may be dispersed by means of a prism spectrograph (Sec. 43.10) or a diffraction grating (Sec. 47.10) into a spectrum giving a wavelength, or frequency, analysis of the light. The detailed study of various spectra has contributed greatly to our knowledge of the structure and dynamical behavior of atoms and molecules. These studies led to the development of the quantum theory and quantum mechanics, and these radically new theories governing the interaction of electromagnetic radiation and matter have in turn been instrumental in predicting and interpreting numerous startling discoveries in atomic physics.

Spectra may be grouped into five main classes:

Continuous emission spectra.

Line emission spectra.

Continuous absorption spectra.

Line absorption spectra.

Band or molecular spectra (both emission and absorption).

Note that the various classifications for absorption spectra are the same as for emission spectra.

Continuous emission spectra come chiefly from incandescent solids, an example being the light from the ordinary tungsten-filament lamp. The laws governing such continuous radiation will be discussed in detail in Chap. 48. Most of the radiant energy produced by such sources lies in the infrared region and only a small part is visible light. The emission spectrum of molecular hydrogen from a lamp similar to a short neon sign, but filled with hydrogen gas at a few millimeters of mercury pressure, however, consists in part of continuous radiation through the entire ultraviolet. It is well established that, upon emitting this continuous ultraviolet radiation, the hydrogen molecule dissociates into two hydrogen atoms. Another source of continuous emission spectra in the far ultraviolet even to X-ray wavelengths is the light viewed "end on" of the discharge of a large capacitance at a high voltage through a narrow capillary tube filled with gas at a low pressure. In such a discharge the current density may be 10,000 amp/cm^2, and consequently the concentration of electrons and atomic ions is large. These recombine and emit energy as light with a continuous wavelength distribution.

Line spectra are characteristic of electric discharges in gases at low current densities and of electric arcs and sparks between metallic electrodes. In the arc or spark some of the electrode material is vaporized into the discharge region so that the light comes from atoms of the metallic elements involved. The spectral *lines* are images of the spectrograph slit, which is illuminated by the light under investigation. Each frequency or wavelength in the given sample of light is represented by its line image, and every element exhibits a unique set of spectral lines. For this reason the spectrograph has become a valuable tool in analytical chemistry. Quantities of an impurity as small as 1 part per million may be rapidly and surely detected by examination of the spectrum of the material for the most prominent lines of that element. Figure 45.1 is a photograph of these prominent lines for a number of the elements.

Absorption spectra are produced by inserting gaseous or liquid material into the path of the light having a continuous emission spectrum. Spectral frequencies characteristic of the atoms and molecules of the absorbing substance are observed as dark lines or bands in the bright emission spectrum. An example of this process is the dark *Fraunhofer lines* in the solar spectrum. These lines in the spectrum of sunlight originate mostly from absorption by the atoms in the relatively cool outer layers of the sun's atmosphere, but there is also absorption by the oxygen, ozone, water vapor, and carbon dioxide in the earth's atmosphere. From comparison of wavelengths of Fraunhofer lines with atomic spectra produced in the laboratory at least 60 of the elements are known to exist in the sun.

Evident series of lines whose frequencies may be related by relatively simple formulas occur in the atomic spectra of hydrogen and the alkalies and alkaline earths that fall in the columns on the left-hand side of the periodic table (Table 49.1). The spectra of the elements in the center and right-hand columns of the periodic table, on the other hand, contain in general large numbers of lines apparently distributed at random. Those lines in a complex spectrum which are closely related may often be located by varying the excitation conditions in the source, changing the pressure of the gas, causing the light emission to occur in a strong magnetic field, etc.; for such variations affect similarly the breadth, frequency, and intensity of all the spectral lines in the same series. We shall return to the subject of regularities in the spectra of atoms and

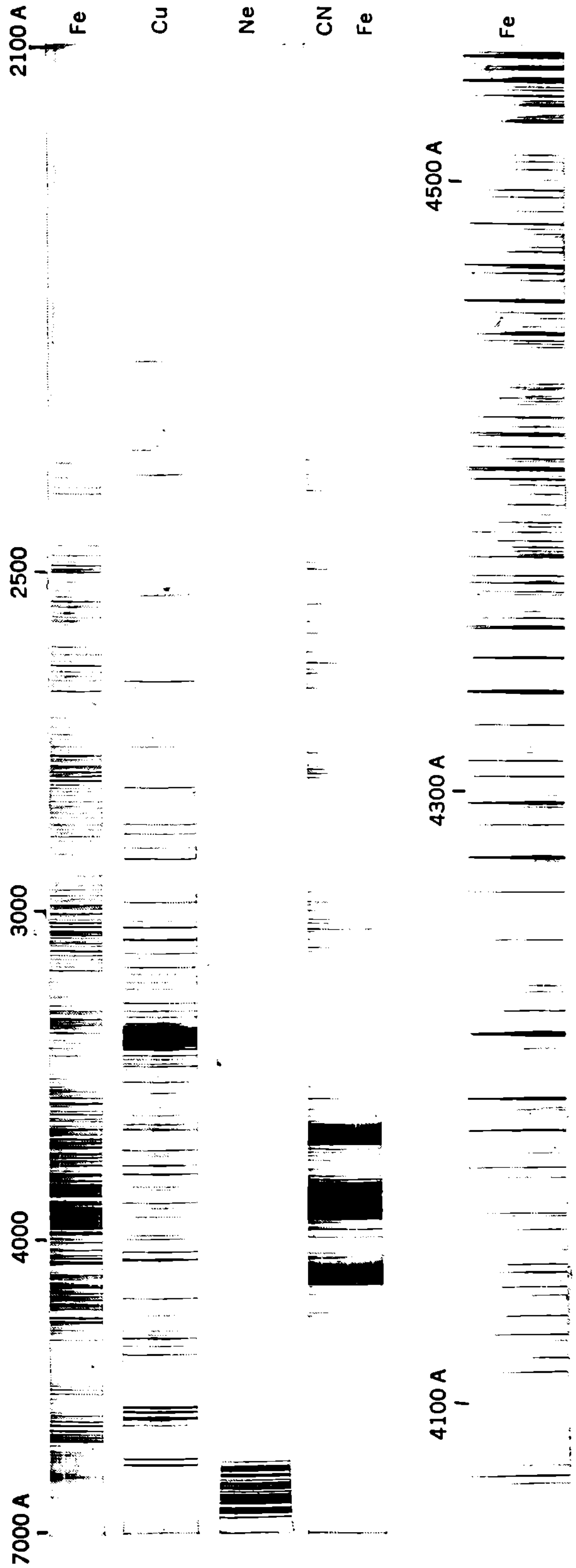


FIG. 45.1. Spectra of several elements.

molecules in Chap. 49. For the remainder of the present chapter we shall consider the quantitative description of color.

45.2. The Sensation of Color. When we say that a prism disperses a beam of light into all its component colors, we are using the term "color" loosely. Color is actually a sensory experience, influenced by association, fatigue, and sharpness of vision and by color blindness. The physicist often contents himself by measuring the luminous flux at each small wavelength interval in a given spectrum and is well aware that the eye is sensitive to but one octave out of the many in the complete range of electromagnetic radiations. Because the use of colored objects has become of such importance in art, in industry, and in the daily lives of all of us, however, it has become necessary to develop a science of color, *colorimetry*. For many purposes it may be sufficient, or even desirable, to call a color "aquamarine," "orchid," or "apple green," but as an aid to industry it is becoming increasingly necessary to describe and measure color in terms of the subjective sensations.

The ear can analyze complex sounds into different frequencies. The eye, even though it can detect a small difference in color when the dispersed light comes directly from the prism, does not have the ability to recognize the component frequencies in a synthetic color. The sensation of yellow, for example, may be produced by superposing red- and green-colored lights on a white screen. Despite this characteristic of our visual sense the science of colorimetry has developed to the point where it can adequately specify colors numerically.

The aspects of the color sensations that are to be *described* as quantitatively as possible are *brightness, hue,* and *saturation*. The first of these qualities is determined by the amount of luminous flux in the light. Descriptive of the hue and saturation of a color are the so-called *dominant wavelength* and *purity,* respectively. These two characteristics will now be considered.

45.3. Color of Reflecting Surfaces. We observe most colored objects by the light reflected from their surfaces. The color of this reflected light depends not only on the color of the surface but also on the color of the light illuminating the surface. Daylight is naturally chosen as the standard for the illumination of colored surfaces. The color of daylight may be simulated rather well with a gas-filled tungsten-filament lamp operating at a temperature of about 2850°K and used with specified filters. Such calibrated lamps and filter specifications may be obtained from the National Bureau of Standards. "Daylight" tungsten lamps with blue-colored glass envelopes give a qualitatively correct spectral distribution.

The *reflectance* of a surface at a wavelength λ is the fraction of the incident light at that wavelength which is reflected. This fraction is independent of the intensity of the light. A body that appears white by daylight has about the same reflectance for all wavelengths. Gray is similar to white, but with a smaller value for the reflectance.

The variation of the reflectance of a surface against the wavelength of the light is obtained with a spectrophotometer. In the best type of automatic recording spectrophotometer a single photoelectric tube is used as the detector. The light from the source, after being dispersed by a prism, falls on a slit, which isolates a narrow band of wavelengths. This

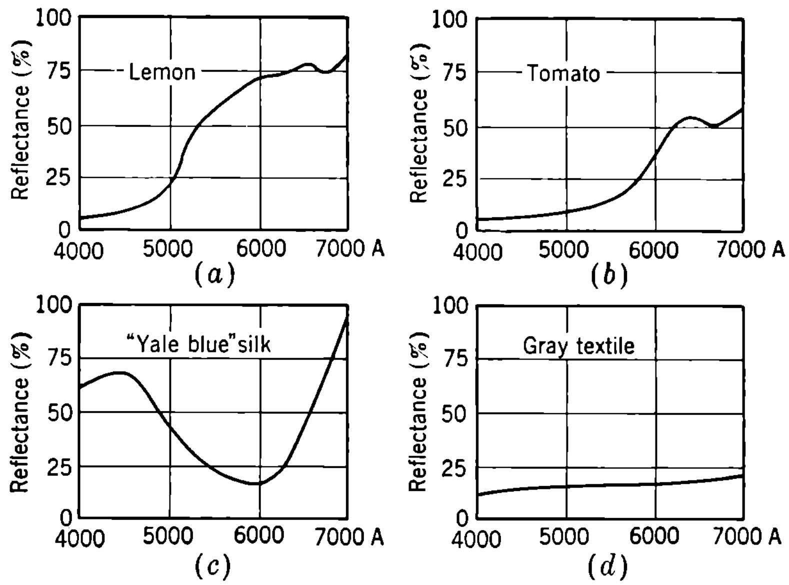


Fig. 45.2. Spectrophotometric curves showing the reflectances throughout the visible region for several surfaces.

beam is then split into two, which are reflected, respectively, from the sample and from a comparison surface of magnesium oxide. The latter forms an almost perfectly white surface, *i.e.*, one for which the reflectance is uniformly 1 for all the wavelengths of the visible spectrum. The two reflected beams are then sent in rapid alternation into the phototube. An electronic servomechanism drives the recorder, which is actuated by the difference of the two reflected beams so that any fluctuations in the original light cancel out.

Figures 45.2*a* to *d* are examples of reflectance curves automatically recorded by this instrument. Notice that all these surfaces reflect some light at every wavelength. The tomato's simple red is seen to be a complex color mixture. Red, orange, and yellow objects reflect highly throughout the longer wavelength regions. The specimen of "Yale blue" silk exhibits higher reflectance in the violet at 4,300A and in the red beyond 6,700A than in the blue at 4,800A, although the visual sensation is a shade of blue. It is characteristic of the purples to show high violet, blue, and red reflectance with low green and yellow. The low, uniform reflectance of the gray cloth for all wavelengths is typical of neutral grays.

Such photometric curves, particularly after multiplication by the luminosity curve (Fig. 44.1) to take account of the variation of the sensitivity of the eye with wavelength, could serve as a complete specification of the color stimulus. The effective dominant wavelength, the purity, and the intensity of these colors are all inherently involved in the data plotted in these curves. Two objects of identical color may have quite different spectrophotometric curves, however, and such curves are not used by color experts directly for color specifications. Rather, they contribute the data to which the *tristimulus system* of color specification is applied. The *tristimulus coefficients*, explained in the next section, form a convenient and accurate specification of color and one that is the same for identical colors even if their spectrophotometric curves are different.

***45.4. Addition of Colors. Tricolor Stimulus.** By *additive mixture* of three *primary* colors—red, green, and blue produce the widest range of colors—in proper proportions, many color sensations can be matched. An example of color mixture by addition is the superposition of beams of colored light rendered visible because of scattering from smoke particles in the air. Where red and green beams overlap, a yellow color results, while at the convergence of visible beams of blue, green, and red light their sum is white. Actually it is true that not *every* color may be matched by a mixture of the three component colors; but, in the event that a color cannot be so matched, either one (or two) of the component colors may be mixed with the unknown color and then a color match may be made with the other two (or one) component colors. Thus in the latter case it is at least possible to specify the unknown color by giving the amount of a component (or components) that, when added to it, produces a match with given amounts of the remaining components. That is, any color may be expressed in terms of *any* three components, the match being expressed by three numbers.

Thus, 1 watt of radiant flux of any color C, is matched by

$$R + G + B$$

where R is some definite flux of a primary red and G and B are fluxes of green and blue primary colors. If amounts of one or two of the components must be added to C to produce a color match, these amounts must be considered negative. It is possible to avoid negative values by using other primaries that lie outside the realm of real colors. These primary colors are unreal in the sense that they cannot be specified by spectrophotometer curves. They can, however, be uniquely defined by stating the relative amounts of each one of the primaries which must be added together to match any pure spectral color.

The International Commission on Illumination (ICI) has prepared tables that give on an arbitrary scale the relative amounts X, Y, and Z of three such primary colors necessary to match any given color of definite wavelength. The plots of these values constitute the standard color-mixture curves sketched in Fig. 45.3. These primaries are chosen so as to make Y the luminosity curve that was given in Fig. 44.1. Since the ordinates are in arbitrary units, they have been adjusted so that the areas under each of the three curves (the X curve has a double maximum) are equal. To every wavelength there correspond ordinates X, Y, and Z, which are, respectively, the heights of the three curves at this wavelength. These three ordinates X, Y, and Z are called *tristimulus coefficients*.

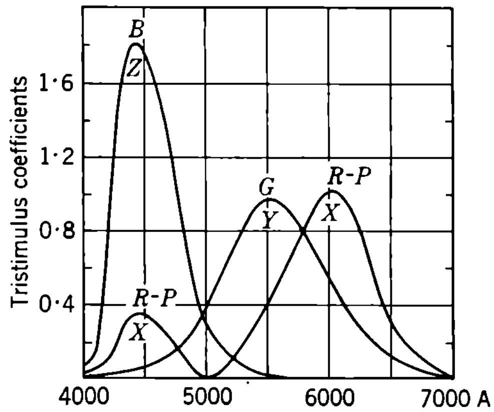


FIG. 45.3. Standard ICI color mixture curves.

A spectral color is thus specified by the values of the tristimulus coefficients and could be represented by a point with these three coefficients as coordinates. The ICI components are not wholly arbitrary; all visible colors can be analyzed with them.

Since a three-dimensional diagram would be required to represent all colors with these tristimulus coefficients, we transform these into so-called *trichromatic coefficients* x, y, z by the equations

$$x = \frac{X}{X + Y + Z} \qquad y = \frac{Y}{X + Y + Z} \qquad z = \frac{Z}{X + Y + Z} \tag{45.1}$$

These new coefficients x, y, and z are not all independent, for $x + y + z = 1$. Hence a two-dimensional graph, plotting any two of these quantities, say x and y, for each color should serve to represent all the spectrum colors. For example, the values of the tristimulus coefficients for yellow-green light of wavelength 5,500A are, from Fig. 45.3,

$$X = 0.450 \qquad Y = 0.990 \qquad Z = 0.010$$

Therefore $X + Y + Z = 1.450$, and

$$x = \frac{0.450}{1.450} = 0.310, \qquad y = \frac{0.990}{1.450} = 0.683$$

Hence $z = 1 - (x + y) = 1 - 0.993 = 0.007$.

A similar calculation may be made for a large number of wavelengths spaced at sufficiently close intervals. Each such pair of x and y values forms a point on the curve shown in Fig. 45.4. This is the curve of trichromatic coefficients, or, better, the *chromaticity diagram.*

The three tristimulus coefficients of any spectral color are defined as the amounts of the three ICI components that together match the color. A given color corresponds to a multiplicity of wavelengths each with its own strength as given by a spectrophotometer curve. Let $f(\lambda)$ be the distribution. Then the

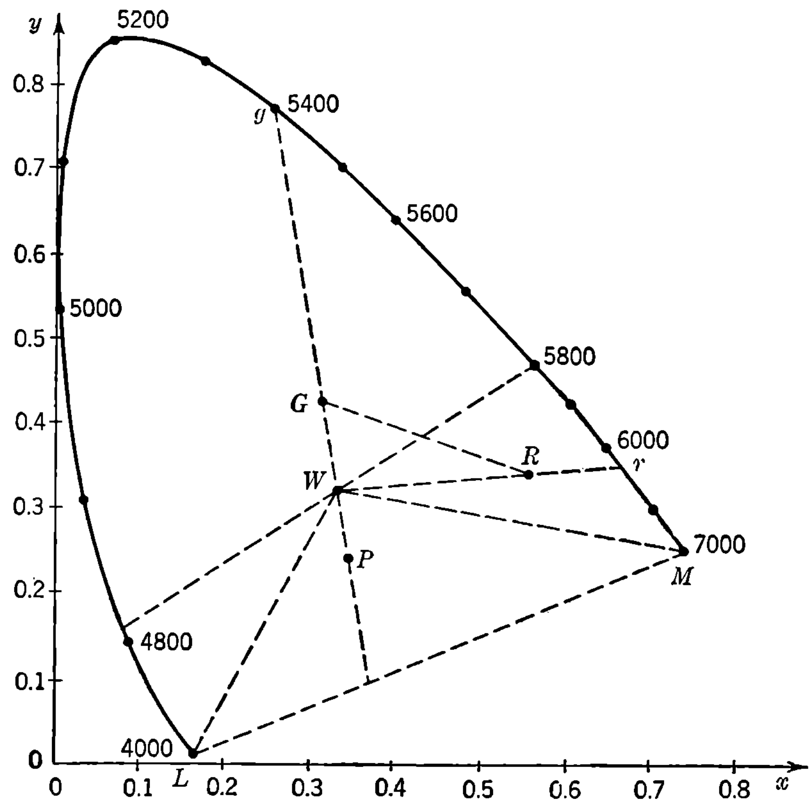


FIG. 45.4. Chromaticity diagram. W is the "white point."

amount of the first component required for the wavelength range $d\lambda$ is $Xf(\lambda)d\lambda$ and therefore the total amount $\bar{X}$ of this primary for a match is

$$\bar{X} = \int_0^\infty Xf(\lambda)d\lambda \tag{45.2}$$

Similarly
$$\bar{Y} = \int_0^\infty Yf(\lambda)d\lambda \quad \text{and} \quad \bar{Z} = \int_0^\infty Zf(\lambda)d\lambda \tag{45.3}$$

The three curves of Fig. 45.3 cannot be expressed analytically, and hence these integrations must be done graphically or numerically.

If the data for the sample of colored light come from a spectrophotometer reflectance curve that gives the reflectance r_λ at each wavelength and if the radiant flux incident on the reflecting material in this small wavelength range is $F(\lambda)d\lambda$, then $f(\lambda)d\lambda = r_\lambda F(\lambda)d\lambda$. The distribution $F(\lambda)d\lambda$ in the illuminant is usually that for daylight, which is well known. The tristimulus coefficients having been determined, the trichromatic coefficients become

$$x = \frac{\overline{X}}{\overline{X} + \overline{Y} + \overline{Z}} \qquad y = \frac{\overline{Y}}{\overline{X} + \overline{Y} + \overline{Z}} \qquad z = \frac{\overline{Z}}{\overline{X} + \overline{Y} + \overline{Z}} \tag{45.4}$$

When this procedure is carried out for a daylight *white* source, the trichromatic coefficients 0.3101, 0.3163, and 0.3736 result. The fact that these coefficients are each nearly equal to $\frac{1}{3}$ indicates that the ICI primaries mixed in about equal proportions produce daylight white. The first two of these numbers are the coordinates of the "white point" W plotted in Fig. 45.4.

All points inside the curve of the chromaticity diagram represent real colors. As the points get nearer to the boundary curve, the colors approach spectral purity, while the closer the points are to W the nearer the colors are to pure white. The x and y coefficients of the standard ICI primaries are $x = 0$, $y = 0$; $x = 1$, $y = 0$; $x = 0$, $y = 1$.

***45.5. Purity and Dominant Wavelength.** The two points G and R in Fig. 45.4 represent a pale green and a bright red color, respectively. All mixtures of these two colors lie on the line GR. Also, each of these colors represents a dilution by white of the wavelength found by drawing straight lines from W through the points G and R to the curve at 5,400A and 6,100A. The ratio of the distance of a given color point from W to the whole distance from W to the boundary curve, measured along one of these straight lines, is called the *purity* σ of the color. For G this ratio WG/Wg is $^{11}\!/_{29} = 38\%$, while, for R, σ is $WR/Wr = ^{16}\!/_{23} = 70\%$. The point of intersection with the boundary of one of these straight lines drawn from W through the color point is called its *dominant wavelength* λ_0.

Thus any color may be represented either by its chromatic coefficients (x,y) or by its purity σ and dominant wavelength λ_0. The (σ,λ_0) designation is the more meaningful of the two. For the particular colors represented by the points G and R in Fig. 45.4 the specifications are

$$G\colon x = 0.28;\ y = 0.46 \qquad \text{or} \qquad \sigma = 38\%;\ \lambda_0 = 5{,}400\text{A}$$
$$R\colon x = 0.56;\ y = 0.32 \qquad \text{or} \qquad \sigma = 70\%;\ \lambda_0 = 6{,}100\text{A}$$

If a color is located on the boundary curve, its $\sigma = 100\%$ and it is completely pure or *saturated*. White is completely impure ($\sigma = 0$), or *unsaturated*.

When two colors can be added to give white, they are called *complementary*. Complementary pure colors lie at opposite ends of a straight line through the white point. For example, 4,800 spectral blue and 5,800 spectral yellow, as indicated in Fig. 45.4, are complementary. But lines from a large segment of the boundary drawn through W do not intersect the boundary a second time but instead strike the dotted line LM. The triangular region LWM contains all the shades of purple and magenta. No colors represented by points in this triangle may be obtained by mixing white with one of the pure spectral colors. Inspection of Fig. 45.4 shows, however, that the purple P and the spectrum color 5,400 could be mixed in proper proportions to match the white light W. Now 5,400 is green, and therefore such a purple is called a *minus green*. These two colors are complementary, 5,400A is the dominant wavelength for P, and its purity is the ratio of WP to the distance from W through P to the line LM.

It is now possible to describe all real colors with this number system, including a number giving the brightness of the color. The latter is possible because the Y curve in Fig. 45.3 conforms to the color-sensitivity curve of the eye. The trichromatic coefficient $\overline{Y}$ is therefore directly proportional to the luminous flux in the sample of light. Actually, the values of $F(\lambda)$ for daylight illuminant are usually given in arbitrary units such that, if the reflectance r_λ of a surface is 1 at all wavelengths,

$$\text{The integral } \overline{Y} = \int_0^\infty Y \times 1 \times F(\lambda) d\lambda = 1$$

Then if the reflectance of a surface is less than unity, $\overline{Y}$ is decreased in the same ratio. That is, $\overline{Y}$ just gives the average reflectance of the surface for the illuminant having these $F(\lambda)$ values. Thus a "baby-pink"-colored surface may now be listed as having $\lambda_0 = 6{,}100$, $\sigma = 10\%$, $\overline{Y} = 50\%$, while a "navy-blue" fabric has $\lambda_0 = 4{,}750$, $\sigma = 20\%$, and $\overline{Y} = 3\%$.

To obtain any color by the proper mixing of three primaries it is obvious from inspection of the chromaticity diagram that the greatest range of colors would be covered by using highly saturated red, green, and blue components. Lines connecting these R, G, and B points would inclose a triangular area, and for any color point C within this area some mixture of R and G could be found such that B, C, and this RG mixture point all lie on a straight line in the chromaticity diagram. That means that a mixture of B and that particular RG mixture in the correct proportions would yield C. But the RGB triangle could not possibly cover the whole area enclosed by the curve of Fig. 45.4. The full range of printing-ink colors covers an area about the white point of only about one-quarter of the total area inside the boundary curve of the chromaticity diagram.

45.6. Colors by Subtraction. Colors may also be mixed to yield new colors by a subtractive process. Thus blue is white minus green and red. The subtracted color or mixture of colors is the complementary of the resulting color. Yellow, which is the mixture of red and green, is the complementary of blue. Thus with a set of three complementary filters—a yellow (minus blue), a red-blue (minus green), and a blue-green (minus red), each of which can subtract a primary from white light—a good range of colors can be produced by subtractive mixture. Thus, when yellow and blue-green are both subtracted from white, green (minus blue and red) results.

The colors obtained with paints and inks result from a subtractive process. White paint is a colorless, transparent liquid, such as linseed oil, containing suspended particles of equally transparent material, such as lead oxide. These particles have an index of refraction considerably different from that of the liquid base, however, and reflection of light occurs from their surfaces. When ordinary white light is incident on a layer of this paint, some is reflected from the surface and the remainder penetrates. This latter portion is partly reflected and partly refracted at the particle surfaces; and since all wavelengths are about equally reflected, the sum total of the light reflected back out of the paint is white.

But if dyed particles are added to the white paint, such particles act as tiny filters. The light reflected in the paint must pass through many of these particles, which will subtract particular colors from the incident white light. If, for example, both blue-green and yellow-dyed particles are mixed into the colorless paint, the red and blue portions of the white light are subtracted, leaving only green to be reflected. Of course, some white light is reflected from the first surfaces, but green will predominate in the total reflected light.

In three-color printing processes the pigments used are the complementaries of the red, green, and blue primary colors. These pigments should then be called green-red (yellow), red-blue, and blue-green. Each of them subtracts one and reflects two of the primaries, the amount of absorbed color being controlled by the concentration of the pigment particles in the base. The ink layer must be thin enough for the light to pass through and be reflected from the white paper underneath. In color printing, subtractive mixture is obtained by printing the inks over each other in half-tone screens, each color being printed by a separate plate. Six different colors, plus black, may be printed on white paper—the three complementary colors minus blue, minus green, and minus red, and the three true primaries, red, green, and blue, which are produced by subtractive mixture of the complementaries. Black is a mixture of all three complementaries, but printers use a separate black-ink plate because the three superposed color layers do not absorb sufficiently to give a good black.

Half-tone printing places each color on the page in tiny, closely spaced dots. The solid colors seen by the eye are an additive mixture produced in the visual mechanism itself. The eye, being unable to distinguish the tiny half-tone screen dots, mixes the colored light that they reflect into a single color.

PROBLEMS

*1. Locate with the aid of Fig. 45.4 the approximate pure color complementary to the hydrogen C line (6,563A), the sodium D line (5,893A), and the hydrogen G line (4,340A).

*2. Three colored glasses R, G, and B produce light having the following trichromatic coefficients:

Glass	x	y
R	0.60	0.30
G	0.10	0.70
B	0.20	0.20

Copy the chromaticity diagram, locate these colors, and indicate the location of all colors that can be matched by mixtures of R and G, B and G, and R and B. Also give the location of colors that if mixed with R can be matched by a mixture of B and G.

***3.** Compute from your chromaticity diagram for Prob. 2 the dominant wavelength and purity of the three colors.

4. Why is a block of ice transparent, while snow is opaque and white? Why does a dark-blue suit seem black when viewed in candlelight?

CHAPTER 46
INTERFERENCE

46.1. Wave Phenomena. Interference, diffraction, and polarization are phenomena that can be explained only if light is a wave motion. Just as two trains of ripples on the surface of a liquid may pass through each other and proceed onward without any apparent effect on each other, so two beams of light may cross each other without modification once the crossing is made. In the region of crossing, however, as for all types of wave motion (Chap. 37), the resultant displacement at every point and at every instant is the algebraic sum of the displacements of the medium from the separate wave trains. The modifications produced when two or more beams of light are superposed are termed *interference*. If the resultant amplitude and intensity are less than that to be expected from one of the beams alone, the interference is *destructive*, while if the resultant intensity is greater it is *constructive* interference. Since light waves are electromagnetic, the displacement in this case means the instantaneous strength of the electric or the magnetic field.

If the two superposed wave trains are *in phase*, the resultant electric field is a maximum. For water waves this means that crest falls on crest and trough on trough. If, on the other hand, the crest from one wave train falls on the trough from the other, the two are *out of phase* and the resultant disturbance is a minimum. Because of the very short wavelength, however, interference between two light beams is not easily observed. It was first demonstrated in 1801 by Thomas Young, a London physician. This work proved the wave nature of light; but, because of the great popularity of the corpuscular theory during the preceding century, people generally were not convinced until the additional evidence from Fresnel's experiments a few years later made the corpuscular theory untenable.

46.2. Young's Experiment. Young allowed sunlight to pass through a pinhole S and then at some distance through two pinholes S_1 and S_2 (Fig. 46.1). In the modern version of this experiment, narrow slits replace the pinholes, and the first slit S is illuminated with monochromatic light such as the yellow-orange sodium light from a flame or a sodium-vapor lamp. By Huygens' principle the slits become new sources from which cylindrical wave fronts, shown in Fig. 46.1, proceed toward the right. If the solid circular lines represent crests and the dashed circular lines troughs, at any one instant the solid lines intersecting the screen AB at P_0 and P_2 are the loci of points of constructive interference (crest on crest, trough on trough), while the dashed lines striking the screen at P_1 connect points of destructive interference (crest on trough). Parallel interference bands or fringes alternately light and dark, such as those shown in Fig. 46.2, are observed on the screen.

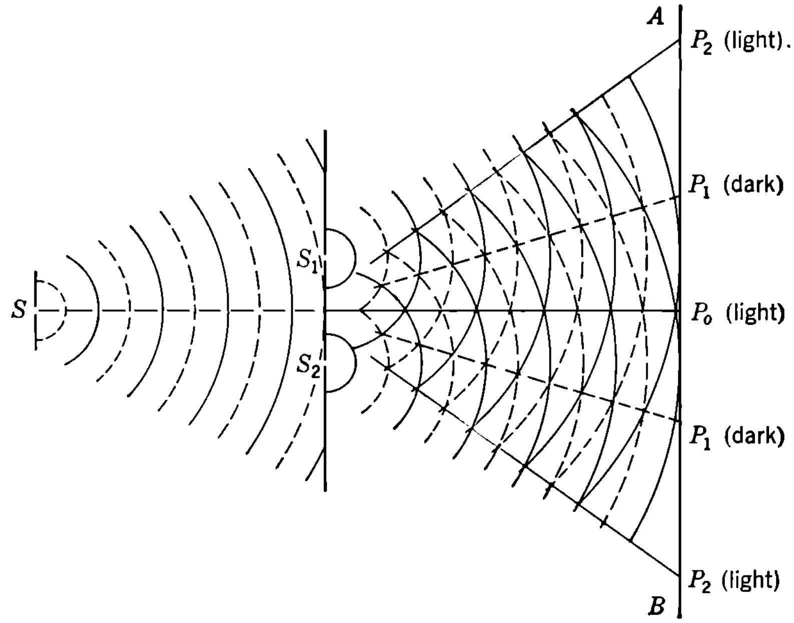


FIG. 46.1. Young's double-slit experiment.

A convenient way to perform this experiment is to use a lamp with a single straight vertical filament as the source S. This is observed at a distance of a few feet through a pair of fine slits scratched in the emulsion of an undeveloped photographic plate. The double slits, which are held just in front of the eye, should be 1 mm or less apart; the smaller their

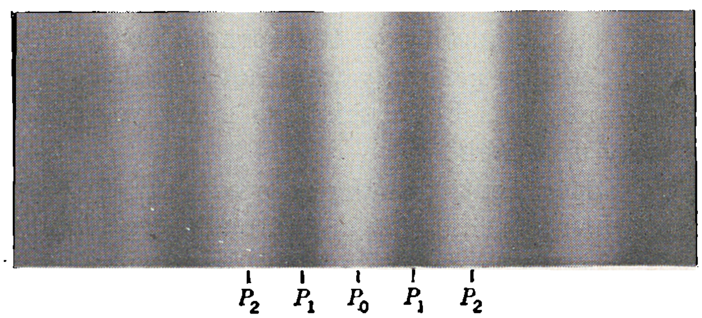


FIG. 46.2. Photograph of double-slit interference fringes.

separation, the wider the spacing of the fringes. If first a red filter and then a green filter are placed in front of the lamp, it will be noticed that the red light produces wider fringes than the green. This, as we shall now demonstrate, arises from the longer wavelength of the red light.

The intensity of the light at any point P on the screen (Fig. 46.3) depends upon the phase difference between the two waves arriving at P from the two slits. We assume that the source S is equidistant from S_1 and S_2 so that the light vibrations at the two slits are in the same phase at any instant. If λ is the wavelength, the number of wavelengths in the distance $\overline{S_1P}$ is $\overline{S_1P}/\lambda$. Since the phase in a wave train increases by 2π radians in each wave length, at an instant when the phase of the wave train is zero at S_1, its phase at P is $2\pi(\overline{S_1P}/\lambda)$. Similarly the phase at P of the wave train from S_2 at the same instant is $2\pi(\overline{S_2P}/\lambda)$. The phase difference δ between the two wave trains is

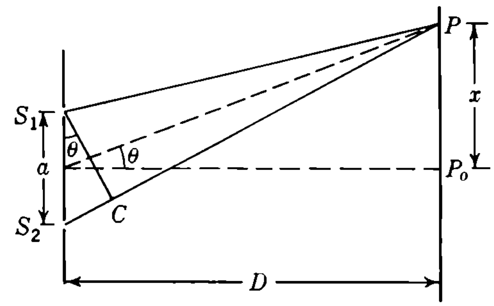


FIG. 46.3. The optical paths in Young's experiment.

$$\delta = \frac{2\pi}{\lambda}(\overline{S_2P} - \overline{S_1P})$$

If P is a point of maximum intensity, δ must be some integral multiple of 2π,

or $$\delta = 2n\pi = \frac{2\pi}{\lambda}(\overline{S_2P} - \overline{S_1P})$$

Hence $$\overline{S_2P} - \overline{S_1P} = n\lambda \qquad \text{(for maxima)} \qquad (46.1)$$

where $n = 0, 1, 2, \ldots .$ If the intensity at P is a minimum, δ must be some odd multiple of π, or

$$\delta = (2n + 1)\pi = \frac{2\pi}{\lambda}(\overline{S_2P} - \overline{S_1P})$$

Therefore $$\overline{S_2P} - \overline{S_1P} = (n + \tfrac{1}{2})\lambda \qquad \text{(for minima)} \qquad (46.2)$$

In Fig. 46.3, S_1C is an arc struck from P as a center with $\overline{PS_1}$ as the radius. Thus, $\overline{S_2C} = \overline{S_2P} - \overline{S_1P}$; and since the distance D is very large as compared with the distance a between slits, $\overline{S_1C}$ may be considered a straight line perpendicular to $\overline{S_2P}$, CS_1S_2 is a right triangle, and

$$\overline{S_2C} = a \sin \theta.$$

Also, the distance $\overline{P_0P}$ or x from the central bright fringe equals $D \tan \theta = D \sin \theta$, since θ is small. If P lies at the center of a bright fringe, then

$$\overline{S_2C} = a \sin \theta = n\lambda$$

$$a\frac{x}{D} = n\lambda$$

$$\lambda = \frac{ax}{nD} \qquad (46.3)$$

where $n = 1$ for the first bright fringe on either side of the central maximum P_0, $n = 2$ for the next bright fringe on either side, and so on. Thus the wavelength of the light may be determined from measurement of the distances x, a, and D.

Determinations of λ by this method yield about 0.000065 cm, or 6,500A, for red light and 0.000043 cm, or 4,300A, for violet light. If white light is used, only the central fringe at P_0 is white, for all wavelengths give a maximum with $n = 0$ at that point. Since x varies with λ for all the other fringes, they are colored and only a few fringes are observable because of overlapping. As an example of this effect, let us compute the distance x along the screen from P_0 to the fourth maximum of $\lambda = 4{,}300$A and the third for 6,500A, with $a = 0.5$ mm and $D = 1$ m. Then,

$$x = \frac{4 \times 4.3 \times 10^{-4}\text{ mm} \times 10^3\text{ mm}}{0.5\text{ mm}} = 3.44\text{ mm}$$

for the fourth violet fringe, while

$$x = \frac{3 \times 6.5 \times 10^{-4}\text{ mm} \times 10^3\text{ mm}}{0.5\text{ mm}} = 3.90\text{ mm}$$

for the third red fringe.

It is thus apparent that already the colored fringes are beginning to overlap. In order to observe a considerable number of fairly sharp fringes as depicted in Fig. 46.2 it is essential that the light be monochromatic.

***46.3. Other Examples of Interference.** Soon after Young performed his experiment, it was thought by some that the bright fringes resulted from a modification of the light by the edges of the slits. A. J. Fresnel (1788–1827), about 1815, convincingly demonstrated the interference between two light beams by some novel experiments. In the Fresnel double-mirror experiment (Fig. 46.4), light from the narrow slit S falls on two mirrors M and M', whose planes make a small angle with each other. In the region P_1P_2 of the screen where the two reflected beams overlap, interference fringes are observed, for the light appears to come from the two virtual sources S_1' and S_2'. The explanation of the fringes is very similar to that of the double-slit experiment.

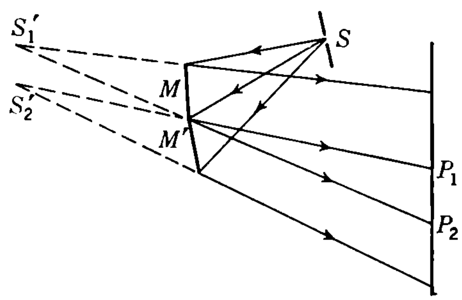


FIG. 46.4. Fresnel double mirror.

Fresnel also produced interference fringes with a biprism (Fig. 46.5), which is equivalent to two prisms of very small angle (about 1°) mounted base to base. The light from the slit S is deviated slightly, so that in the region of overlap P_1P_2

on the screen the light appears to come from two virtual sources S_1' and S_2', with the result that interference fringes are observed in this region.

From measurements on the interference fringes produced by either of these methods the wavelength of the light may be determined. The fringes are much brighter than with the double-slit experiment because of the much wider cone of light transmitted (or reflected).

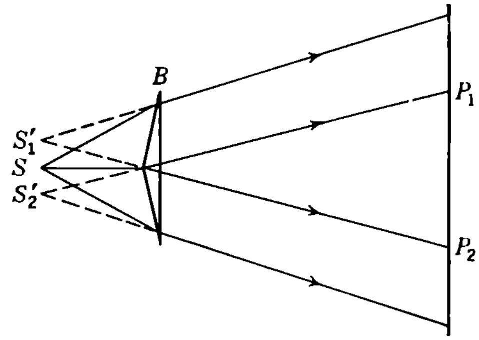


FIG. 46.5. Fresnel's biprism.

It is to be noted that in all three of these methods of demonstrating interference of light the interfering beams come from a common source. It is impossible to obtain interference fringes from two separate sources, such as two lamp filaments side by side. The reason for this is that each emitting atom in the source sends out pulses of light of very short duration (about 10^{-8} sec); hence the phase of the light jumps erratically from pulse to pulse. If the light comes from two sources, interference fringes exist on the screen for such extremely short intervals but will be shifting their position continually and as a result no fringes at all will be observed. But if the light from S_1 and S_2 comes from the same source S, then phase shifts occur simultaneously in both, the difference in phase between any pair of points in the two sources remains constant, and the interference fringes are stationary. These two sources are then said to be *coherent sources.*

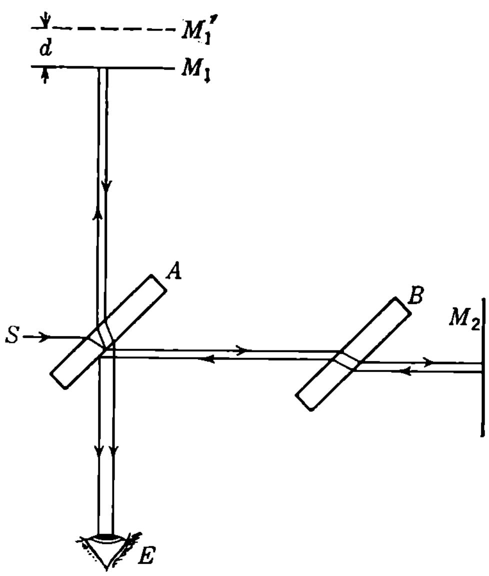


FIG. 46.6. The Michelson interferometer.

***46.4. The Michelson Interferometer.** This instrument, in which interference fringes are produced for the purpose of making extremely accurate measurements of lengths and angles, is shown in Fig. 46.6. The plane-parallel plate of glass A is "half-silvered" on its rear side; *i.e.*, it is given a silver or aluminum coating of such thickness that about half the light incident on this surface is reflected, the other half being transmitted. Light from point S of a source such as a sodium flame falls on this surface and is in part reflected to the mirror M_1 and in part transmitted to the mirror M_2. From these mirrors the light is reflected back and into the eye as shown. Since the light in traveling the path A to M_1 and back passes through the plate A three times, the similar but unsilvered plate B is usually introduced into the other path to render the two paths *in glass* the same. Without this compensating plate B, fringes would still be produced by interference between the light beams that have traveled the two paths provided that the light is monochromatic. With white light, however, no fringes will be seen if the difference between the two paths exceeds a few wavelengths. If M_1 and M_2 are exactly perpendicular to each other and at 45° with plate A, and if the distances from the back surface of A to the two mirrors are exactly equal, the eye sees M_1 directly and a reflected image of M_2 superposed. There is then no interference between the two beams.

If now M_1 is moved normally to its surface a distance $\lambda/4$, the ray that travels this path has gone $\lambda/2$ farther than the ray in the other path when they combine to enter the eye. This path difference means a phase difference of π radians, and hence destructive interference in the center of the field of view. Further movement of M_1 by an amount $\lambda/4$ restores the light. Therefore, by slowly displacing M_1 and counting the number N of times the light reaches a maximum, the distance d through which M_1 has been moved is given by

$$d = \frac{N\lambda}{2}$$

This distance may be measured to a fraction of a wavelength of the light.

The interference here is between the rays reflected from the two sides of the "air film" between M_1 and the image of M_2 in A. When the rays at the center of the field of view show zero phase difference, other rays will in general not be in phase and there will be circular fringes about the central part. These fringes change in diameter as M_1 is moved. If M_1 and M_2 are not quite perpendicular to each other, the air film is wedge-shaped and the fringes are still circular but very eccentric. As M_1 is moved, one counts the number of curved fringes that pass any point in the field of view. •

With this interferometer Michelson measured the length of the standard meter in Paris in terms of the wavelength of the red spectrum line emitted by a cadmium-vapor lamp. The final mean result of his and later measurements is

$$1 \text{ m} = 1{,}553{,}164.1\lambda \qquad \text{or} \qquad \lambda = 6{,}438.4696\text{A}$$

The accuracy of this result is about 1 part in 2 million. Since such spectral wavelengths are probably the most permanent and unchangeable standards of length that it is possible to obtain, the standard meter could now be accurately reproduced if ever destroyed. With an interferometer of this type, numerous other very precise measurements may be carried out.

46.5. Interference in Thin Films. Beautiful color effects are caused by interference resulting from multiple reflections of white light between the two surfaces of a thin transparent film. Common examples are the colors of thin films of oil on water, and of soap bubbles. We shall examine this type of interference phenomenon in some detail for the case of reflection of monochromatic light from a film with plane-parallel sides. Figure 46.7 illustrates the multiple reflections that may occur when a ray of light is incident on such a film. At A this ray is partly reflected and partly refracted. At B the refracted ray is in part reflected back to the top surface and in part refracted out of the film. At C a portion is refracted to form ray 2, and the remainder is refracted to D. The repetition of this process produces a set of parallel rays on each side of the film, but with the intensity decreasing rapidly from one ray to the next in each set. If the set of parallel reflected rays is brought to a focus by a lens, the phase relations between the rays, which have all traveled different distances, determine whether there is constructive or destructive interference at P.

To find this phase difference we must calculate the difference in *optical path* for two successive rays, such as 1 and 2. The optical path of a

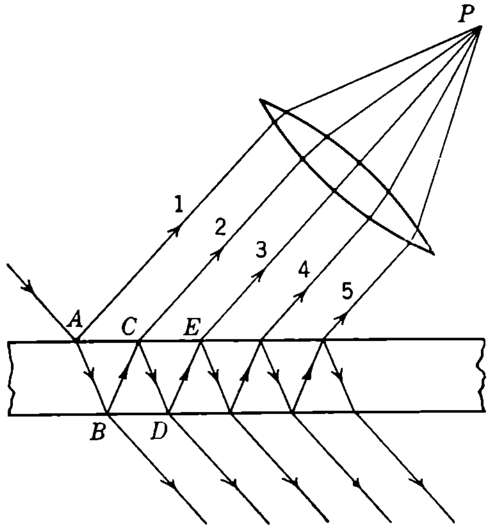


FIG. 46.7. Multiple reflections in a thin film.

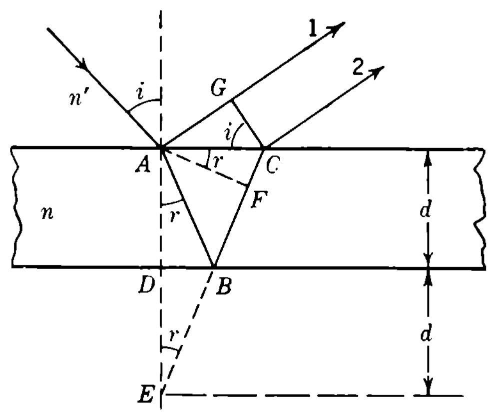


FIG. 46.8. Optical path difference between two reflected rays.

ray of length l in a medium of index n is defined as the product nl. It is the distance in air containing the same number of wavelengths as the actual geometrical path in the medium. In Fig. 46.8 we let d be the thickness of the film, n its index of refraction, and λ the wavelength of the light. If CG is perpendicular to ray 1, the optical paths from C and G to the focus of the lens are equal. While ray 1 travels the path AG in the upper medium of index n', ray 2 follows the path ABC in the film. The difference Δ in these optical paths is given by

$$\begin{aligned}\Delta &= n(AB + BC) - n'(AG)\\ &= n(EC) - n'(AG)\\ &= n(EF + FC) - n'(AG)\end{aligned}$$

$$\frac{AG}{FC} = \frac{\sin i}{\sin r} = \frac{n}{n'}$$

Hence $$nFC = n'AG$$

Therefore, $$\Delta = n(EF) = n(2d \cos r) \qquad (46.4)$$

It might be expected that, if this path difference is an integral number of wavelengths, rays 1 and 2 would be in phase and hence would interfere constructively. But ray 1 is reflected at a denser medium and hence undergoes a phase change of π (Sec. 37.12). Ray 2, however, is reflected from a less dense medium, so that no phase change occurs. Hence the relation

$$2nd \cos r = N\lambda \qquad \text{(minima)} \qquad (46.5)$$

where $N = 0, 1, 2, \ldots$, is the condition for *destructive* interference of rays 1 and 2.

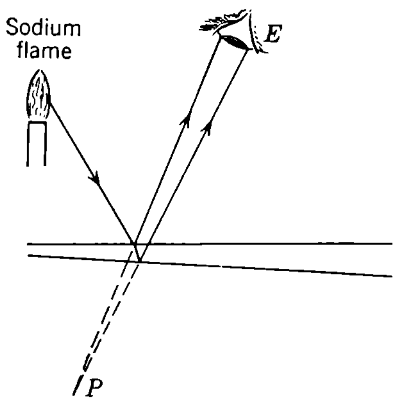


FIG. 46.9. Interference fringes from wedge-shaped film.

Because the distances and angles are the same, Eq. (46.4) also gives the path difference between rays 3 and 2 and between all succeeding pairs of reflected rays. But for all these rays the reflection is from a less dense medium, as for ray 2, and hence rays 3, 4, . . . are in phase with ray 2 and out of phase with ray 1. For the minima of intensity given by Eq. (46.5) it can be shown that complete darkness results. Although ray 2 is less intense than ray 1, the addition of rays 3, 4, 5, . . . , which are all in phase with ray 2, is just sufficient to make up the difference. On the other hand, if

$$2nd \cos r = (N + \tfrac{1}{2})\lambda \qquad \text{(maxima)} \qquad (46.6)$$

rays 1 and 2 should be in phase but 3 is out of phase with 2, 4 is out of phase with 3, etc. Since ray 2 is more intense than 3, 4, 5, . . . , these pairs cannot cancel each other and their net effect combined with the strongest ray 1 is to give a maximum of intensity. Since color is a function of the wavelength λ and since all wavelengths are present in white light, Eq. (46.6) shows that white-light interference fringes should exhibit a considerable variety of colors from violet to red. For a particular path difference, some color is always missing by destructive interference. Therefore the color observed is the result of a subtractive process.

If the two plane surfaces of the film make a small angle with each other (Fig. 46.9), the interfering rays entering the eye appear to diverge from a point near the film. The fringes then are similar to those seen in the Michelson interferometer when its two mirrors are not exactly perpendicular to each other. The fringes, which appear to be formed in the film itself, are practically straight and parallel to the thin edge of the wedge-shaped film. Then the path difference for any pair of rays is practically the same as that given by Eq. (46.4); and if one observes close to the normal to the film, cos r is approximately 1, so that for bright fringes

$$2nd = (N + \tfrac{1}{2})\lambda \tag{46.6'}$$

From one fringe to the next, N increases by 1, and hence the optical thickness nd of the film changes by $\lambda/2$.

These thin-film fringes are easily obtained by taking two pieces of ordinary plate glass placed together with a strip of paper between them at one edge. The film is thus of air with $n = 1$. If a sodium flame is then viewed as in Fig. 46.9, orange-yellow fringes are clearly seen. It is essential that the film be thin and that the source be broad if an extended system of fringes is to be seen.

Such fringes are used in testing optical surfaces for planeness. If the surface being polished is placed in contact with a surface known to be flat, the air film between the two viewed in monochromatic light will exhibit fringes of irregular shape if the new surface is not flat. A given fringe, being characterized by a particular value of N, in Eq. (46.6), will follow those parts of the film where d is constant. Thus the fringes form *contour lines*, showing the hills and valleys of the uneven surface. Since for this film $n = 1$, the interval between contours is $d = \lambda/2$. When, after continued polishing of the high spots, this test shows the fringes to be straight, the new surface is also known to be optically flat.

46.6. Newton's Rings. Newton described in his "Opticks" an experiment in which he placed the convex side of a plano-convex lens of long focus against a flat glass surface and observed, in the light reflected from the air film, colored rings concentric about the point of contact. He performed thus a striking demonstration of the interference of light, but he nevertheless continued to advocate the corpuscular theory.[1] If monochromatic light is used, many sharp fringes are seen both in reflected and transmitted light. Figure 46.10 is a photographic reproduction of these rings in reflection (eye at E_1 in Fig. 46.11) with monochromatic light. *The rings are loci of equal path difference in the air film.* If white light is employed, only a few of the circular fringes are observed, all highly colored except for the central spot, which is black in the reflected light. Colored rings are likewise observed in the transmitted light (eye at E_2 in Fig. 46.11), but they are not so brilliant as those due to reflected light because of an admixture of unmodified white light. The colors in the

[1] In his discussion of this experiment Newton came close to the wave hypothesis when he wrote about the "Fits of easy Reflection," the "Fits of easy Transmission," and the "Interval of its Fits," referring to the multiple reflections and refractions of a ray at the two surfaces.

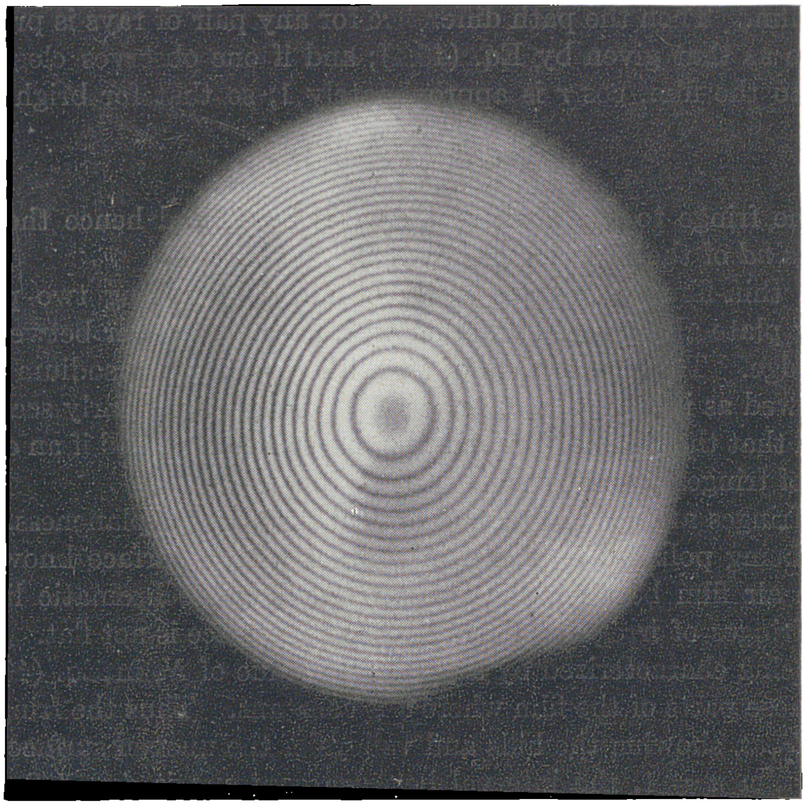


FIG. 46.10. Newton's rings in reflected light.

two sets of rings are complementary, for at a given point the color in the transmitted light is white minus the color most strongly reinforced in the reflected light at that same point.

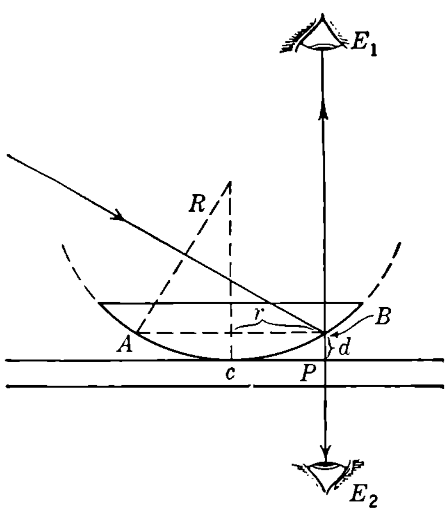


FIG. 46.11. The formation of Newton's rings.

To derive the relation between the radius r of a ring, λ, and the radius of curvature R of the convex surface, consider the arrangement of Fig. 46.11. The thickness d of the air film at a distance r from the point of contact c is the *sagitta* (Sec. 40.7) of the arc AB. Then

$$d = \frac{r^2}{2R - d}$$

but this may be written

$$d = \frac{r^2}{2R} \tag{46.7}$$

since d is but a small fraction of a millimeter, while R is of the order of

1 m. A bright fringe will be observed at P if, according to Eq. (46.6),

$$2nd = (N + \tfrac{1}{2})\lambda$$

for again $\cos r \cong 1$ since we observe close to the normal to the film. Substituting this value of d in Eq. (46.7), we find

$$\frac{(N + \frac{1}{2})\lambda}{2n} = \frac{r^2}{2R}$$

or

$$r^2 = \frac{R(N + \frac{1}{2})\lambda}{n} \qquad \text{(bright rings)} \qquad (46.8)$$

For the dark fringes, from Eq. (46.5),

$$r^2 = \frac{RN\lambda}{n} \qquad \text{(dark rings)} \qquad (46.9)$$

The index n is closely 1 for air, and hence it is necessary only to measure r and R in order to calculate λ. Since there will be some distortion of the surfaces at the point of contact, it is better to measure two radii, say r_N of the dark ring N and r_{N+s} of the dark ring $N + s$. Then, from Eq. (46.9) with $n = 1$,

$$r_{N+s}^2 - r_N^2 = R(N + s)\lambda - RN\lambda = Rs\lambda$$

Hence

$$\lambda = \frac{r_{N+s}^2 - r_N^2}{Rs} \qquad (46.10)$$

If it were not for the relative phase change of π radians between rays reflected from the upper and lower surfaces of the air film, the central spot of the reflected Newton ring system would be bright rather than black. For, at the central contact point, $d = 0$; and hence, without this phase change, the rays reflected from the two surfaces would interfere constructively. The central spot in the transmitted system is bright.

***46.7. Nonreflecting Films.** An important modern application of this phenomenon of interference is the coating of optical surfaces with nonreflecting films. This serves, for example, to reduce the loss of light by reflections in multisurface optical systems. A thin layer of transparent material of index intermediate between that of air and glass is deposited on the glass surface (Fig. 46.12). Since at both the top and bottom surfaces of the film the reflection is from a denser medium, the same phase change of π radians occurs at both reflections. But if the optical film thickness nd is made just one-quarter wavelength, then for normal incidence rays 1 and 2 will be exactly π radians out of phase with each other and therefore destructive interference between them results. Since the nonreflecting film can have the correct thickness for one wavelength only, a wavelength in the yellow-green where the eye has its maximum sensitivity is chosen. Some reflection then occurs for both longer and shorter wavelengths, and hence the slight residual reflected light has a purple hue when white light is incident.

This treatment of lens and prism surfaces reduces the loss of light by reflection at each surface from 4 or 5 per cent to a fraction of 1 per cent. When there are many air-glass surfaces in succession, as in range finders, periscopes, and fast camera lenses, it is now common practice to place such films on all surfaces to decrease light loss. Also, glare due to reflections from the lens surfaces is so

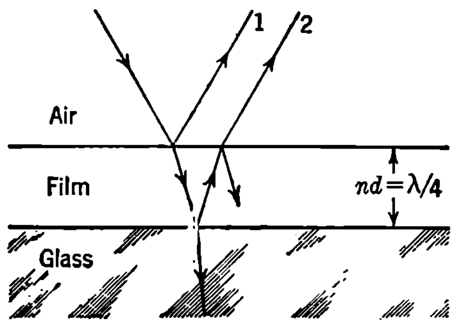


FIG. 46.12. Nonreflecting film. Rays 1 and 2 cancel each other by destructive interference.

reduced by this coating process that photographs may then be taken with the camera pointed almost directly into the sun.

PROBLEMS

1. If sodium light of wavelength 5,893A falls on a double slit of separation 0.2 mm, what will be the distance between successive interference fringes on a screen 1 m away?

2. Interference fringes formed on a screen 150 cm from a double slit of separation 0.5 mm are measured to be 1.5 mm apart. Calculate the wavelength of the light. What is its color?

3. In moving the mirror M_1 of a Michelson interferometer a distance of 0.5 mm, 1,910 fringes are counted. Calculate the wavelength of the light, and state its color.

4. A wedge-shaped air film is formed by slipping a strip of paper at one edge between two slabs of plate glass. Interference fringes are produced with sodium light λ 5,893A and are viewed normally to the surface. If there are 10 fringes per centimeter, calculate the angle of the wedge.

5. If in a set of Newton's rings the tenth bright ring for sodium light λ 5,893A is 18 mm in diameter, what is the radius of curvature of the surface of the plano-convex lens? Take $n = 1$ for air.

6. In a Newton-ring experiment, R for the lens was 733 cm, the radius of the Nth dark ring was 4 mm, and that of the $(N + 5)$th dark ring was 6 mm. Calculate the wavelength of the light used and the ring number N.

7. The air film in a crack in glass may be extremely thin. Will it then appear bright or dark by reflected light? Why? What will be the appearance of a very thin $(d \ll \lambda)$ liquid film (*a*) in air, (*b*) between two glass surfaces? Why?

8. The surfaces of a glass lens of index 1.50 are made nonreflecting by coating with a thin layer of material of index 1.22. Calculate the correct thickness of the layer for $\lambda = 5{,}500$A. What is the phase difference between light reflected normally from the first and second surfaces of the layer, if (*a*) $\lambda = 4{,}000$A and (*b*) $\lambda = 7{,}000$A?

CHAPTER 47
DIFFRACTION

47.1. Diffraction Patterns. When light passes through a small opening or by the edge of an obstacle and then falls on a screen, it is found that some of the light has been deviated into the region of the geometrical shadow. We have already noted that this phenomenon of *diffraction* is to be expected according to Huygens' principle. In the present chapter we shall study in some detail the *diffraction patterns*, or distributions of light intensity on the screen. When the wave fronts are limited by a narrow slit or a straight edge, diffraction bands are observed on the screen, or, if the wave front is cut off by a round hole or obstacle, circular, alternately dark and bright rings occur in the patterns. These diffraction effects result from interference of all the secondary wave trains arising at every point in the wave fronts where these pass through the aperture. Their study contributes not only to our knowledge of the nature of light but also to an understanding of the functioning of optical instruments. The images formed by a telescope or microscope, for example, are diffraction patterns, for the objective lens is a limiting aperture.

It is convenient to divide diffraction phenomena into two groups, (1) those in which both the source of light and the screen are effectively at an infinite distance from the diffracting aperture and (2) those in which the source or the screen or both are at finite distances from the aperture. The first group is called *Fraunhofer diffraction*, while the second is known as *Fresnel diffraction*. Fraunhofer diffraction is easily arranged experimentally by using one lens to render the light incident on the aperture plane-parallel and another lens to focus the light from the aperture onto the screen. This type is also simpler to treat theoretically. To observe Fresnel diffraction, only a small source, the diffracting obstacle, and a screen are needed, but the fact that the light is divergent makes the theoretical treatment more difficult. We shall limit our discussion to a few of the simpler cases of diffraction.

***47.2. Fresnel's Half-period Zones.** In Fig. 47.1, AB represents an instantaneous section of a plane wave of monochromatic light traveling toward the right. We wish to find the resultant effect at P of all the secondary wavelets originating at every point in the present position of the wave. As a useful approximation we divide the wave front into a number of zones of finite area in the following manner: circles of radii r_1, r_2, r_3, . . . , r_n are described about O, so that each circle is half a wavelength farther from P. If the distance $OP = R$, the circles are at distances $R + \lambda/2$, $R + 2\lambda/2$, $R + 3\lambda/2$, . . . , $R + n\lambda/2$ from P. The areas between successive circles are called *half-period zones.* Since each zone is on the average $\lambda/2$ farther from P than the one just inside it, the secondary wavelets emanating from successive zones arrive at P with an average phase difference of π.

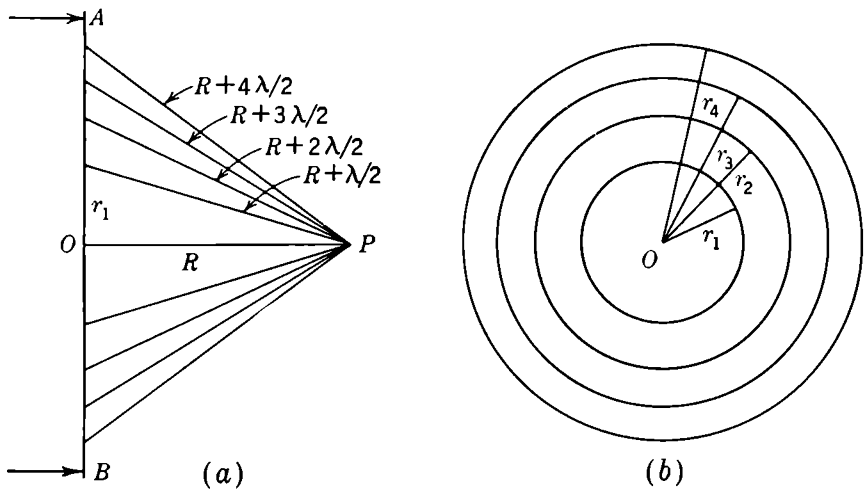


FIG. 47.1. Half-period zones constructed on plane wave front AB.

The areas of all the zones are approximately equal. For in two successive right triangles of Fig. 47.1a,

$$r_n^2 + R^2 = \left(R + \frac{n\lambda}{2}\right)^2$$

and

$$r_{n-1}^2 + R^2 = \left[R + \frac{(n-1)\lambda}{2}\right]^2$$

These two relations yield, for the areas of the circles of radii r_n and r_{n-1},

$$\pi r_n^2 = \pi\left(R + \frac{n\lambda}{2}\right)^2 - \pi R^2$$

$$\pi r_{n-1}^2 = \pi\left[R + \frac{(n-1)\lambda}{2}\right]^2 - \pi R^2$$

Squaring the terms in brackets and subtracting the second of these equations from the first give for the area of the zone

$$\pi r_n^2 - \pi r_{n-1}^2 = \pi R\lambda + \pi\frac{(n - \frac{1}{2})\lambda^2}{2} \tag{47.1}$$

If R is large compared with the wavelength λ, as it is for all the cases we shall discuss, the term in λ^2 may be neglected. Hence

$$\text{Area of any zone} = \pi R\lambda \tag{47.2}$$

Because of this equality of the areas the amplitudes of the secondary wave trains contributed by each zone to P are very nearly equal.

Denote by $a_1, a_2, a_3, \ldots, a_n$ the amplitudes at the point P of the disturbances from successive zones. These amplitudes slowly diminish with increasing radius of the zone because of the increasing obliquity and because of the increasing distance from the zone to P. The latter effect is just counterbalanced by the slight increase in zone area with n. Furthermore they alternate in sign, since changing the phase by π means reversing the direction of the amplitude. (We are here using amplitude as synonymous with maximum displacement that alternates in sign, whereas, in our previous discussions of wave motions, amplitude has denoted just the magnitude of the maximum displacement.) Therefore the resultant amplitude A at P may be written as the sum

$$A = a_1 - a_2 + a_3 - a_4 + a_5 - \cdots (-1)^{n-1}a_n \tag{47.3}$$

Taking n to be odd, and using a procedure first employed by Lord Rayleigh (1842–1919), we may evaluate this series by grouping its terms in the following two forms:

$$A = \frac{a_1}{2} + \left(\frac{a_1}{2} - a_2 + \frac{a_3}{2}\right) + \left(\frac{a_3}{2} - a_4 + \frac{a_5}{2}\right) + \cdots + \frac{a_n}{2}$$

$$\text{or } A = a_1 - \frac{a_2}{2} - \left(\frac{a_2}{2} - a_3 + \frac{a_4}{2}\right) - \left(\frac{a_4}{2} - a_5 + \frac{a_6}{2}\right) - \cdots - \frac{a_{n-1}}{2} + a_n$$

Since the amplitudes $a_1, a_2, \ldots$ decrease slowly at first, then more rapidly, each amplitude is greater than the arithmetic mean of the preceding and following amplitudes. Therefore the quantities in parentheses in the above equations are all negative, and consequently

$$\frac{a_1}{2} + \frac{a_n}{2} > A > a_1 - \frac{a_2}{2} - \frac{a_{n-1}}{2} + a_n$$

But a_1 very nearly equals a_2, and a_{n-1} very nearly equals a_n,

Hence
$$\frac{a_1}{2} + \frac{a_n}{2} \cong A \cong \frac{a_1}{2} + \frac{a_n}{2} \tag{47.4}$$

The same result is obtained if n is even. Therefore the resultant amplitude at P due to n zones is half the sum of the amplitudes produced by the first and last zones. If the wave front is not obstructed, n approaches infinity, a_n becomes negligible, and $A = a_1/2$. That is, the amplitude at P due to the whole wave of infinite extent is *only half of that which would be produced by the first zone acting alone.* Although every point on the large wave front sends out secondary waves in all directions, almost all these waves annul each other by interference. Since the energy transported by a wave disturbance per square centimeter per second, or the intensity, is proportional to the square of the amplitude, just as in acoustics (Sec. 38.4), the intensity of the light at P is then only one-fourth that which would be produced by the central zone acting alone.

***47.3. Fresnel Diffraction by a Circular Aperture.** Suppose that a diaphragm with a small circular opening at O (Fig. 47.2) blocks off the wave. We are interested in the effect of the opening on the intensity. If the aperture has a radius r_1 just equal to that of the first half-period zone, the amplitude at P will be a_1, or twice the amplitude from the entire unobstructed wave. If now the radius of the opening is increased until it includes the first two zones, the amplitude at P falls to $a_1 - a_2$, or practically zero. Thus increasing the size of the opening *decreases* the intensity of the light at P. Further increase in r causes the intensity at P to pass through maxima and minima when the number of zones included is, respectively, odd or even.

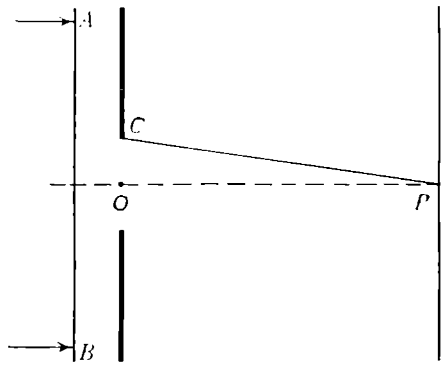


FIG. 47.2. Fresnel diffraction by a circular aperture.

As P is moved continuously along the axis toward or away from the opening, the intensity at P alternates through maxima and minima. This alteration in the distance to the screen changes the size of the zones; for if originally $PC - PO$ in Fig. 47.2 is $\lambda/2$, moving P toward the aperture first increases this path difference to $2\lambda/2$ (two zones included), then to $3\lambda/2$ (three zones), etc. Maxima and minima thus occur along the axis of the aperture. This emphasizes the point that the number and sizes of the zones in a given limiting aperture are not fixed but depend upon the distance to the observing point.

The theory of the intensity variations for observing points off the axis is too complex to be included here. A detailed treatment shows that the point P is surrounded by a system of circular diffraction fringes. These may be observed on a screen or photographed at some distance behind circular holes of different

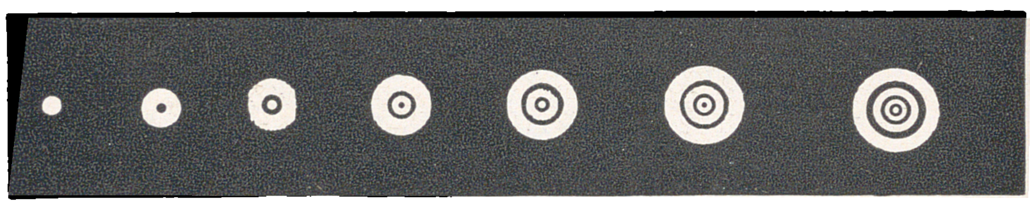


FIG. 47.3. Fresnel diffraction by circular openings exposing one to seven zones.

sizes illuminated by monochromatic light from a distant point source. The appearance of these patterns as the holes expose one, two, . . . seven zones is sketched in Fig. 47.3. Note the alternation of the center of the pattern from bright to dark. Some light also reaches the screen in the region of the geometrical shadow, but this quickly drops to practically zero intensity with increasing distance from the axis.

***47.4. Diffraction by a Circular Obstacle.** If the circular aperture is replaced by a circular disk so that a few half-period zones as viewed from P are covered, the method of Fresnel leads to the conclusion that there should be a bright spot in the center of the shadow. For suppose that the disk obstructs the first three zones. Then the resultant amplitude at P will be $A = a_4 - a_5 + a_6 - a_7 +$

. . . $= a_4/2$ nearly, as can be shown by an analysis similar to that leading to Eq. (47.4). The intensity at P is then proportional to $a_4^2/4$. Since a_4 differs very little from a_1, the intensity at the center of the shadow will be nearly as intense as if the disk were removed. This holds only for a point on the axis, however, while off the axis the intensity is small, with faint concentric rings showing if the obstacle is small. The complete analysis of this diffraction pattern shows that there should also be bright circular fringes surrounding the shadow of the disk.

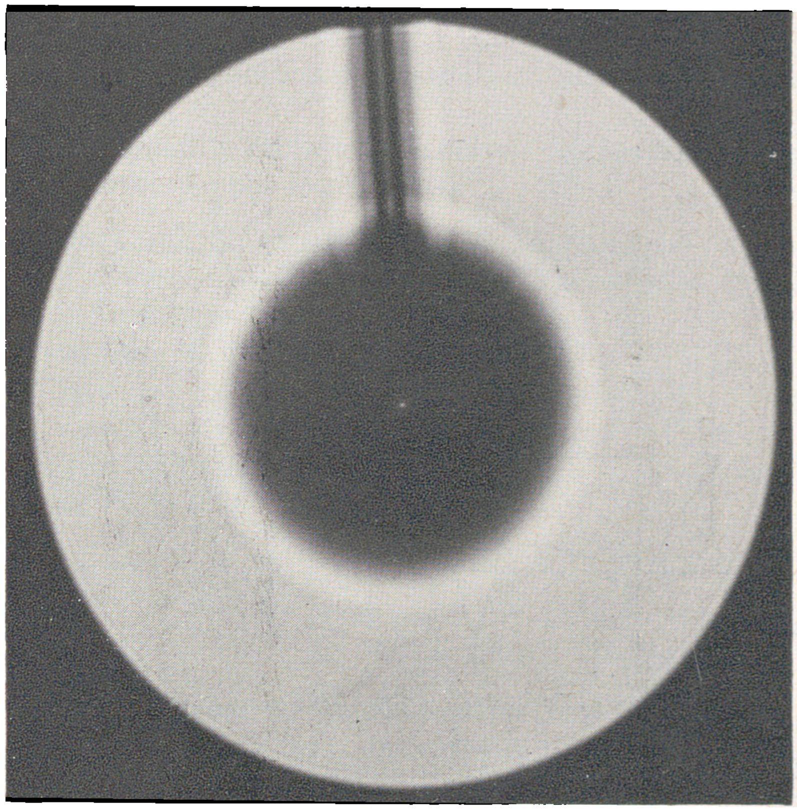


FIG. 47.4. Shadow of a small steel ball. Note the spot of light at the center.

This diffraction pattern may be observed by examining the shadow cast by a small ball bearing mounted in the parallel beam from an arc light. It is well to use a magnifier in order to see clearly all the details of the pattern. Figure 47.4 is a photograph of this interesting diffraction effect.

47.5. Diffraction by a Single Slit. The Fresnel diffraction produced by an aperture in the form of a slit may be studied by using the device of dividing the wave surface in the plane of the opening into strip zones. The diffraction pattern on the screen consists of alternate bright and dark bands, and the center of the pattern may be either bright or dark. As the slit becomes narrower, or the screen more distant, a situation will be reached where the opening is less than two zones in width. Then it is necessary to divide the opening into infinitesimal elements, for the approximation of summing the contributions from a number of Fresnel zones is valid only if several zones are involved. The resultant amplitude at any point on the screen is obtained by integration over the contributions of all the infinitesimal elements in the opening. If the distance to the screen as compared with the width of the slit is great enough to make this method necessary, the lines from these elements to any point on the screen may all be considered parallel and hence the problem is one of Fraunhofer diffraction. The introduction of a lens to bring the diffracted light to a focus in a shorter distance introduces no fundamental modification.

Figure 47.5 represents in section a slit of width a, illuminated by parallel light from the left. Let the strip element of the wave front in the plane

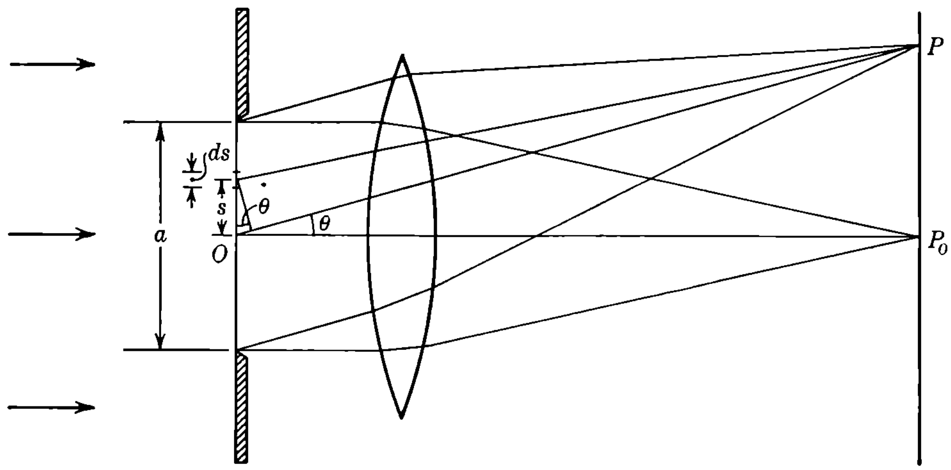


FIG. 47.5. Fraunhofer diffraction by a single slit.

of the slit be ds at a distance s from the origin O at the center of the slit. The amplitudes of all the wave trains reaching P will be proportional to the widths ds of the elements from which they originate. All other factors affecting the amplitudes will be essentially the same for all elements, and so we shall represent all these factors by the single symbol A. Consider first the secondary wave train emitted by an element ds situated at the origin. We may represent the instantaneous displacement at P produced by such waves at some time t by

$$A\ ds \sin 2\pi ft$$

or simply $A\ ds \sin \varphi$, since for this integration over all slit elements, t may be considered a constant.

For an element at a distance s above O the wave travels a lesser distance $s \sin \theta$, so that its phase at P will differ from that of the wave from the central element by $2\pi(s \sin \theta)/\lambda$. The displacement produced at P by this wave is therefore

$$A\,ds\,\sin\left(\varphi - 2\pi\,\frac{s\sin\theta}{\lambda}\right)$$

To obtain the resultant displacement at P from all elements, we integrate this expression over the width of the slit. If we set $\beta = (\pi/\lambda)\,a\sin\theta$, this integral is

$$\int_{-a/2}^{+a/2} A\sin\left(\varphi - 2\frac{\beta}{a}s\right)ds = \left[\frac{Aa}{2\beta}\cos\left(\varphi - 2\frac{\beta}{a}s\right)\right]_{-a/2}^{+a/2}$$
$$= \left(Aa\,\frac{\sin\beta}{\beta}\right)\sin\varphi \qquad (47.5)$$

A strong central maximum comes at P_0, where all the secondary wave trains arrive in phase. For this point θ and β are zero, and $\sin\beta/\beta = 1$ for $\beta = 0$, because $\sin\beta = \beta$ in the limit as $\beta \to 0$. The coefficient of $\sin\varphi$ is a measure of the relative amplitude at every point on the screen. The relative intensities I (ratio of the intensity at any angle θ to that at P_0) are then given by

$$I = \frac{\sin^2\beta}{\beta^2} \qquad (47.6)$$

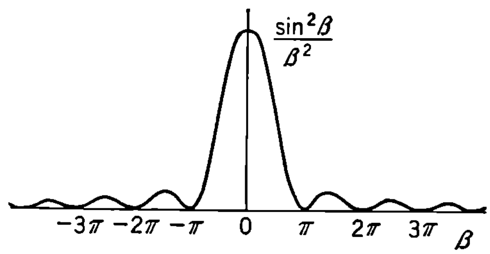


FIG. 47.6. Intensity distribution for Fraunhofer diffraction of a single slit.

Note that the quantity β is one-half the phase difference in radians between the contributions from the edges of the slit. Figure 47.6 is a plot of $\sin^2\beta/\beta^2$ vs. β. From the *principal maximum* the intensity falls to zero at $\beta = \pm\pi$. There are several *secondary maxima*, with equally spaced points of zero intensity at $\beta = \pm 2\pi, \pm 3\pi$, etc. The secondary maxima do not fall exactly halfway between these minima but are displaced toward the center by an amount that decreases with increasing β.

The principal maximum has double the width of the secondary maxima. The relative intensities may be calculated closely by finding the values of $\sin^2\beta/\beta^2$ at the halfway positions where $\beta = 3\pi/2, 5\pi/2, 7\pi/2, \ldots$. This gives 1/22.2, 1/61.7, 1/121, . . . of the intensity of the principal maximum. For the first minimum, $\lambda = a\sin\theta_1$; and since in practice θ is small, the angle θ_1 is given by

$$\theta_1 = \frac{\lambda}{a} \qquad (47.7)$$

If a lens is used close to the slit, the slit-screen distance is the focal length

f of the lens and then the linear distance d between successive minima on the screen is

$$d = \frac{f\lambda}{a} \tag{47.8}$$

Since the width of the pattern is proportional to λ, for red light it is roughly twice that for violet light, for the same f and a. If white light is used, only the central maximum is white and even that is reddish on its outer edges. Figure 47.7 is a photograph of a single-slit diffraction pattern.

Worked Example. A lens of 1 m focal length images on a screen the diffraction pattern of a slit 0.1 mm in width, when plane waves of sodium

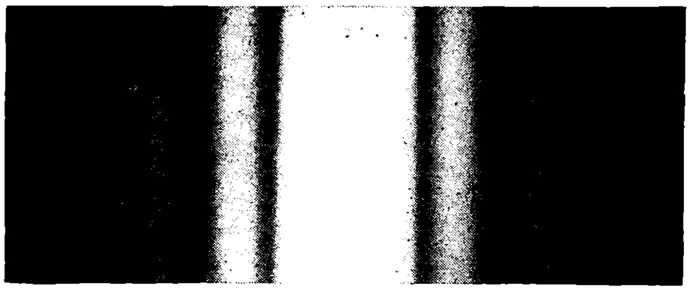


FIG. 47.7. Single-slit diffraction pattern.

light, $\lambda = 5{,}893$A, fall on the slit. Find (*a*) the width of the central maximum; (*b*) the distance between the fourth dark bands on either side.

a. From Eqs. (47.7) and (47.8), $2\theta_1 = 2\lambda/a = 2d/f$. Therefore the distance $2d$ between the two innermost minima is

$$2f\lambda/a = 2 \times 10^3 \times 5.893 \times 10^{-4}/10^{-1}, \text{ or } 11.8 \text{ mm},$$

which is the total width of the central maximum.

b. From Eq. (47.8),

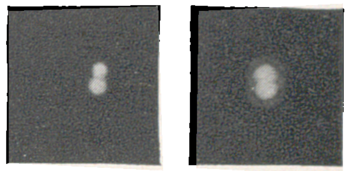
$$8d = 8f\lambda/a = 8 \times 10^3 \times 5.893 \times 10^{-4}/10^{-1}, \text{ or } 47.1 \text{ mm}$$

is the distance between the fourth dark bands.

47.6. Circular Aperture. The diffraction pattern formed by plane waves from a point source passing through a circular aperture is of considerable importance, but the calculations are difficult, requiring a double integration over the surface of the aperture. The pattern consists of a bright central disk surrounded by one or more fainter rings with both disk and rings shading off gradually at the edges. Calculation shows that the angular radius (angle subtended at the center of the aperture by the half-diameter of the image disk) of the central disk is

$$\theta_1 = \frac{1.22\lambda}{D} \tag{47.9}$$

where D is the diameter of the opening. This is to be compared with the formula $\sin\theta_1 = \lambda/a$ for the half-angular width of the central maximum of the single-slit pattern. On calling the intensity of the central disk 1, that of the first bright diffraction ring is only 0.017.



(a) (b)

FIG. 47.8. Images of two sources just resolved. (a) Short exposure, full aperture of camera lens. (b) Long exposure, pin-hole over camera lens.

The aperture in front of the lens need not be small to have Eq. (47.9) apply. We are here considering Fraunhofer diffraction, for which any opening of finite size includes only a fraction of a zone. The diameter D may be that of the lens of an optical instrument, for example, 40 in. for the Yerkes telescope. Then the half angle of the central disk is extremely small, but it is not zero.

47.7. Resolving Power of a Telescope. By the resolving power of any optical system is meant its ability to produce separate images of objects very close together. We have seen that, when parallel light passes through any aperture, it cannot be focused to a point image but rather gives a diffraction pattern. It is evident, therefore, that the images of two objects will not be resolved if their separation is much less than the width of the central diffraction maximum. (See Fig. 47.8.)

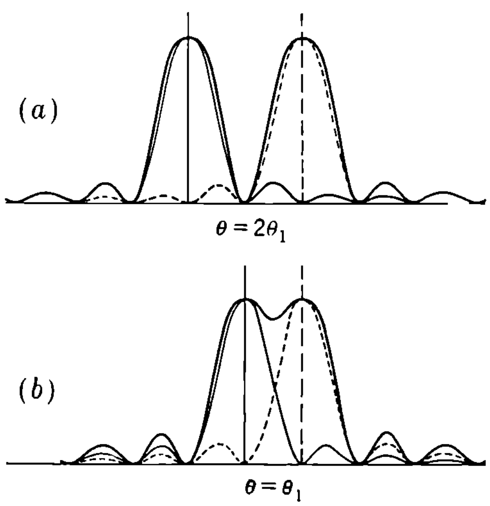


FIG. 47.9. Diffraction images of two close sources, (a) well resolved, (b) just resolved.

The intensity distribution diametrically across the rings and central disk of a circular-aperture diffraction pattern is similar to the intensity plot for the single-slit pattern shown in Fig. 47.6. In Fig. 47.9 are sketched two such diffraction patterns with (a) an angular separation $\theta = 2\theta_1$ and (b) $\theta = \theta_1$, where $\theta_1 = 1.22\lambda/D$ is the angular radius of the central disk. These double diffraction patterns are the images of two close point sources separated by this same angle θ. When $\theta = 2\theta_1$, each principal maximum falls exactly on the second minimum of the other pattern. This is the smallest possible angular separation θ of the sources that produces zero intensity between the two strong maxima in the resultant pattern. When the maximum of one pattern falls exactly on the first minimum of the other, as in Fig. 47.9*b*, the intensities of the two maxima in the resultant pattern are equal to those of the separate maxima, while the intensity at the center of the resultant minimum may be shown to be 0.81 of the intensity of the maxima. Since it would be impossible to resolve the two images if they were much closer than this, Lord Rayleigh suggested that this separation when the central maximum of one pattern falls on the first minimum of the other be arbitrarily set up as the criterion for resolution of two diffraction patterns. The angle θ_1 is called the *minimum angle of resolution.*

Worked Examples. *a.* Calculate this angle θ_1 for the giant Yerkes telescope for which $D = 40$ in. White light has an effective wavelength of 5.6×10^{-5} cm, so that $\theta_1 = 1.22 \times (5.6 \times 10^{-5})/(40 \times 2.54) = 6.7 \times 10^{-7}$ radian $= 0.14$ second of arc. This is then the smallest angular separation of a double star that could be theoretically resolved by this telescope. The linear radius of the first dark ring in an image is this angle multiplied by the focal length of 65 ft, or $6.7 \times 10^{-7} \times 65 \times 30.48 = 1.3 \times 10^{-3}$ cm. The central disk of a star image for this Yerkes telescope is then 0.26 mm in diameter.

b. Find the resolving power of the eye. Taking the pupillary diameter as 3 mm, the minimum angle of resolution for the eye is 47 sec, but actually the eye of the average person is not able to resolve objects of angular separation less than about 1 minute of arc.

Returning to the telescope, we note that increasing the size of the image by increasing the power of the eyepiece does not increase the amount of detail that can be seen. Since each point in the object becomes a small circular diffraction pattern in the image, when an eyepiece of very high power is used, the image appears blurred and no greater detail is seen. The normal magnifying power of a telescope is the ratio of its *resolving power*, or $1/\theta_1$, to that of the unaided eye. For the Yerkes telescope this ratio is $60/0.14 = 428$, which is also the ratio of the diameter of the objective lens to that of the pupil of the eye (Sec. 43.6).

***47.8. Resolving Power of a Microscope.** For a microscope we are interested in the smallest distance d between two points O_1 and O_2 in the object for which the images I_1 and I_2 (Fig. 47.10) are just resolved. This distance d is called the *limit of resolution* of the microscope. The objective subtends a large angle $2i$ at the plane of the object, with $D = 2p \tan i$. Now each of the images I_1 and I_2 consists of a circular diffraction pattern, and the angular separation of the two central disks when they are on the limit of resolution is $\theta_1 = 1.22\lambda/D$. The minimum value of $O_1O_2 = d$ is therefore $p \tan \theta_1 \cong p\theta_1$. Hence the smallest linear separation of two point objects that will be resolved is

$$d = p \frac{1.22\lambda}{2p \tan i} = \frac{1.22\lambda}{2 \tan i}$$

Actually the angle i is always so large (nearly 90°) that this simple derivation is not quite correct. It can be shown that $\tan i$ should be replaced by $\sin i$, so that

$$d = \frac{1.22\lambda}{2 \sin i} \tag{47.10}$$

is more nearly correct.

When the space between object and lens is filled with an oil, the resolving power is increased because of the smaller value of the wavelength in the oil. If the index of the oil is n, then λ (oil) $= \lambda$ (air)$/n$ and hence Eq. (47.10) becomes

$$d = \frac{1.22\lambda}{2n \sin i} \tag{47.11}$$

Thus, the smaller the wavelength, the smaller the separation of two object points that may be just resolved. The product $n \sin i$ is called the *numerical aperture*. Even with oil immersion the largest value of the numerical aperture obtainable is about 1.6. With white light of effective wavelength 5.6×10^{-5} cm, Eq. (47.11) gives $d = 2.1 \times 10^{-5}$ cm. This limit of resolution may be decreased further by using shorter wavelength (ultraviolet) light, but then the image must be photographed.

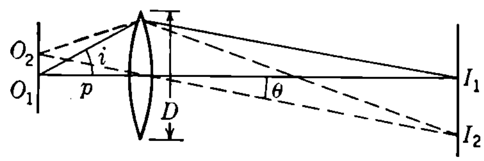


FIG. 47.10. Limit of resolution of a microscope.

47.9. The Electron Microscope. The limit of resolution may be made smaller by another factor of nearly 1,000 with an electron microscope. We have already seen in Chap. 36 how a divergent beam of electrons may be focused by either a magnetic or an electrostatic field. In the electron microscope, electrons pass through a thin section of the object and are absorbed more in its thicker portions, just as light is more strongly absorbed in the denser parts of a slide. The paths of these electrons are then changed by magnetic or electrostatic lenses that are the analogues of the objective and ocular of an optical microscope. The final image is formed on a fluorescent screen or on a photographic plate. By photographing the fluorescent-screen image a further gain in magnification may be attained. A recent model of the electron microscope is shown in Fig. 47.11, while Fig. 47.12 is an example of its high magnification.

To understand how the electron microscope produces the great gain in resolving power, it is necessary to discuss briefly the wave nature of the electron. It was first proposed by De Broglie in 1923, and has since been completely verified by several types of experiment, that any particle of mass m, moving with velocity v, is equivalent to a wave of wavelength λ

given by the equation

$$\lambda = \frac{h}{mv} \tag{47.12}$$

where h is Planck's constant (Chap. 49). The higher the electron's velocity, the smaller its "De Broglie wavelength."

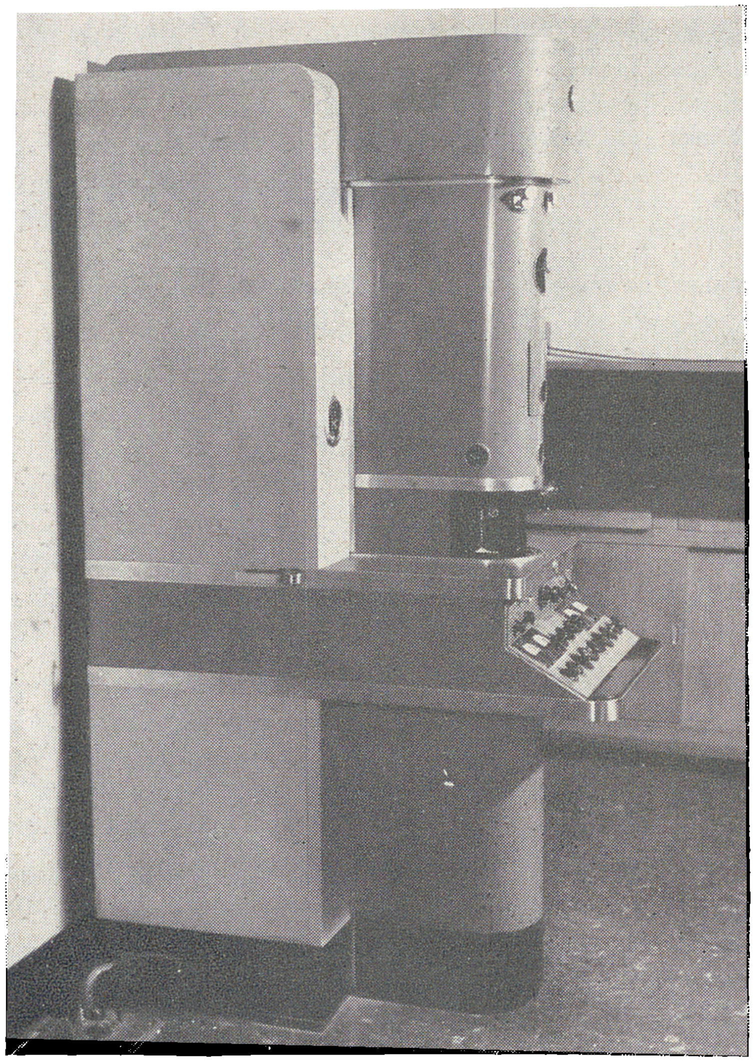


FIG. 47.11. R.C.A. research-type electron microscope. (*Courtesy of Dr. J. Hillier, R.C.A. Laboratories.*)

The velocity v is given to the electron by accelerating it through a potential difference V in an electron gun. Then $\frac{1}{2}mv^2 = eV$, or $v = \sqrt{2eV/m}$. Hence $mv = \sqrt{2meV}$, and thus

$$\lambda = \frac{h}{\sqrt{2meV}} \tag{47.13}$$

Now $h = 6.62 \times 10^{-27}$ erg sec, $m = 9 \times 10^{-28}$ gm, $e = 1.60 \times 10^{-19}$ coulomb. Therefore

$$\lambda = \frac{12.24}{\sqrt{V}} \times 10^{-8} \text{ cm} \tag{47.14}$$

If V is 100 volts, $\lambda = 1.22$A, while, if $V = 10{,}000$ volts, $\lambda = 0.122$A. It is thus seen that electrons of wavelength 10,000 times smaller than

FIG. 47.12. Electron micrograph of freshly prepared vanadium pentoxide fibers. These fibers are about 25A wide. The limit of resolution in this micrograph is about 20A. (*Courtesy of Dr. J. Hillier, R.C.A. Laboratories.*)

that of ordinary light are easily obtained. The limit of resolution is then about 10,000 times smaller than with an optical microscope. Since the whole system must be well evacuated, it is impossible to examine living objects but this is only a small limitation to its usefulness. Useful magnifications, electronic plus photographic, of 100,000 × have been obtained with this instrument.

47.10. Diffraction Gratings. A large number of closely spaced parallel slits of the same width constitute a *diffraction grating*. This valuable tool for the investigation of spectra is made by ruling with a diamond point many parallel lines on a glass surface. The spaces between the scratched lines on the glass serve as the parallel openings. In Fig. 47.13 we represent a magnification of a few successive rulings of such a grating. Plane waves are incident normally from the left. These light waves come from a lens that has at its principal focus an illuminated slit oriented parallel to the rulings of the grating. If the openings are narrow, diffracted light from each spreads out to such an extent as to interfere with the diffracted light from all the others.

Since the secondary wave trains start from every element of every opening in phase, these wave trains will reinforce each other in the forward direction and come to an image at P in the focal plane of the lens. This direction is not the only one to the right of this grating, however, along which reinforcement of the secondary disturbances from all the openings occurs. Consider the plane CD making an angle θ with the surface of the grating and starting from the element at the lower edge of the bottom opening. The angle θ is so chosen that CD is distant one wavelength from the lower edge of the first opening above this one, 2λ from the lower edge of the second opening, etc. The waves from all these elements are in phase along CD and therefore are brought to a focus at P_1. Also, the disturbance from other corresponding elements in each opening will be in phase along CD, for the distance of CD from any element of one of the openings is exactly λ greater than its distance from the corresponding element of the opening next below it. Therefore there will be general reinforcement of the diffracted light from all the openings in this direction and a maximum of intensity at P_1.

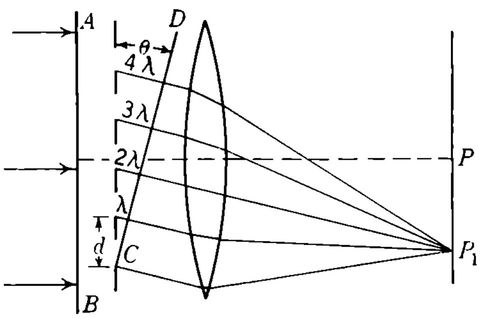


FIG. 47.13. Plane diffraction grating.

If the plane CD is distant 2λ from the lower edge of the second opening 4λ from the next opening, etc., the disturbances along CD are again in phase and another maximum occurs. Further maxima result when CD is distant 3λ, 4λ, . . . from the edge of the second opening. Similar maxima are also observed at the same angles above the normal to the grating. In Fig. 47.13, if d is the distance between successive openings, the values of the angle θ for the production of maxima are given by

$$d \sin \theta = n\lambda \tag{47.15}$$

where $n = 0, 1, 2, 3,$ etc. When $n = 1$, we have the first-order maximum, $n = 2$ gives the second-order, etc.

Provided that the number of openings in the grating is large, these maxima are extremely sharp. For constructive interference along surfaces such as CD, Eq. (47.15) must be *exactly* fulfilled. The slightest departure of the surface CD from the angles given by this equation results in complete destructive interference, for the disturbance from an element of one opening is then necessarily just out of phase with the disturbance from a corresponding element of another opening. The maxima are called *spectrum lines* because, as mentioned above, the parallel beam incident on the grating is usually produced by a lens at the focus of which is an illuminated slit parallel to the rulings of the grating. If the light is monochromatic, each of the maxima is a sharp image of the slit in that color. If the slit is illuminated by light consisting of a mixture of wavelengths, *there will be an image of the slit for each wave length in each "order."* That is, a grating produces, not one spectrum, but two first-order spectra, two second-order spectra, etc. If white light is used, these spectra are, of course, continuous.

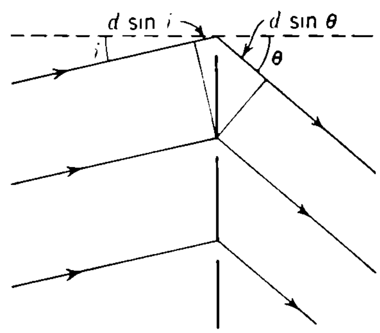


FIG. 47.14. Angle of incidence $= i$, angle of diffraction $= \theta$.

If the parallel light is incident on the grating at an angle $i \neq 0$ (Fig. 47.14), Eq. (47.15) takes the more general form

$$d(\sin i + \sin \theta) = n\lambda \qquad (47.16)$$

As a measure of the separation between the spectrum lines of wavelengths λ_1 and λ_2, we compute the *angular dispersion* $d\theta/d\lambda$, or the rate of change of angle with wavelength. On differentiating Eq. (47.16) with respect to λ and remembering that i is a constant independent of λ, there results

$$\frac{d\theta}{d\lambda} = \frac{n}{d \cos \theta} \qquad (47.17)$$

Usually θ is a small angle and therefore $\cos \theta$ is nearly equal to 1.

This equation shows that, for a given small wavelength difference $d\lambda$, the angular separation $d\theta$ is directly proportional to the order n. Hence the second-order spectrum is twice as wide as the first, the third three times as wide as the first, etc. Second, the smaller the grating space d, the greater $d\theta$ and the more widely spread the spectrum. Third, if θ is small, $\cos \theta$ differs little from unity and the different spectrum lines in one order will differ in angle by amounts that are directly proportional to their wavelength difference. Such a spectrum is called a *normal spectrum*, or *rational spectrum*. This simple linear scale for wavelengths is to be contrasted with the *irrational spectrum* produced by a prism in which the violet is much more spread out than the red.

Worked Examples. *a.* Parallel light consisting of two wavelengths 4,000A (violet) and 7,500A (red) falls on a grating for which $d = 0.001$ cm. Calculate the difference in the angular deviations of these two spectral lines in the second order.

From Eq. (47.15), for the 4,000A line,

$$\sin \theta = (2 \times 4 \times 10^{-5})/(1 \times 10^{-3}) = 0.08,$$

or $\theta = 4°35'$.

For the 7,500A line, $\sin \theta = (2 \times 7.5 \times 10^{-5})/(1 \times 10^{-3}) = 0.15$, or $\theta = 8°38'$.

The angular separation of these two spectral lines is thus 4°3′.

b. Show that the violet of the third order overlaps this second-order red.

For the third order at 4,000A,

$$\sin \theta = (3 \times 4 \times 10^{-5})/(1 \times 10^{-3}) = 0.12,$$

or $\theta = 6°54'$.

This is a smaller angle of diffraction than that for the second-order 7,500A line; hence overlapping occurs. This overlapping of orders may be an aid in wavelength measurements and if found undesirable may usually be eliminated by use of suitable color filters.

The grating discussed so far is called a *transmission* grating. Diffraction gratings for spectroscopic research are now almost always ruled on a polished concave spherical surface, a development initiated by H. A. Rowland (1848–1901) at Johns Hopkins University. The rulings are intersections with the surface of equidistant planes parallel to the principal axis of the surface. This ruled surface is both a grating and a concave mirror that forms images of its diffracted light without the use of lenses. If is thus possible to photograph spectra in wavelength regions —for example, far into the ultraviolet—where glass and other lens materials are not transparent. The most powerful concave gratings have 6 in. of ruling, 30,000 lines to the inch, and have a radius of curvature of 35 ft.

***47.11. Diffraction cf X Rays and Electrons.** It was suggested by von Laue in 1912 that the indicated wavelengths of X rays, about 10^{-8} cm, are about the same as the distance between atoms in a crystal and that therefore a crystal might act as a three-dimensional diffraction grating for X rays. The experiment was successfully performed by Friedrich and Knipping. In 1926 Davisson and Germer discovered the almost identical phenomenon of electron diffraction by crystals. These experiments prove both that X rays and electrons are wavelike in character and that the atoms in crystals are arranged in an orderly, regular manner. X-ray and electron diffraction are now much used in the study of the structure of crystalline substances.

Figure 47.15 is a diagram of the arrangement of atoms in rock salt in which the sodium and chlorine atoms form a simple cubic lattice. It is evident that various parallel planes exist in this crystal, along which the atoms are distributed in a regular manner. These sets of equidistant planes act somewhat like the succession of equispaced slits in a plane grating.

If a beam of X rays is incident on a crystal, some reflection occurs at each of these planes. The X rays cause the electrons in the atoms to vibrate and send out secondary scattered waves. These will all be in phase and give relatively intense reflected beams only in the directions for which mirrorlike reflection takes place from the many equidistant planes. Let the horizontal lines in Fig. 47.16 represent two parallel layers of atoms a distance d apart, upon which a narrow beam of X rays is incident at a "glancing angle" θ. If the secondary waves from A and E are to be in phase at B and D, forming part of a reflected wave front traveling in the direction of rays 1 and 2, the path difference

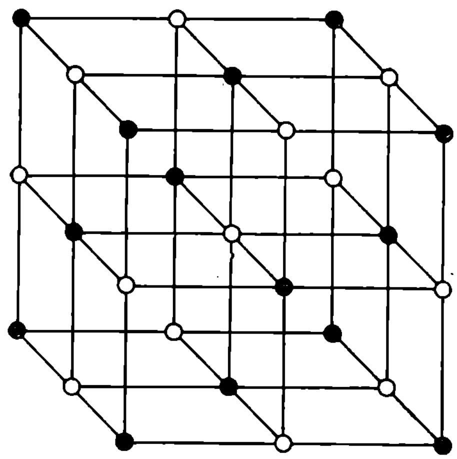


Fig. 47.15. Crystal lattice of rock salt.

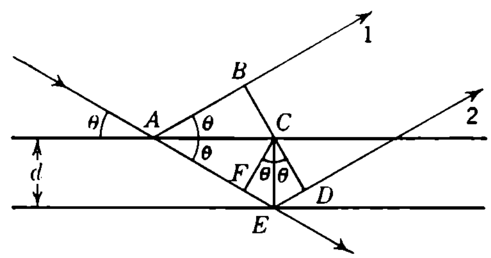


Fig. 47.16. Reflection of X rays from crystal when $2d \sin \theta = N\lambda$.

$$(AE + ED) - AB = FE + ED = 2d \sin \theta$$

must equal $N\lambda$, where N is an integer. Therefore, for strong reflection,

$$2d \sin \theta = N\lambda \tag{47.18}$$

This relation is known as Bragg's law and is identical with the condition for constructive interference in reflection of visible light from the sides of a thin film. If the angles θ for strong reflection are measured for X rays of known wavelength λ, the spacings d between the several kinds of atom planes in the crystal may be calculated. If the incident X rays pass first through a narrow slot in a lead block, X-ray "line" spectra are produced by these reflections from the planes of a crystal, and if d is known for the crystal the wavelengths of the X-ray lines can be computed. Each element of the periodic table when used as the target of an X-ray tube displays a characteristic X-ray spectrum (Chap. 49) that may be photographed with this crystal spectrograph.

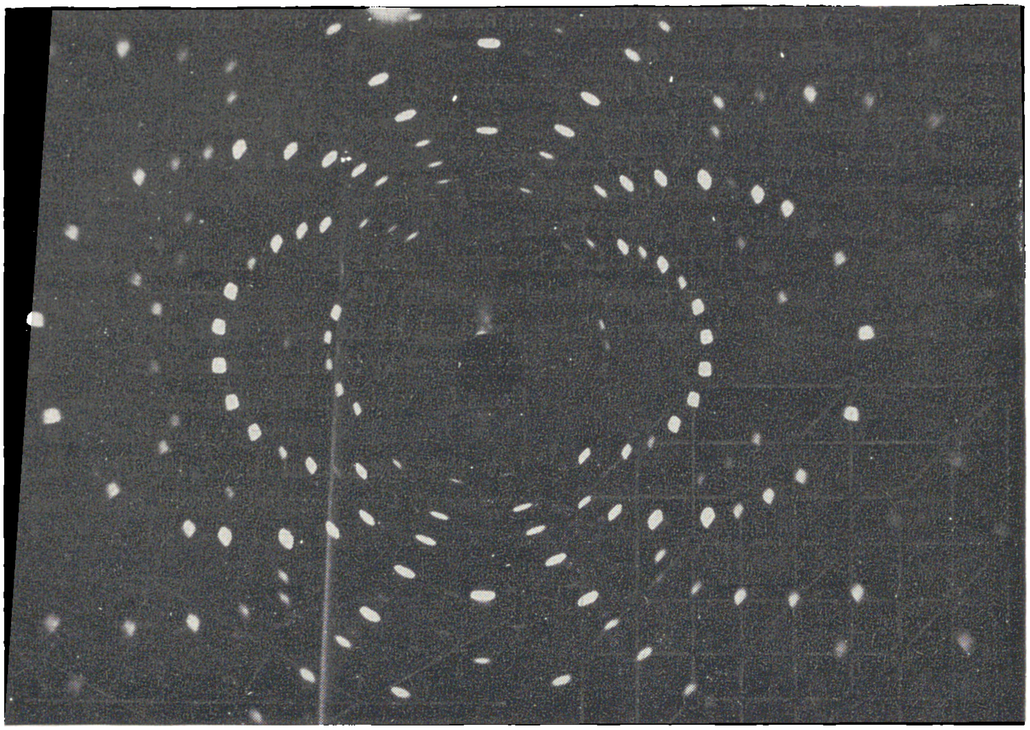


FIG. 47 17. Von Laue-spot X-ray diffraction pattern of anhydrite (calcium sulphate.) (*Courtesy of H. Winchell.*)

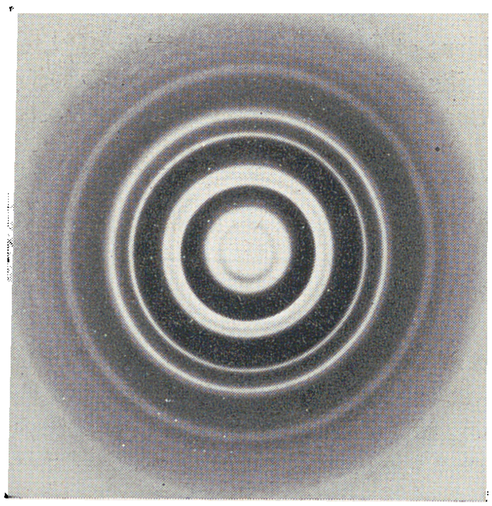


FIG. 47.18. Electron diffraction pattern from a gold film. (*Photographed by L. H. Germer.*)

When a narrow beam of X rays is sent through a thin section of a single crystal and thence onto a photographic plate, a Laue spot diffraction pattern such as that shown in Fig. 47.17 is formed. Each spot represents reflection from a particular set of planes in the crystal lattice. The exact electron-diffraction analogue of such a pattern was first produced by the Japanese Kikuchi, using 68,000-volt electrons passing through a thin mica crystal. If the narrow beam of electrons or X rays is sent through a thin foil (polycrystalline, *i.e.*, with all its tiny crystals oriented at random), a diffraction-ring pattern (Figs. 47.18 and 47.19) is

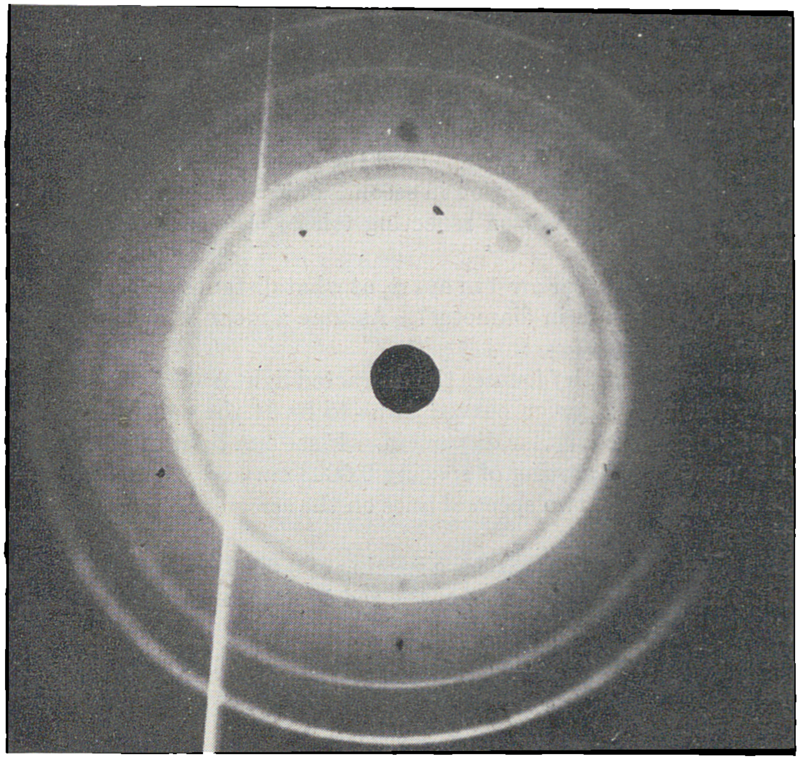


FIG. 47.19. Diffraction pattern produced by transmission of an X-ray beam through sodium chloride powder. (*Courtesy of H. Winchell.*)

produced. Each crystal forms its own pattern of spots, but these are turned at all angles with equal probability, so that rings result. The same ring pattern would be obtained by spinning a spot pattern of a single crystal like Fig. 47.17 about its central spot.

PROBLEMS

1. Light of wavelength 5,500A from a distant point source passes through a circular aperture and falls on a screen 1 m beyond the aperture. If the aperture exposes the first five Fresnel zones, what is its diameter?

2. Through what distance must the screen of Prob. 1 be moved toward the opening so that the center of the diffraction pattern changes from bright to dark?

3. A single-slit diffraction pattern is formed with mercury light of wavelength 5,461A. A lens of focal length 200 cm focuses the pattern on a screen. If the slit width is 0.5 mm, what is the width of (*a*) the central maximum; (*b*) one of the secondary maxima?

4. Calculate the intensity of the fourth secondary maximum relative to that of the principal maximum in the diffraction pattern of a single slit. (Assume this secondary maximum to be midway between the adjacent minima.)

5. Calculate the angular and linear radii of the first dark ring of the image of a star in an ordinary field glass if the diameter of the objective is 35 mm and its focal length is 25 cm.

6. What must be the aperture of the objective of a telescope if two sources with an angular separation of 1 sec may be just resolved?

7. Compute the relative limits of resolution, assuming the numerical aperture to be the same, for an optical microscope using white light and an electron microscope using 1,000-volt electrons.

8. An oil-immersion microscope will resolve a set of test lines 100,000 to the inch, using blue light of wavelength 4,800A. Find the numerical aperture.

9. Compute the angular separation in seconds of arc of the closest double star that can be resolved by the Mt. Wilson reflecting telescope, whose objective is a mirror 100 in. in diameter.

10. If the headlights of a car are 1 m apart, at what distance will they be resolved by the eye, if the pupil is 3 mm in diameter? Assume a mean wavelength of 6,000A for the light in air.

11. If a lens of focal length f focuses the diffracted light from a transmission grating, the linear dispersion on a screen curved so as to be at the constant distance f from the lens is just f times the angular dispersion. If sodium light of wavelengths 5,890A and 5,896A is used with a grating of spacing 0.0003 cm and a lens with $f = 2$ m, find the distance between these two spectral lines on the screen (*a*) in the first order; (*b*) in the third order.

12. A plane transmission grating has a spacing of 0.0002 cm. Calculate the angular separation in the second order between the red hydrogen line 6,563A and the violet line at 4,340A.

13. For rock salt the spacing d of one set of planes in the lattice is 2.8×10^{-8} cm. If palladium X rays of wavelength 0.586A are used, find the glancing angle for Bragg reflection.

14. In the spectrum of light of a certain color formed by a grating with 250 lines per millimeter the angular deviation of the second-order spectrum from the central image is 15° when the incident light is normal to the grating. Calculate the wavelength of the light.

15. How many orders of a wavelength 6,000A can be produced by a diffraction grating having 5,000 lines per centimeter? [The maximum value of sin θ in Eq. (47.15) is 1.]

16. The grating of Prob. 11 is used with a light source of unknown spectral composition. It shows lines at distances 67.60, 67.70, and 68.32 mm from the zero position. Assuming them to be first order lines, calculate their wavelengths.

CHAPTER 48
RADIANT ENERGY

48.1. Polarization of Light. The experiments in interference and diffraction prove that light is a wave motion. These phenomena occur for all types of wave. There is another group of experiments, demonstrating the *polarization of light,* which show that light is a transverse wave motion. For transverse waves the displacement is transverse, *i.e.*, at right angles to the direction of propagation, in contradistinction to longitudinal waves (cf. Sec. 37.6). If the displacement in a transverse wave is confined to one plane, the wave is called *plane-polarized.* For example, a wave set up in a rope by random displacement of one end in all directions is transverse but unpolarized. However, if the rope is horizontal and the end moves up and down, the transverse wave is polarized.

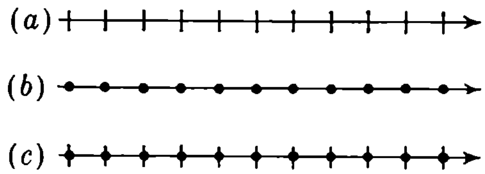


FIG. 48.1. Representation of (a) and (b) plane-polarized light, (c) ordinary light.

Light is an electromagnetic disturbance, in which the electric and magnetic field strengths are in the transverse plane. Experiment shows that it is the electric vector **E** which produces the optical polarization effects we are about to consider. When we speak of the "vibrations" in the light, we refer, then, to the variations in the electric vector in planes transverse to the direction of propagation. Ordinary light consists of wave trains about a meter in length, each coming from a separate atom, and with the waves polarized in one plane. A beam of ordinary light consists of millions of such wave trains, however, from the very large number of radiating atoms and molecules in the source, and hence the vibrations are in all transverse directions with equal probability. But all the amplitudes of the electric vectors may be resolved into components along any two transverse directions at right angles to each other, and therefore ordinary light may be thought of as consisting of two kinds of vibrations only, with the two sets of waves vibrating in planes at right angles to each other. A convenient pictorial representation of these vibrations is shown in Fig. 48.1. In (*a*) and (*b*), plane-polarized light is traveling to the right, in (*a*) vibrating with the electric vector in the plane of the page, while in (*b*) the vibrations are perpendicular to the page. The simultaneous presence of both these vibrations as in (*c*) is typical of ordinary light.

The common methods of effecting the polarization of light are by means of (1) reflection, (2) refraction, (3) selective absorption, (4) double refraction, and (5) scattering. We shall consider the first four of these in turn.

48.2. Polarization by Reflection. The French physicist Malus discovered in 1808 that, if a beam of light is incident on the polished surface of a plate of ordinary glass at an angle of about 57°, the reflected light is plane-polarized. This may be demonstrated by the following very simple experiment: In Fig. 48.2, AB represents a beam of ordinary light incident on the glass surface B at this particular angle. The reflected beam from B is then allowed to fall on another glass plate C parallel to plate B. Light from this second reflection is observed in the direction CE. Now if the upper plate is rotated about BC as an axis, with the angle of incidence kept constant, the intensity of the reflected light from C is found to decrease, reaching zero when C has been rotated through 90°. If the upper surface is rotated further about BC, the reflected light reappears and reaches a maximum of intensity at 180°.

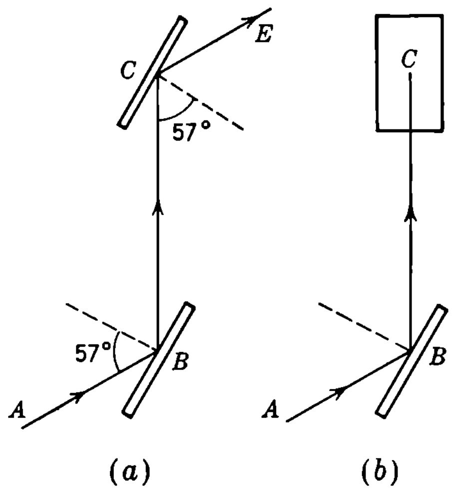


FIG. 48.2. Polarization by reflection from glass surfaces.

If the angle of incidence on either surface is not 57°, the twice-reflected beam will go through a minimum for 90° rotation of C but it will not go to zero intensity. The critical value of the angle of incidence that produces the zero minimum is called the *polarizing angle*. Sir David Brewster discovered experimentally in 1812 that at this polarizing angle the reflected and refracted rays are perpendicular (Fig. 48.3). Calling the polarizing angle i_p and calling the corresponding angle of refraction r_p, $\sin i_p/\sin r_p = n$ (Snell's law); also, if the angle between the reflected and refracted rays is 90°, $\sin r_p = \cos i_p$. Therefore

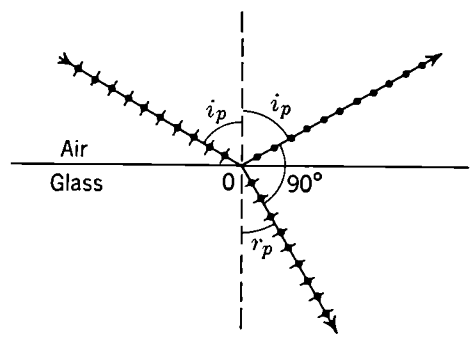


FIG. 48.3. For polarizing angle the reflected and refracted rays are at right angles to each other.

$$\frac{\sin i_p}{\sin r_p} = \frac{\sin i_p}{\cos i_p} = \tan i_p = n \tag{48.1}$$

This relation is known as *Brewster's law.* It shows that the angle of maximum polarization for the reflected light depends on the refractive index and hence varies with the wavelength and the kind of dielectric material used as the reflector. If the reflector is a metal surface, the polarizing angle is 90°, which corresponds to an infinite value of n.

It is clear that, if light consisted of longitudinal vibrations, rotation about a vertical axis of the second mirror even if it is inclined at the 57° angle would have no effect on the intensity of the reflected rays E. Investigation shows that at the polarizing angle the reflected beam consists entirely of vibrations perpendicular to the *plane of incidence* (the plane defined by the incident ray and the normal to the surface at O in Fig. 48.3). Some of these perpendicular vibrations are also refracted, as

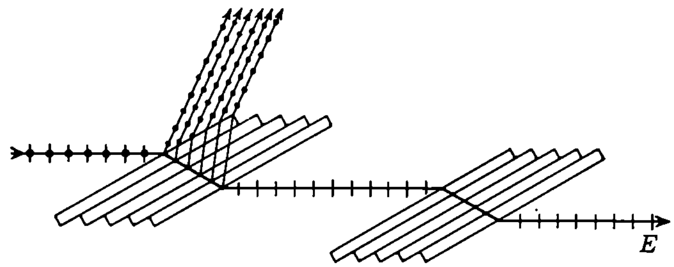


FIG. 48.4. Pile of glass plates for production and analysis of plane polarized light.

are, of course, all the vibrations in the plane of incidence. If a pile of plates, as shown in Fig. 48.4, is used at the polarizing angle, some of the $\perp$ vibrations are reflected at each surface and all of the $\parallel$ vibrations are refracted. The reflected beams are thus all plane-polarized in the same plane, and the refracted beam, having lost more and more of its $\perp$ vibrations, is partly plane-polarized. The larger the number of plates, the more nearly plane-polarized is this transmitted beam. If the second pile shown in Fig. 48.4 is rotated 90° about an axis parallel to the direction of the incident light, the light vibrations will now be mostly perpendicular to the plane of incidence and hence will be reflected at the first surface. In this crossed position, very little light is transmitted by the second pile. In intermediate positions the intensity I of the light transmitted by the second pile is

$$I = I_0 \cos^2 \theta \tag{48.2}$$

where I_0 is the intensity of the light incident on it and θ is the angle through which the pile has been rotated from the position of maximum transmission. Equation (48.2) is called the *law of Malus.* Its proof is simple. The amplitude A of the incident plane-polarized light may be resolved into a reflected component $A \sin \theta$ and the transmitted component $A \cos \theta$. Since the intensity is proportional to the square of the amplitude, Eq. (48.2) follows.

The first pile of plates that produces polarized light is called a *polarizer*, while the second pile is called the *analyzer*. An optical system composed of a polarizer and an analyzer is called a *polariscope*.

★48.3. Double Refraction. Many crystalline substances such as calcite (a hydrated calcium carbonate) and quartz exhibit the phenomenon of *double* refraction. A ray of light incident on a crystal of calcite emerges as two rays; and if a rhomb of calcite is placed over a black dot on a piece of paper, in general two dots will be seen. Natural calcite crystals have a rhombohedral form, each face being a parallelogram with two acute angles of 78°5′ and two obtuse angles of 101°55′. Two trihedral angles of the crystal are formed by the junction of the

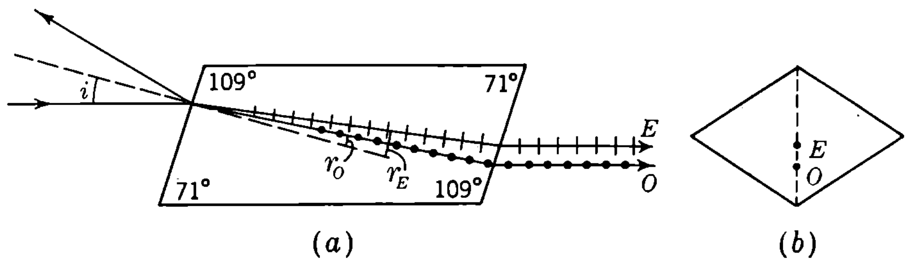


FIG. 48.5. (*a*) Principal section of calcite crystal showing double refraction. (*b*) End view.

obtuse angles of three faces. A line into the crystal at one of these solid angles and equally inclined to the three faces is called a "crystallographic axis." Only when viewed through the crystal in the direction of this axis will an object *not* appear double. *Any direction in the crystal parallel to the crystallographic axis is called an optic axis.*

If a thin slab having its plane faces perpendicular to the crystallographic axis is cut from such a crystal and mounted between metal plates, measurement of the capacitance of this capacitor yields the specific inductive capacity K of the crystal. If the slab is cut with its faces parallel to the crystallographic axis, however, it is found that a different value of K results. The atoms of the crystal are so ordered that for all directions in the planes perpendicular to the axis there is complete symmetry of the electrical vibrations associated with light traveling along the axis. This symmetry is lacking for the vibrations in any other sets of parallel planes in the crystal. Since K equals the square of the index of refraction [Eq. (39.5)], this asymmetry in general produces double refraction for light passing through the crystal in any direction other than along an optic axis.

A plane including an optic axis and the normal to a crystal surface is called a *principal section.* A principal section always cuts the surfaces of a calcite crystal in a parallelogram with angles of 71° and 109°, as shown in Fig. 48.5. When a beam of ordinary light is incident on one of these crystal faces, two refracted beams are produced. Snell's law holds for one of these but not for the other. The ray for which Snell's law holds is called the *ordinary ray O*, while the other,

E, is called the *extraordinary ray*. If the incident light is normal to the surface, the ordinary ray passes through without deviation but the extraordinary ray is refracted at some angle and comes out parallel to but displaced from the incident beam. Rotating the crystal about the O ray in this case causes the E ray to rotate around the stationary O ray.

When the two refracted rays are tested for polarization, it is found that they are both plane-polarized but at right angles to each other. This test can be made, for example, with a pile of glass plates at the polarizing angle. The vibrations in the O ray are perpendicular to the principal section and the plane of incidence, as shown in Fig. 48.5, while in the E ray the vibrations are in this plane.

Plane-polarized light by *selective absorption* is produced by a number of minerals and organic compounds. *Tourmaline*, a mineral crystal having the shape of flat, elongated hexagonal slabs, absorbs the light vibrations perpendicular to the long

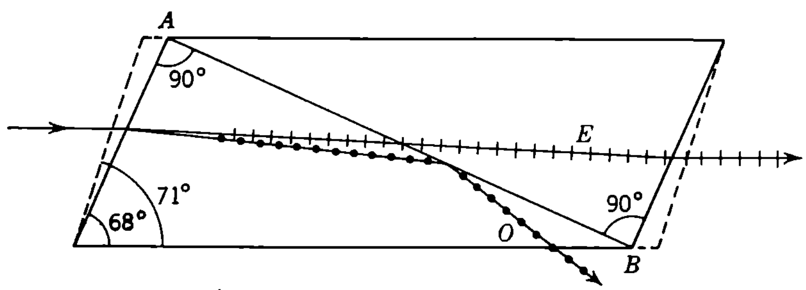


FIG. 48.6. The Nicol prism.

side of the crystal, so that the transmitted light is plane-polarized with the vibrations parallel to the long side. This can be verified with a second tourmaline crystal; for when the latter is rotated 90° with respect to the first, no light gets through. Because these crystals are colored, they are not used in optical instruments.

The modern, much-used *polaroid* film consists of thin sheets of nitrocellulose packed with ultramicroscopic crystals of the organic compound iodosulfate of quinine (herapathite), with their optic axes all parallel. These double-refracting crystals completely absorb one component of the polarization, transmitting the other with little loss and with very little coloration. Since these films may be made of large size and are relatively inexpensive, they find numerous applications as in devices for diminishing glare (reflected light is always partly polarized, with the vibrations parallel to the reflecting surfaces predominating) and in three-dimensional motion pictures.

The *Nicol prism* is an ingenious optical device made from a calcite crystal, until recently much used in optical instruments for producing and analyzing plane-polarized light, but now largely supplanted by polaroid film. The construction of the Nicol prism is illustrated in Fig. 48.6. First the ends of the crystal are cut down from 71° in the principal section to an angle of 68°. The crystal is next sliced along the plane AB perpendicular both to the principal section and to the end faces. The two cut surfaces are then polished and cemented together with Canada balsam, which is a transparent substance with an index about midway between the index of the O and E rays.[1] Therefore the O ray for large angles of incidence is totally reflected at the balsam surface, whereas the E ray is refracted into the balsam and on through the crystal.

48.4. Stress Analysis by Polarized Light. Polarization effects are also observed when specimens of transparent substances such as glass, celluloid, or bakelite under mechanical stress are placed between

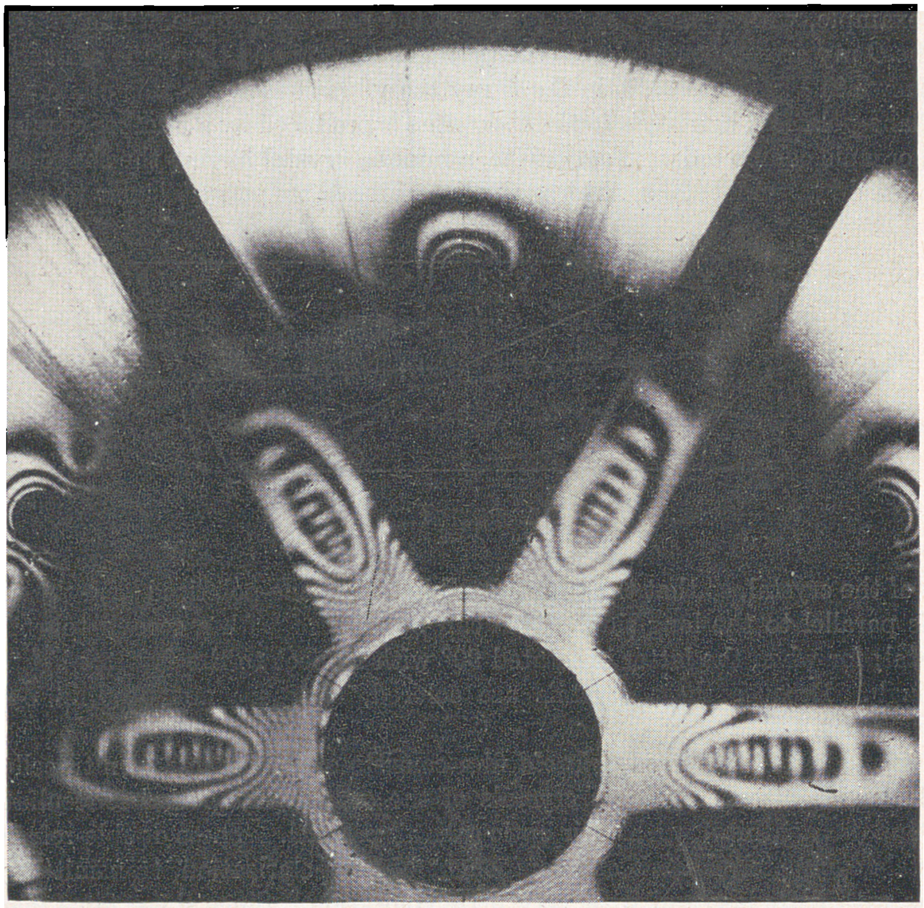


FIG. 48.7. Photograph of stress lines in a rotating model of an impeller of a supercharger of a modern airplane engine. (*Photographed by J. L. Meriam.*)

a polarizer and analyzer. Normally these substances are not doubly refracting, and if they are placed in stress-free condition between "crossed" nicols or polaroids they will leave the field of view dark. If, now, stress is applied to the specimens, the light emerging from them is in general (elliptically) polarized and is therefore partly transmitted by the analyzer. The stress distribution is made very evident by colored bands of light on the final image.

[1] Although Snell's law does not hold for the E ray, there is still a definite velocity and hence a definite index of refraction for this ray.

In this manner, glass-to-metal seals may be examined for stress in the glass, and optical glass may be checked for residual stress after annealing. If scale models of engineering materials are constructed of transparent bakelite, these may be analyzed for stress distribution when under load, giving valuable information not obtainable from mathematical analysis because of the complexity of the shapes. This field of study is known as *photoelasticity*. Figure 48.7 is a photograph of a model under stress.

Most liquids are not normally doubly refracting, but they become so when an electric field is set up in them. This phenomenon is known as the *Kerr effect*. Light is passed between two capacitor plates inserted in a glass cell containing the liquid (nitrobenzene gives the largest effect). Such a cell placed between crossed polarizer and analyzer constitutes an "optical shutter" that will operate even at radio frequencies. When the electric field is off, no light is transmitted by the analyzer. When the electric field is on, however, the liquid becomes doubly refracting and the light is restored.

48.5. Emission of Radiation from a Hot Body. When a gas is made luminous by the passage of an electric discharge through it, as in a neon sign, the light dispersed by a prism or grating consists in general of sharp spectral lines characteristic of the atoms and molecules present in the gas. The description and interpretation of such spectra form the subject matter of the following chapter, but in the remainder of the present chapter we shall continue the discussion of *thermal radiation* begun in Sec. 22.5. As mentioned there, the electromagnetic radiation emitted by a hot solid is identical with light in all respects except that it consists mostly of wavelengths much longer than those to which the eye is sensitive. At temperatures above about 700°C, however, the radiation emitted by bodies is in part visible light. When examined with a spectrograph, the light is found to have a continuous spectrum, the visible fraction constituting but a small portion of the total *radiant energy* distributed over all wavelengths. At extremely short as well as at the very longest wavelengths the intensity in this continuous spectrum drops effectively to zero.

To determine the distribution of radiant energy emitted by a hot body in the region of infrared wavelengths, one method is to allow the radiation to pass first through a rock-salt prism or some other device to disperse the radiation into a spectrum and then perhaps to cause portions of this spectrum to fall successively on a diffraction grating with a relatively coarse spacing of its slits or rulings. The dispersed radiation is then focused onto a slit in front of one set of junctions in a sensitive thermopile. The deflection of the galvanometer in the thermopile circuit is a measure of the energy in the particular wavelength interval $d\lambda$

isolated by the thermopile slit. When the radiation from hot solid and liquid bodies at various temperatures is systematically explored in this manner, wavelength distribution curves similar to those sketched in Fig. 48.9 for a black-body radiator are obtained.

48.6. Kirchhoff's Law. Black-body Radiation. In Sec. 22.5 we discussed the experiment of suspending a small body at temperature T_2 inside an evacuated enclosure whose walls are maintained at a temperature T_1. Regardless of its initial temperature T_2, whether T_2 is greater than or lower than T_1, and regardless of the nature of the surface of the body, the object always comes to the same temperature T_1 as that of the walls. In this condition of thermal equilibrium there must be equal amounts of radiation streaming both ways between the body and the walls. Of the radiant energy from the walls falling on the body, part is reflected and part is absorbed. But when thermal equilibrium is established and the temperature of the body remains constant at the value T_1, it must be emitting radiation at the same rate as it is absorbing it. That is, if the body is a good absorber, it is also a good emitter, and vice versa. At the constant temperature T_1 the amount of radiation from the walls incident on the body per second is always the same, and hence if the body is a poor absorber it must be a good reflector. It follows that, if the body is a good reflector, it is a poor emitter.

The radiant energy falling on a surface or emitted by a surface per second is called *radiant flux* and is measured in watts. Radiant flux incident on a surface *per unit area* is the intensity, or *irradiance*, denoted by R, for which a convenient unit is the watt per square centimeter. The fraction of the incident flux reflected by the surface is called its *reflectance* r, while the fraction absorbed is called its *absorptance* a. If the surface is opaque, $r + a = 1$. The total radiant energy *emitted* by a surface per unit area per second is called the *radiant emittance* W of the surface, and again 1 watt/cm² is a convenient unit.

Let the surface area of the small body in the enclosure be S, and when equilibrium is established let its rate of absorption of energy per unit area be aR, while its rate of emission of energy per unit area is W. Then, if the temperature of the body is constant,

$$WS = aRS$$

or

$$R = \frac{W}{a} \tag{48.3}$$

Now the irradiance R depends only on the temperature of the enclosing walls, not on how good an absorber the small body is. Therefore *the ratio of the radiant emittance to the absorptance is the same for all bodies.* This is a statement of *Kirchhoff's law.*

A body for which $a = 1$ is called an *ideal black body.* Therefore Kirchhoff's law should also include the additional statement that the ratio of the radiant emittance to the absorptance for any body is equal to the radiant emittance of a black body at the same temperature. That is,

$$\frac{W}{a} = W \text{ (black body)} \tag{48.4}$$

and consequently the absorptance a of a body is equal to the ratio of its radiant emittance to that of a black body at the same temperature. This ratio is also called the *emissivity* e of a surface; $e = a$.

No real surface absorbs perfectly all the radiation falling on it, though this may be very closely approximated by a small hole in an enclosure lined with poorly reflecting walls (Fig. 48.8). Practically all the radiation entering this hole will be absorbed after a few reflections from the walls and hence will never get out of the enclosure. The interior walls are also emitting radiant energy, of course, and some of this escapes through the opening. Since the walls have some reflectance, their radiant emittance is less than that of a black body but the energy they reflect just compensates for their smaller emittance. Therefore the hole will be very nearly a perfect radiator at the temperature of the interior walls of the box. For very similar reasons, an open window appears black when seen from the outside even in daytime.

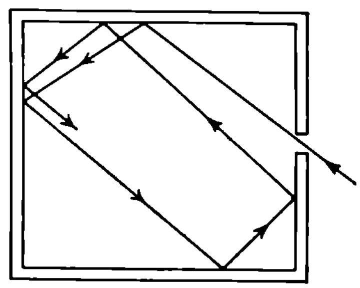


FIG. 48.8. A small opening in a box lined with absorbing walls is very nearly a perfect absorber of radiation.

***48.7. Planck's Radiation Law.** Toward the end of the last century one of the most challenging problems facing physicists was to give a theoretical deduction of the distribution of the black-body radiation among wavelengths (Fig. 48.9). One line of reasoning that was tried was to treat the radiation in the black-body enclosure as a gas and to discuss its properties by means of thermodynamics. Another method was to apply the laws of the kinetic theory to this radiation "gas." These attempts succeeded in deriving correct formulas for the *total* radiation from a black body at any temperature and for the distribution of the radiant emittance with λ for the longer wavelengths. But the exact form of the curves giving the distribution of black-body radiation over the entire range of wavelengths was not explained by any of this work, and it seemed clear that the classical concepts of theoretical physics are not adequate to explain the experimental curves.

Max Planck, in 1900, succeeded in producing a formula giving correctly the distribution of this radiation with wavelength, but he had to abandon one of the fundamental ideas of the classical electromagnetic theory. In that theory one assumes that the electrically charged particles, whose motions in the radiator cause the *continuous* emission of radiation, can have a *continuous* range of energies. Planck decided to give up this idea. He assumed rather that for any one of the

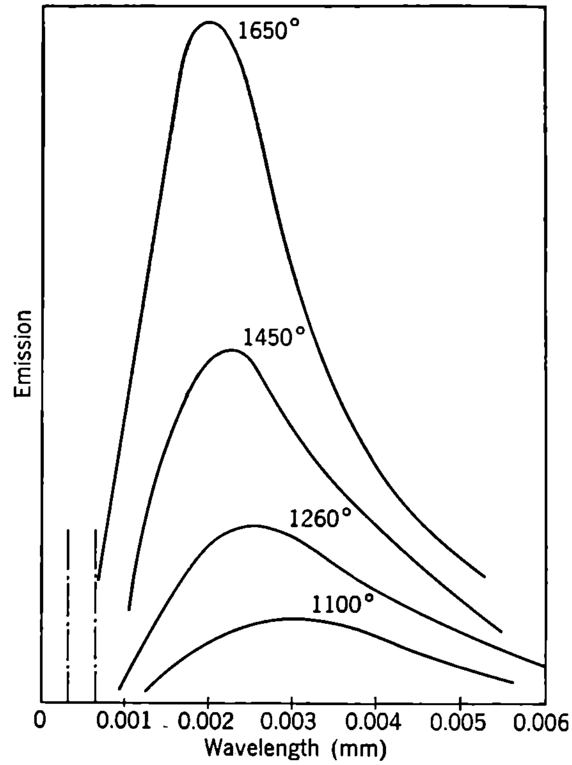


FIG. 48.9. Spectral emittance curves for a black body. Vertical dashed lines indicate limits of the visible spectrum.

very large number of elementary oscillators the energy E is proportional to some integral multiple of its frequency f. That is,

$$E = nhf \tag{48.5}$$

where n is any integer and h is now called Planck's *constant of action*. We have already referred a number of times to this fundamental constant of atomic physics. It has the dimensions of energy times time, *i.e.*, erg seconds in the cgs system.

From this radical postulate made by Planck there developed the *quantum theory*, which has played a major role in the development of twentieth-century physics. In the quantum theory the radiant energy, too, is *quantized*, the energy of the discrete packets, or *quanta*, of radiation being hf. From every application of the quantum concept—for instance, from the study of the photoelectric effect (Chap. 35)— the fundamental constant h may be shown to have the value

$$h = 6.62 \times 10^{-27} \text{erg sec} \tag{48.6}$$

Since the frequency f of visible light of wavelength 5,000A is 6×10^{14} sec^{-1}, the products hf are evidently very small energies. But the quantum theory deals only with individual atoms and molecules and their interaction with radiation. Many extremely important discoveries in atomic and nuclear physics have resulted from the impetus given to these fields of experimentation by the quantum theory.

The details of this theory, which we have very inadequately sketched, form the subject matter of advanced courses in physics. We do not derive Planck's formula in this book but state it without proof because of its fundamental nature. From it, as will be seen, all the properties of the black-body spectrum can be deduced with almost miraculous accuracy.

Planck's formula for the radiant emittance dW of a black body within a range of frequencies between f and $f + df$ is

$$dW = \frac{2\pi h}{c^2} \frac{f^3}{e^{hf/kT} - 1} df \tag{48.7}$$

where c is the velocity of the light, k is the Boltzmann constant, or gas constant per molecule, and T is the absolute temperature of the black body. Equation (48.7) may be expressed in terms of wavelengths by the following substitutions: Since $c = f\lambda$, $hf/kT = hc/\lambda kT = c_2/\lambda T$, where c_2 is a constant equal to hc/k.

Also, $f^3 = c^3/\lambda^3$, and $df = -(c/\lambda^2)d\lambda$. Hence

$$\begin{aligned} dW &= -\frac{2\pi h}{c^2} \frac{c^3/\lambda^3}{e^{c_2/\lambda T} - 1} \frac{c}{\lambda^2} d\lambda \\ &= -\frac{c_1\lambda^{-5}}{e^{c_2/\lambda T} - 1} d\lambda \end{aligned} \tag{48.8}$$

where $c_1 = 2\pi hc^2$. We shall disregard the minus sign since we are interested only in the magnitude of the quantities. If we denote by W_λ the ratio $dW/d\lambda$, or the radiant emittance of a black body per unit range of wavelength, then

$$W_\lambda = \frac{c_1\lambda^{-5}}{e^{c_2/\lambda T} - 1} \tag{48.9}$$

The quantity W_λ is called the *spectral emittance.* Numerically,

$$c_1 = 3.740 \times 10^{-12} \text{ watt cm}^2$$
$$c_2 = 1.4385 \text{ cm deg}$$

Figure 48.9 is a plot to scale of the spectral emittance of a black body at four different temperatures. Since Planck's law gives these distributions accurately, it should be possible to derive from this law all the other relations holding for black-body radiation. We derive the most important of these radiation laws in the following section.

48.8. Other Black-body Radiation Laws. Recalling that

$$e^x = 1 + x + (x^2/2!) + (x^3/3!) + \cdots,$$

suppose that λ is so large that $c_2/\lambda T \equiv x$ is negligible for all powers higher than the first. For such long wavelengths, Eq. (48.9) becomes

$$(W_\lambda)_{\lambda\to\infty} = \frac{c_1\lambda^{-5}}{c_2/\lambda T} = \frac{c_1 T}{c_2 \lambda^4} \tag{48.10}$$

which is the *Rayleigh-Jeans radiation law* and was first derived from kinetic-theory considerations applied to the radiation "gas." The law holds well for the long wavelengths, considerably beyond the maxima of the curves.

Again set $c_2/\lambda T = x$, and hold T constant. Then $\lambda = c_2/Tx$ and Eq. (48.9) becomes

$$W_\lambda = \frac{c_1 T^5 x^5}{c_2{}^5(e^x - 1)}$$

Set $x^5/(e^x - 1) = \phi_\lambda$. To find the wavelength λ_m at which Eq. (48.9) is a maximum, we must differentiate ϕ_λ with respect to λ and set the result equal to zero. Thus

$$\frac{d\varphi_\lambda}{d\lambda} = \frac{d\phi_\lambda}{dx}\frac{dx}{d\lambda} = \frac{(e^x - 1)5x^4 - (x^5)(e^x)}{(e^x - 1)^2}\frac{dx}{d\lambda} = \frac{x^4[(5 - x)e^x - 5]}{(e^x - 1)^2}\frac{dx}{d\lambda} = 0$$

Consequently either $(5 - x)e^x = 5$, or $x = 0$, or $x = \infty$. Considering the first of these possibilities,

$$\left(1 - \frac{x}{5}\right)e^x = 1$$

and by graphical solution, the value of x at the maximum, x_m, is 4.965. Thus, $c_2/\lambda_m T = 4.965$, or

$$\lambda_m T = \frac{c_2}{4.965} = \text{const}$$

Since $c_2 = 1.4385$ cm deg if λ is in centimeters and T in degrees Kelvin,

$$\boxed{\lambda_m T = 0.2897} \tag{48.11}$$

This is *Wien's displacement law*, which explains the shift of the maximum of the emittance curve to shorter wavelengths as the temperature is raised. This shift is evident for the four curves plotted in Fig. 48.9. The limits of the visible spectrum are there indicated by the vertical dashed lines. This shift of λ_m toward the blue with increasing T accounts for the fact that as incandescent bodies become hotter their color changes from red to a "blue-white." At 6000° K, which is very nearly the surface temperature of the sun, the maximum of the spectral-emittance curve falls approximately at 5,000A.

What is the physical interpretation of $x = 0$ and $x = \infty$ which are also solutions of $d_{\varphi_\lambda}/d\lambda = 0$? Now $x = c_2/\lambda T = hc/k\lambda T$, and therefore $x = 0$ means $\lambda = \infty$, whereas $x = \infty$ means $\lambda = 0$. These are of course *minimum* values of the φ_λ vs. λ radiation distribution curves.

To obtain the total radiant emittance W, which is represented by the area under the spectral-emittance curve, we must evaluate the integral

$$W = \int_0^\infty W_\lambda \, d\lambda$$

Placing $c_2/\lambda T = x$, $d\lambda = -(c_2 dx/Tx^2)$. Hence

$$\int_0^\infty W_\lambda d\lambda = -\int_\infty^0 \frac{c_1 T^5 x^5 c_2 dx}{c_2{}^5(e^x - 1)Tx^2} = \frac{c_1 T^4}{c_2{}^4} \int_0^\infty \frac{x^3 dx}{e^x - 1}$$

The definite integral

$$\int_0^\infty \frac{x^3 dx}{e^x - 1}$$

has the value $\pi^4/15$. Therefore, replacing the constant $\frac{1}{15}\,(\pi/c_2)^4 c_1$ by σ, we have

$$\boxed{W = \sigma T^4} \qquad \text{(black body)} \qquad (48.12)$$

where the constant σ is found to have the value 5.67×10^{-12} when W is in watts per square centimeter of the black-body surface and T is in degrees Kelvin. Equation (48.11) is the famous *Stefan-Boltzmann law*, first proposed by Stefan in 1879 and later derived theoretically by Boltzmann.

On combining Eqs. (48.4) and (48.11) the radiant emittance from any real body is given by

$$W = a\sigma T^4 \qquad (48.13)$$

and since the absorptance a equals the emissivity e, this equation may also be written

$$W = e\sigma T^4 \qquad (48.14)$$

48.9. Radiation Exchange. Equation (48.13) gives the rate at which radiant energy is *emitted* by a body per unit area of its surface. The body will also be absorbing radiation incident on it, and the net gain or loss of energy is the difference between the rates of emission and absorption.

Consider once again the small body suspended in the evacuated enclosure, and suppose that thermal equilibrium at the wall temperature T_1 has been attained. Since the absorptance a equals the emissivity e for the small body, Eq. (48.3) may be written

$$R = \frac{e\sigma T_1^4}{e} = \sigma T_1^4$$

and thus we see that *the irradiance R on the surface of the body depends only on the temperature* T_1.

If the body is at a temperature T_2 different from that of the walls and if it is small enough so that its changed temperature has a negligible effect on the radiant energy in the large box, the irradiance R is not changed but the rate of emission of radiant energy is now $SW = Se\sigma T_2^4$, while its rate of absorption remains $SeR = Se\sigma T_1^4$. The *net* rate of emission is thus

$$Se\sigma(T_2^4 - T_1^4) \tag{48.15}$$

If the body is not small compared with the enclosure and if the walls are not perfectly absorbing, some of the radiation emitted by the body is reflected back to it from the walls, so that the amount of radiant energy incident on the body is changed. It is thus impossible to treat the general case quantitatively, for real objects of finite size with partly reflecting walls are exchanging radiant energy at all times with all surfaces in the "line of sight."

The distribution curve for the radiation from the sun is closely that of a black-body radiation curve at 6000°K. The average rate at which this radiant energy is received at the top of the earth's atmosphere, known as the "solar constant," is 0.033 cal cm^{-2} sec^{-1}, or 1.15 kw/yd^2. As mentioned in Sec. 23.1, about 10 per cent of this radiant energy is absorbed by our atmosphere. The remainder falls on the earth's surface and is partly absorbed and partly reflected. The earth's surface also emits radiation at a rate corresponding to its lower temperature and the average emissivity of its surface. This emitted and reflected radiation is in part absorbed by our atmosphere and in part transmitted into space. The net result of all these processes is that, of the radiant energy received from the sun, 17 per cent is absorbed by the atmosphere and 40 per cent by the earth's surface and the remaining 43 per cent reflected and radiated back into outer space.

48.10. Summary of Radiant Energy Terms

Radiant flux is the total radiant energy falling on a surface or emitted by a surface per second. It is measured in watts.

Irradiance is the radiant flux incident on unit area of a surface. We denote it by the symbol R and use 1 watt/cm^2 as the unit.

Radiant emittance is the total radiant energy emitted by a surface per unit area per second. It is denoted by W, and we use 1 watt cm^{-2} sec^{-1} as the unit.

Reflectance r is the fraction of the incident flux reflected by a surface.

Absorptance a is the fraction of the incident flux absorbed by a surface. If the surface is opaque, $r + a = 1$.

Emissivity e is the *ratio* of its radiant emittance W to that of a black body at the same T. Also $e = a$.

Spectral emittance W_λ is $dW/d\lambda$, or the radiant emittance per unit range of wavelength. A convenient unit of W_λ is 1 watt cm^{-2} sec^{-1} $m\mu^{-1}$.

PROBLEMS

1. For a certain dielectric material the polarizing angle is found to be 53°. What is the index of refraction?

2. Calculate the polarizing angle for dense flint glass of index $n = 1.655$.

3. Calculate the limits of the polarizing angle for white light 4,000 to 7,500A for crown glass with $n = 1.525$ and 1.505, respectively, at these wavelengths.

4. Two Nicol prisms are mounted as polarizer and analyzer, first with an angle of 10° between their principal sections, then with this angle increased to 70°. What is the ratio of the intensity of the transmitted light in the two cases? (Use the law of Malus.)

★5. A beam of plane-polarized light is incident on a calcite crystal, with its vibrations making an angle of 30° with the principal section of the crystal. Calculate the relative amplitudes and intensities of the two refracted beams.

6. The irradiance on an opaque surface of area 2 m^2 is 1 watt/cm^2, of which one-fourth is reflected. Find (*a*) the absorptance, (*b*) the reflectance, (*c*) the total incident radiant flux, (*d*) the radiant emittance from the surface if it is in thermal equilibrium.

7. Calculate the radiant emittance of a black body at a temperature of (*a*) 300 °K; (*b*) 3000 °K.

8. For the black body at 3000 °K calculate the ratio of its spectral emittance W at 10,000A in the near infrared and at 5,000A in the visible.

9. Compute the wavelength at which the spectral emittance of this black body at 3000 °K is a maximum.

10. Calculate the temperature of a black body that has the maximum of its spectral-emittance curve at 5,500A, where the eye is most sensitive.

11. A cube of copper, 1 cm on edge, is suspended within a large evacuated enclosure whose walls are at 300 °K. If the emissivity of copper is 0.3, what power input is required to maintain the cube at 600 °K, neglecting heat conduction?

12. In an evacuated 100-watt lamp bulb the tungsten filament operates at 2450 °K. If the emissivity of tungsten is 0.35 and the glass surface is at 350 °K, find the surface area of the filament. Assume that all of the radiant energy is absorbed by the glass.

13. If the power input to the copper cube of Prob. 11 ceased, how long would it take for the cube to cool to 599 °K? The density of copper is 8.9 gm/cm^3, and its specific heat is 0.093 cal gm^{-1} deg C^{-1}.

14. Assuming that the surface temperature of the sun is exactly 6000°K and that it is a perfect radiator, at what rate does it lose energy? The radius of the sun is 4.3×10^5 miles.

15. Show that Newton's law of cooling (Sec. 17.9) is obeyed by a body losing energy by radiation to its surroundings if the temperatures are nearly equal. [Hɪɴᴛ: Differentiate Eq. (48.14)].

CHAPTER 49
ATOMIC STRUCTURE

49.1. The Periodic System of the Elements. If the elements are arranged in order of increasing atomic weight, there becomes apparent a periodicity in the properties of the elements. This was first discovered by Mendelyeev with regard to the chemical properties of the elements, but the physical properties show the same periodicity. The order number of an element in this arrangement, starting with hydrogen as 1, is

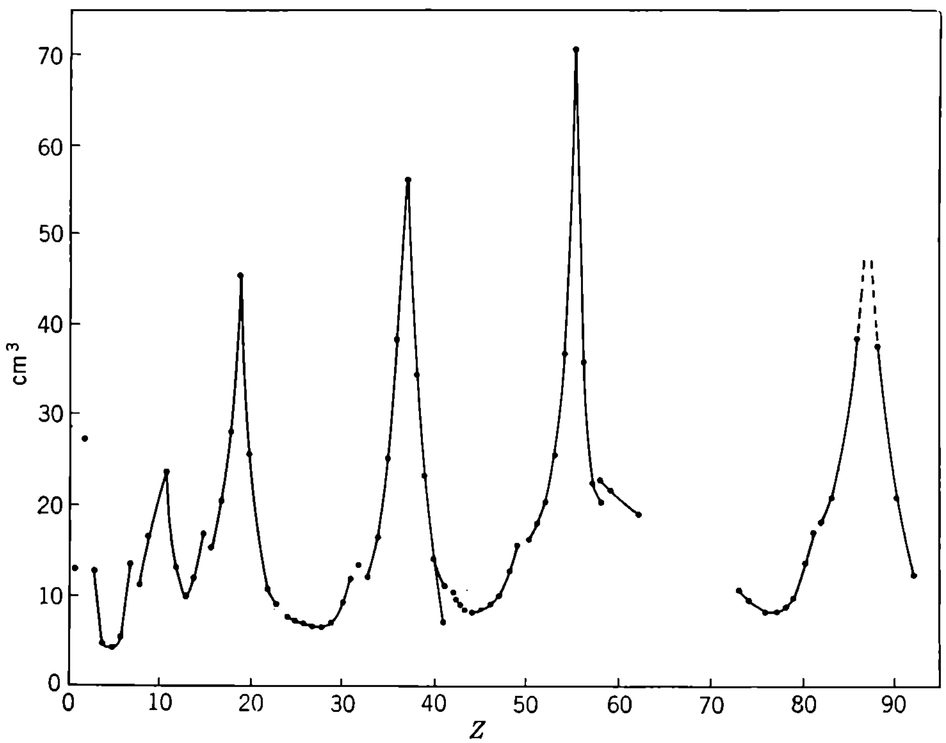


FIG. 49.1. The atomic volumes of the elements.

called the *atomic number* Z of the element. The chemical properties of hydrogen ($Z = 1$), lithium ($Z = 3$), sodium ($Z = 11$), potassium ($Z = 19$), rubidium ($Z = 37$), and cesium ($Z = 55$) are very nearly the same. The successive differences in the values of Z in this series are 2, 8, 8, 18, 18, and these differences are therefore the periods for the recurrence of a given property. Thus the series of inert gases, He, Ne, A, Kr, Xe, has this same series of differences of atomic numbers, with the exception of the first difference. Many other examples can be found as well, such as the ability of a series of elements to form an oxide or a hydride of a given type.

The same periodicity is revealed in the physical properties. The simplest example of this is the fact that the inert gases are all in the gaseous state at ordinary temperatures. Figure 49.1 shows the atomic volume (the volume in cubic centimeters of one mole of atoms) of the elements, plotted against atomic number. Some other physical properties are shown in Fig. 49.2. The curves for all these properties are strikingly similar, and all show the fundamental periodicity that the chemical properties show. Much more striking examples of the perio-

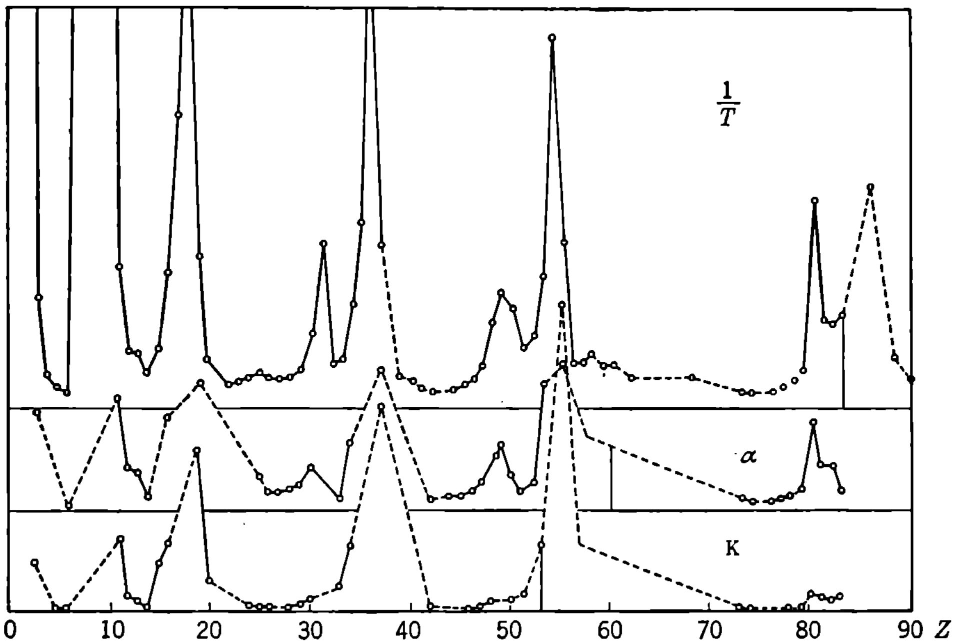


FIG. 49.2. The reciprocal of the melting point $1/T$, the coefficient of thermal expansion α, and the compressibility K, of the elements.

dicity of physical properties are afforded by the spectra of the various elements, both in the optical and in the X-ray region.

Mendelyeev believed that the period 8 should prevail through the whole series, and as a result the periodic table is usually arranged as in Table 49.1. This arrangement has persisted because other schemes based on the correct periodicity are much less convenient and compact. Above the element barium ($Z = 56$) apparent irregularities occur. The whole series of elements from lanthanum ($Z = 57$) to lutecium ($Z = 71$), the rare earth elements, are assigned to a single place in the table, and a similar behavior appears to exist for the very heaviest elements ($Z = 89$ to 96). We shall postpone the discussion of these irregularities for the moment.

The scattering experiments with α particles described in Sec. 24.7

Table 49.1. *Periodic Chart of the Atoms*

Atoms grouped according to the number of outer (valence) electrons

	I	II	III	IV	V	VI	VII	VIII		
1	1 H 1.008							2 He 4.003		
2	3 Li 6.94	4 Be 9.02	5 B 10.8	6 C 12.01	7 N 14.01	8 O 16.00	9 F 19.00	10 Ne 20.18		
3	11 Na 23.00	12 Mg 24.32	13 Al 26.97	14 Si 28.06	15 P 30.98	16 S 32.06	17 Cl 35.46	18 A 39.94		
4	19 K 39.10	20 Ca 40.08	21 Sc 45.10	22 Ti 47.90	23 V 50.95	24 Cr 52.01	25 Mn 54.93	26 Fe 55.85	27 Co 58.94	28 Ni 58.69
	29 Cu 63.57	30 Zn 65.38	31 Ga 69.72	32 Ge 72.60	33 As 74.91	34 Se 78.96	35 Br 79.92	36 Kr 83.7		
5	37 Rb 85.48	38 Sr 87.63	39 Y 88.92	40 Zr 91.22	41 Cb 92.91	42 Mo 95.95	43 Tc 99	44 Ru 101.7	45 Rh 102.7	46 Pd 106.7
	47 Ag 107.9	48 Cd 112.4	49 In 114.8	50 Sn 118.7	51 Sb 121.8	52 Te 127.6	53 I 126.9	54 Xe 131.3		
6	55 Cs 132.9	56 Ba 137.4	57* La 138.9	72 Hf 178.6	73 Ta 180.9	74 W 183.9	75 Re 186.3	76 Os 190.2	77 Ir 193.1	78 Pt 195.2
	79 Au 197.2	80 Hg 200.6	81 Tl 204.4	82 Pb 207.2	83 Bi 209.0	84 Po 210	85 At 211	86 Rn 222		
7	87 Fa 223	88 Ra 226.0	89** Ac 227.0							
*6	58 Ce 140.1	59 Pr 140.9	60 Nd 144.3	61 147	62 Sm 150.4	63 Eu 152.0	64 Gd 156.9	65 Tb 159.2		
	66 Dy 162.5	67 Ho 164.9	68 Er 167.2	69 Tm 169.4	70 Yb 173.0	71 Lu 175.0				
**7	90 Th 232.1	91 Pa 231	92 U 238.1	93 Np 237	94 Pu 239	95 Am 241	96 Cm 242			

* The *6 group, called the rare earths, is to be inserted in the same box as La 57.
* The **7 group, which is similar to the rare earths, is to be inserted in the same box as Ac 89.

made possible a determination of the number of elementary charges on the nucleus of an atom, and it was found that this number is equal to the atomic number of the element in the periodic table. The nucleus of an atom thus has a positive charge equal to Ze; and since the atom as a whole is electrically neutral, the number of electrons is equal to Z. Thus hydrogen has 1 electron, helium 2, lithium 3, and uranium 92 electrons, each of charge e. The number of electrons in an atom can be found, of course, in many other ways, all of which agree. Perhaps the simplest of these methods is the study of spectra.

49.2. Spectra of the Hydrogen Atom.[1] Since hydrogen is the simplest of all atoms, consisting of just the proton with unit positive charge

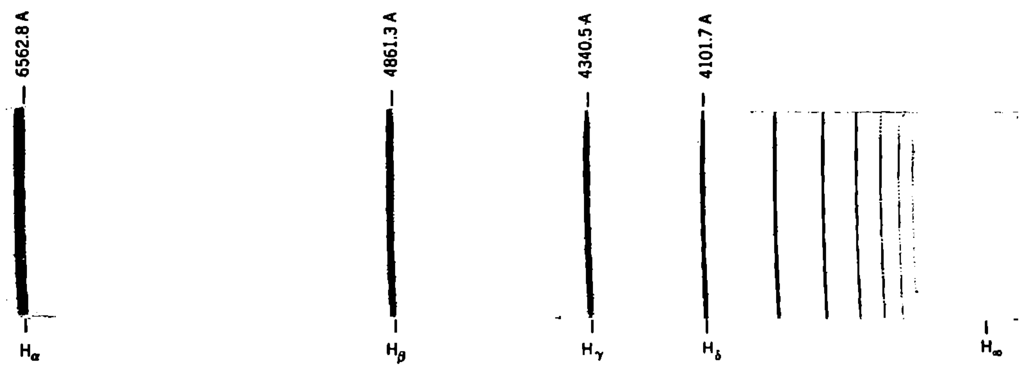


FIG. 49.3. The Balmer series of atomic hydrogen. *(Reproduced by permission from "Atomic Spectra and Atomic Structure," by G. Herzberg, Dover Publications, New York,* 1946.)

around which a single negative electron circulates, it might be expected to produce a simple line spectrum. This is indeed the case. Figure 49.3 is a reproduction of a spectrogram showing the so-called Balmer series of atomic hydrogen. The first line in the red at 6,563A is known as H_α, the second line in the blue at 4,861A is the H_β line, the next line is designated H_γ, etc. J. J. Balmer (1825–1898) in 1885 discovered that the wavelengths λ of these lines are given accurately by the formula

$$\frac{1}{\lambda} = R\left(\frac{1}{2^2} - \frac{1}{n'^2}\right) \tag{49.1}$$

where R is a constant called the Rydberg constant and n' has the value 3 for H_α, 4 for H_β, etc. If λ is in centimeters, $1/\lambda$ gives the number of waves per centimeter, which is called the *wave number* of the spectral line. In these units, $R = 109{,}678\ \text{cm}^{-1}$. As n' increases, the successive lines of the series become closer and closer together and converge to a limit H_∞.

[1] Review Sec. 11.9.

If $n' = \infty$, the convergence limit for the series is given by Eq. (49.1) at $\lambda = 3{,}646$A, which is in the near ultraviolet. A rather faint continuous spectrum proceeds from this limit farther into the ultraviolet.

Following Balmer's discovery, other formulas of a more general nature were developed for series found in the spectra of the alkali and alkaline-earth elements. All these formulas had the common feature that the wave numbers or frequencies ($c = f\lambda$, and hence $1/\lambda = f/c$) of the lines were given as the difference between two quantities or *terms*. For example, in the Balmer formula the two terms are $R/2^2$ and R/n'^2. But until the work of Niels Bohr in 1913 no satisfactory theoretical explanation of the production of sharp spectral lines and of these series relationships had been produced. In that year Bohr brought out his famous and revolutionary theory of the hydrogen atom, which we have already discussed in Sec. 11.9 in so far as it was an example of motion under an inverse-square law of force.

It will be recalled that Bohr adopted the Rutherford model of the atom, which has the electrons revolving in orbits about the more massive, positively charged nucleus. Now an electron moving in a circular path has an acceleration toward the center of the circle, and according to classical electromagnetic theory an accelerated electron radiates a continuous spectrum of energy. Classically, then, the electron would gradually lose kinetic energy and would spiral into the nucleus of the atom. Actually, however, a line spectrum is emitted. To get out of this dilemma, Bohr made his famous postulate that *an electron in an atom can revolve in certain fixed orbits without the emission of radiant energy* and that furthermore *only those orbits are allowed for which the angular momentum is an integral multiple of $h/2\pi$*, where h is Planck's constant. To explain the emission of light by the atom, Bohr made a second postulate that *the atom radiates only when an electron jumps from one orbit to another of lower energy and that the frequency f of the emitted radiation is determined by the difference in energy between the initial and final orbit, or "state," according to the relation*

$$E' - E'' = hf \tag{49.2}$$

where E' is the energy of the initial state and E'' that of the final state. The quantity of radiation possessing this energy hf is called a *photon*.

It was shown in Sec. 11.9 that, as a result of the first postulate, the radii r of the circular orbits of the single electron of charge $-e$ and mass m rotating about the single proton of charge $+e$ in the hydrogen atom are given by

$$r = \frac{n^2h^2}{4\pi^2me^2} \tag{49.3}$$

which is Eq. (11.24). To calculate the radius r_0 of the first Bohr orbit, which corresponds to the normal state of hydrogen, we substitute $n = 1$, $h = 6.62 \times 10^{-27}$ erg sec, $m = 9.11 \times 10^{-28}$ gm, and $e = 4.80 \times 10^{-10}$ esu, so that

$$r_0 = \frac{(6.62 \times 10^{-27})^2}{4 \times 9.87 \times 9.11 \times 10^{-28} \times (4.80 \times 10^{-10})^2} \text{ cm}$$
$$= 0.528 \times 10^{-8} \text{ cm}$$

This is in fair agreement with the indications from other methods, such as X-ray diffraction, which show that the diameter of an atom is about 10^{-8} cm. The other allowed orbits, according to Eq. (49.3), have radii $4r_0$, $9r_0$, $16r_0$, etc.

The total energy E of the electron in any of its orbits is given by Eq. (11.27),

$$E = -\frac{2\pi^2 me^4}{n^2h^2} \tag{49.4}$$

These energies are negative because the zero of potential energy is taken with the electron at an infinite distance from the nucleus. When $n = 1$, the energy has the largest negative value. The greater the value of n, the less negative the energy. Therefore the energy is *lowest* in the normal state, $n = 1$, and the energy is *higher* the larger the orbit. The atom is always in the normal state, unless either by collision with a fast-moving electron or by absorption of a photon of the correct energy its electron is temporarily raised to one of the outer orbits. The atom, however, stays in this "excited" state for only about 10^{-8} sec before returning directly or by steps to the normal state by emission of one or more photons.

By substituting Eq. (49.4) in Eq. (49.2) the frequency of the emitted photon becomes

$$f = \frac{2\pi^2 me^4}{h^3}\left(\frac{1}{n''^2} - \frac{1}{n'^2}\right) \tag{49.5}$$

where n' is the quantum number of the larger orbit from which the electron jumps and n'' is the quantum number of the smaller orbit to which it jumps. The wave number $1/\lambda$ corresponding to any of these frequencies is

$$\frac{1}{\lambda} = \frac{2\pi^2 me^4}{ch^3}\left(\frac{1}{n''^2} - \frac{1}{n'^2}\right) \tag{49.6}$$

Comparison of this equation with the Balmer equation, Eq. (49.1), shows that the two are identical if

$$R = \frac{2\pi^2 me^4}{ch^3} \tag{49.7}$$

and if $n'' = 2$. If we substitute the known values of m, e, c, and h in Eq. (49.7), we obtain precisely the observed value of the Rydberg constant in the Balmer equation. The lines of the Balmer series of hydrogen are therefore given *very exactly* by the Bohr theory.

Bohr was also able to predict other series of lines in the spectrum of the hydrogen atom. By substituting $n'' = 1$ and $n' = 2, 3, 4, \ldots$ one calculates a spectral series in the far ultraviolet region. This series was discovered by T. Lyman of Harvard University, and the measured wavelengths were found to agree exactly with the predicted values from the Bohr theory. The *Lyman series* arises, then, from electrons returning from the various excited states of hydrogen directly to the normal state. This and other possible series of lines are shown on a diagram of the Bohr circular orbits of hydrogen in Fig. 49.4. If n'' is set equal to 3 and n' to 4, 5, 6, . . . , the calculated spectrum lines lie in the infrared. These lines were looked for and found by F. Paschen at precisely the computed wavelengths. Another series of lines arising from electron jumps to orbit $n = 4$ was predicted and observed by Brackett. Pfund observed further in the infrared the first lines of yet another series arising from electron jumps to the $n = 5$ orbit.

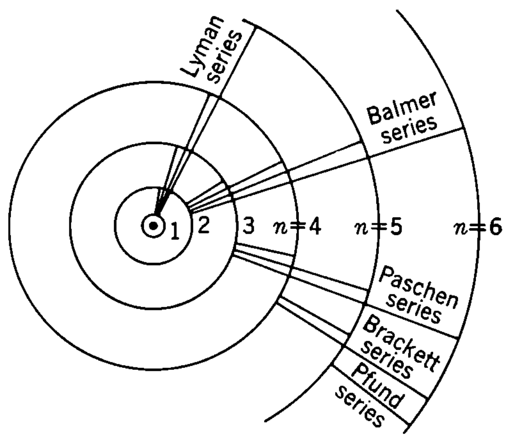


FIG. 49.4. The Bohr circular orbits for hydrogen, showing the electron jumps giving rise to exactly predicted spectral series.

★49.3. Deuterium vs. Hydrogen. In 1932, Urey, Brickwedde, and Murphy used a further prediction of the Bohr theory of the hydrogen spectrum to aid them in making the important discovery that hydrogen has a heavy isotope, now called deuterium, of atomic weight 2. In deriving Eq. (49.4) we have assumed that the electron mass m rotates about a nucleus of infinite mass, *i.e.*, a stationary nucleus. Actually, however, the nucleus has a finite mass M, and the two masses m and M rotate about their center of mass. Equations (49.4), (49.5), and (49.7) may then be shown to be still valid if m is replaced by the *reduced mass* μ, defined by the equation

$$\mu = \frac{mM}{m + M} = \frac{m}{1 + (m/M)} \tag{49.8}$$

Obviously μ approaches the value m as $M \to \infty$. For ordinary hydrogen, M is the proton mass equal to 1,840 times the electron mass, whereas for deuterium the nucleus, called the *deuteron*, has double this mass. The ratio of the Rydberg constant R for deuterium to that for hydrogen is equal to the ratio μ_D/μ_H of their reduced masses, or $(1 + 1/1{,}840)/(1 + \frac{1}{2} \times 1{,}840) = 3{,}682/3{,}681$.

This variation in R produces a small but detectable wavelength shift for all corresponding lines of these two varieties of hydrogen. According to Eq. (49.6) the D_α line should have a wavelength 1.793A shorter than that for the H_α line, and this was exactly the shift first observed by Urey and his collaborators within the limits of accuracy of their measurements. In ordinary hydrogen the heavy variety is present to only 1 part in 4,000, and consequently the close-lying deuterium line accompanying every hydrogen line is too weak to be detected. In so-called "heavy water," obtained by repeated electrolysis of ordinary water, the light hydrogen atoms are largely replaced by deuterium atoms, however, and the spectrum of the hydrogen gas obtained from such water shows at high dispersion all the Balmer-series lines as doublets with the deuterium components in exactly their predicted positions.

Deuterium has proved to be a valuable research tool. With a high-energy deuteron beam in a cyclotron, many artificially radioactive materials may be made. Heavy water is also an effective slowing agent for neutrons in a uranium nuclear reactor.

49.4. Spectra of Alkali Atoms. Ionization Potential. If a small amount of one of the alkali metals, lithium, sodium, potassium, rubidium, or cesium, is placed in a glass tube containing an inert gas such as nitrogen (to prevent oxidation of the alkali) and if the tube is then heated moderately, enough of the alkali is vaporized so that its absorption spectrum may be obtained. The absorption tube is placed between the source emitting a continuous spectrum and the slit of the spectrograph. Figure 49.5 is a spectrogram showing the series of absorption lines (except for the strong yellow line 5,893A) so obtained with sodium vapor. All elements in the gaseous state produce some line absorption, usually in the ultraviolet, but only the alkalies show a long series of this simple type.

The similarity of the arrangement of lines in this series to that in the Balmer series for hydrogen is evident. This fact, among others, led Bohr and Stoner to propose that the filling of each new shell of electrons (Sec. 49.5) starts with the alkali elements. Thus each alkali atom has one electron

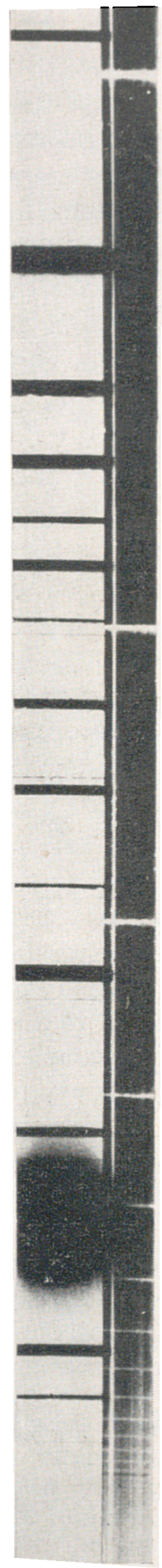


FIG. 49.5. *A*. Emission spectrum of mercury vapor. *B*. Absorption spectrum of sodium vapor.

outside all the others in an orbit by itself. This electron is called the *valence electron*, and for sodium its orbit has $n = 3$. When light of just the correct frequency f strikes the normal sodium atom, the energy hf of this quantum raises this outer electron from its normal orbit to some outer orbit. Thus by light absorption the atom becomes *excited*. At the limit of this absorption series the energy hf of the incident light is sufficient to raise the electron to the orbit $n = \infty$, that is, to knock the electron completely away from the atom. This process is known as *ionization*, and the energy that will just separate the valence electron from the remainder of an atom is called its *ionization energy*.[1]

The value of hf corresponding to the series limit of 2,413A gives a spectroscopic value of the ionization energy of the sodium atom. Now since $E = hf$, there is a conversion ratio between frequency, or wave-number units, and any energy units. In fact, 8,066 $cm^{-1} = 1$ ev. For the limit 2,413A, $1/\lambda = 41{,}472\ cm^{-1}$, and therefore the ionization potential of sodium is $41{,}472/8{,}066 = 5.14$ ev. This is by far the most accurate method of determining this important atomic constant.

When an electric discharge is sent through sodium vapor this same *principal series* of lines is observed in emission, for then many of the atoms become excited and subsequently emit these frequencies when the valence electron drops back into the $n = 3$ orbit. In addition many other transitions of the electron back to intermediate orbits, and then finally to the normal state, are possible. Therefore the emission spectrum of sodium is much more complex than this simple, one-series absorption spectrum.

***49.5. Shells of Electrons.** The properties of an atom of an element can be divided into two classes. In the first class the property in question (the atomic weight, for example) is a monotonic function of atomic number Z; in the second class the property (chemical behavior, for example) is a periodic function of Z. Now the atomic weight is almost wholly the weight of the inner core (the nucleus) of the atom, and it might be expected that other monotonic properties depend upon the nucleus and those electrons that lie near to it. On the other hand the chemical properties, which are manifestations of the forces between atoms, must depend upon the outermost electrons. The fact that the chemical properties are periodic suggests that similar structures of the outer electrons in an atom recur periodically in the progression from the light to the heavy elements.

In Sec. 49.2 we have shown that a series of electron orbits exist in the hydrogen

[1] Ionization energies are often measured by bombarding the atom with electrons. The potential difference through which the bombarding electron must fall in order to attain a velocity necessary to ionize the atom is known as the *ionization potential*. Thus the ionization energy in electron volts (ev) is equal numerically to the ionization potential in volts.

atom, each having a radius proportional to the square of an integer called the total quantum number n. In atoms with more than one electron a similar situation obtains, but there may be several orbits with a given value of n, with an electron in each orbit. The electrons in a complex atom are thus arranged in shells, each shell characterized by a value of n and having a radius proportional to n^2 The number of electrons that can move in a given shell is strictly limited, however, and the maximum number is equal to $2n^2$. For $n = 1$ there are two electrons: for $n = 2$, eight electrons, and so forth. The various shells have been labeled $K, L, M, N, \ldots$ for $n = 1, 2, 3, 4, \ldots$. Since the number of electrons in an atom increases as Z increases, the new electrons usually have the smallest possible value of n. When a shell becomes filled, the next electron must have the next larger value of n and a new shell of electrons is started. The properties of the shells are summarized in Table 49.2.

When a new shell is started, with one of the alkalies, the atom has one electron that is moving far outside all the others and this results in characteristic properties of the atom that are periodic with the atomic number. Let us examine in detail how this occurs. In hydrogen the single electron is in the K shell; in helium both electrons are K electrons, and the first shell is completed. In lithium two electrons fill the K shell, and the third starts the new L shell. Hydrogen and lithium each have a single outer electron, and this results in a number of very similar properties. As Z increases, the new electrons are added to the L shell until with neon ($Z = 10$) the L shell is complete. Neon and helium thus have the complete shells of electrons, which are responsible for the properties of the series of inert gases. The shell structures of atoms of low atomic number are given in Table 49.3.

Table 49.2. *The Electron Shells in an Atom*

Name of shell	Value of n	Max number of electrons in shell	Total number of electrons when shells are full
K	1	2	2
L	2	8	10
M	3	18	28
N	4	32	60
O	5	50	110

The rules of chemical combination are also to be understood in terms of the behavior of the outer electrons in the atom. For the elements on the left side of the periodic table, the valence is just equal to the number of outer electrons. The alkali metals have a valence of 1 and have one outer electron; the alkaline earths have a valence of 2 and have two electrons near the boundary of the atom. When a chemical compound is formed, such as sodium chloride, the single "valence" electron of sodium becomes attached to the chlorine atom, a subshell of eight electrons is formed in the chlorine atom, and the sodium ion is also left with a subshell of eight electrons. As we saw earlier, such octets of electrons have a high stability and the two ions Na^+ and Cl^- are attracted by strong electrical forces. In a similar manner the two outer electrons of calcium become attached to chlorine atoms when calcium chloride, $CaCl_2$, is formed, and each of the three atoms has the stable octet subshell in the compound. Chemical compounds of this type are called *polar* compounds since each pair of ions forms an electric dipole. The inert gases, then, enter into no chemical combination because they already have this extremely stable shell structure. Not all chemical compounds are held together by electrical forces of this kind. In organic compounds and in molecules like H_2 and N_2, for example, the binding forces arise from other, more complex causes. Nevertheless the structures of the outer electron shells still determine these forces.

***49.6. Elements of Higher Atomic Number.** From sodium to argon the electrons are added to the M shell, since both the K and L shells are complete. In argon the M shell contains eight electrons. Although this does not complete the shell, these eight electrons form a stable configuration, which behaves as a subshell and which is responsible for the inert behavior of argon. Next in order to argon is potassium, and the new electron is added, not to the still incompleted

Table 49.3. *The Shell Structure of Atoms*

Element	Z	K shell	L shell	M shell	N shell
H	1	1			
He	2	2			
Li	3	2	1		
Be	4	2	2		
B	5	2	3		
C	6	2	4		
N	7	2	5		
O	8	2	6		
F	9	2	7		
Ne	10	2	8		
Na	11	2	8	1	
Mg	12	2	8	2	
Al	13	2	8	3	
Si	14	2	8	4	
P	15	2	8	5	
S	16	2	8	6	
Cl	17	2	8	7	
A	18	2	8	8	
K	19	2	8	8	1
Ca	20	2	8	8	2
Kr	36	2	8	18	8

M shell, but to the N shell. Thus potassium, with the properties of an alkali, has a single outermost electron.

Above potassium the new electrons are added some to the N shell and some to the M shell, with the result that for krypton ($Z = 36$) the M shell has its complement of 18 electrons and a subshell of 8 electrons has been formed in the N shell.

In elements of higher atomic number the arrangement of the electrons seems to become irregular; electrons are added to several outer shells before the inner shells are filled. It is this circumstance which explains the chemical behavior of the rare earths. All the elements from lanthanum ($Z = 57$) to lutecium ($Z = 71$) are assigned a single place in the periodic table because their chemical properties are almost identical. Each of these elements has two outer electrons in the P

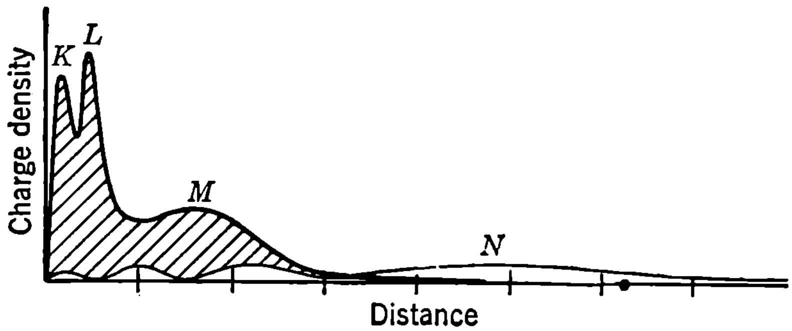


FIG. 49.6. Charge distribution in potassium as a function of the distance from the center of the atom.

shell and is, therefore, chemically similar to calcium. As the atomic number is increased, the additional electrons are added to complete the N shell. Lutecium has the K, L, M, N shells complete, nine O electrons, and two P electrons.

In the elements beyond lutecium more electrons are added, but neither the O shell nor the P shell is completed, since no elements of sufficiently high atomic number exist. The Q shell appears to be started with element 87.

Although a detailed study of the motions of the electrons in a complicated atom is not possible, it is possible to calculate by the methods of *quantum mechanics* (Sec. 49.11) the average charge density as a function of the distance from the nucleus. Figure 49.6 shows the charge distribution for potassium, one of the alkali metals. The inner core of electrons is represented by the shaded area; the single outermost electron is represented by the unshaded area under the curve. Peaks that result from the individual shells are clearly visible, although there is no evidence for sharp boundaries between the shells. The unshaded area extends to very small distances, which means that the outermost electron is sometimes found as close to the nucleus as a K electron. At large distances from the nucleus the curve bounding the unshaded area approaches zero rapidly, but no definite point can be chosen as the outer boundary of the atom. In consequence the shape or size of an atom cannot be defined precisely.

***49.7. Ionized Atoms.** We have seen how an atom such as sodium can lose an electron by the process of light absorption. Atoms may also become ionized by collisions with fast-moving electrons in an electric discharge, and these collisions may in addition excite one of the remaining electrons of the atom to an outer orbit. This excited electron in the ionized atom may then drop back into an orbit of lower energy, emitting the energy difference as a light quantum. Or an atom might lose two electrons, becoming *doubly ionized;* one of its remaining outer electrons can then be excited to an outer orbit and upon dropping back can emit a photon.

Since esu are used in Eq. (49.4), the energy E of the electron in any of its orbits in the hydrogen atom is related to the energy V in electron volts by

$$V = \frac{300E}{e} = -\frac{600\pi^2 me^3}{h^2n^2} \tag{49.9}$$

In order to remove the electron from a hydrogen atom an amount of energy equal to V must be supplied to the electron. This energy is therefore the *ionization energy*. The quantity V is a good measure of the tightness of binding of an electron. For hydrogen in the most stable state, $n = 1$,

$$V = 13.58 \text{ ev}$$

If an atom has one or more electrons removed from it, it is said to be *ionized*, or to become a *positive ion*. In this condition the atom has a net charge which is positive and equal to that of the missing electrons. Some atoms can catch and hold an extra electron and become *negative ions*. The simplest negative ion of this type is one of hydrogen, which has two electrons moving about the nucleus.

In Table 49.4 are listed the values of the ionization energies of several atoms. It is evident that the outer electrons of the alkali metals are relatively easy to remove, whereas much more energy is required to remove one electron of the set of

Table 49.4. *Ionization Energies of Some Elements*

Element	Atomic number	Ionization energy
H	1	13.58 ev
Li	3	5.37
Na	11	5.14
K	19	4.32
Rb	37	4.16
Cs	55	3.88
He	2	24.48
Ne	10	21.47
Ar	18	15.69
Kr	36	13.3
Xe	54	11.5
Mg	12	7.61
Ca	20	6.08
Hg	80	10.39

eight that occurs in each of the inert-gas atoms. It is also evident from Table 49.4 that the binding energy of a given series decreases with increasing atomic number. Thus helium requires 24.48 ev, but this value drops to 11.5 ev for the heavy element xenon. This is principally the result of the fact that the outermost electrons are farther away from the nucleus in the heavy elements.

Every element has one line spectrum associated with the neutral atom, another spectrum with the singly ionized atom, still another with the doubly ionized atom, etc. Because of the similarity in electron structures the spectrum of the once-ionized atom resembles that of the neutral atom of atomic number 1 less. For example, since copper has atomic number 29 whereas sodium has atomic number 11, stripping of 18 electrons from a copper atom in a "hot-spark" source should give a highly ionized atom showing a spectrum similar to that of sodium. This particular spectrum, the nineteenth spectrum of copper, has been observed by Edlén in the extreme ultraviolet, the displacement to higher frequencies being a result of the greater charge on the copper nucleus.

49.8. X-ray Spectra. X rays also originate from the jumping of an electron from an orbit of higher energy to one of lower energy, but the electron transitions are to vacancies in the innermost K and L shells rather than to the orbits of an outer valence electron. When high-speed electrons from the cathode of an X-ray tube strike the target, which is usually made of a high-melting-point metal such as tungsten or molybdenum, they ionize many of the atoms in the surface layers of the metal. The very energetic electrons from the cathode penetrate these atoms and by collision remove an electron from one of their inner shells. If this ejected electron is a K electron, an electron from the next, or L, shell jumps into this vacant place, simultaneously emitting the difference in energy between these two orbits as a photon hf. X rays originating by this process are called "characteristic X rays" since their frequencies are characteristic of the target element, and independent of the energy of the bombarding electron.

Such X rays may be photographed as "lines" with an X-ray spectrograph that employs a crystal as a diffraction grating (Sec. 47.11). These K lines are actually multiple, with the components designated as $K_{\alpha 1}$, $K_{\alpha 2}$, $K_{\beta 1}$, $K_{\beta 2}$, because electron transitions to the vacancy in the K shell can also occur from the M and N shells. A plot of the K lines from titanium ($Z = 22$) to zinc ($Z = 30$) is given in Fig. 49.7. The study of these strong K lines led Moseley in 1913 to announce the law that *the frequencies of corresponding lines are approximately proportional to* Z^2, where Z is the atomic number of the emitting element. Moseley's formula for the frequency f of any $K\alpha$ line is

$$f = R(Z - 1)^2 \left(\frac{1}{1^2} - \frac{1}{2^2}\right)$$

The regular stepwise progression of these X-ray lines in this plot constituted the first certain proof of the regular variation of properties of the elements with atomic number.

There are also L-series X-ray lines due to the jumps of M- and N-shell electrons into the L-shell vacancies, and there are even M and N series arising from electrons falling into the vacancies in those shells. Since the electrons in these outer shells are not so tightly bound to the nucleus, less energetic cathode-ray electrons will eject them from the atom and the photons emitted are of lower frequency or longer wavelength. As explained in Sec. 36.1, in addition to these lines there exists a continuous X-ray spectrum arising from the slowing down of high-speed electrons as they pass close to the nuclei of the atoms in the target of the X-ray tube.

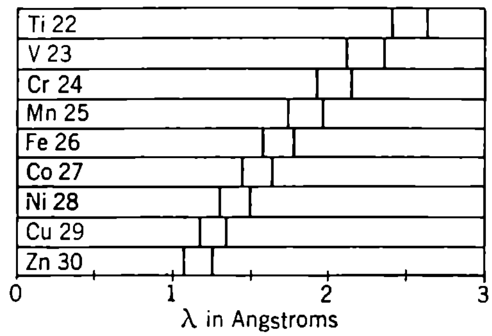


FIG. 49.7. Wavelength of K-series X-ray lines (*after Moseley*).

49.9. Molecular (Band) Spectra. In molecules the electrons are also confined to shells and orbits. The innermost electrons are bound to one atom alone, while the outer electrons circulate about two or more atoms. When an electric discharge is sent through a gas composed of diatomic molecules, the outer electrons of the molecules are excited to larger orbits and also the molecules are made to vibrate and rotate faster. If there are polyatomic molecules in the gas, spectroscopic evidence indicates that the discharge creates diatomic fragments which are largely responsible for the observed molecular spectra. Thus a discharge through the vapor of any organic compound produces CH fragments which show a characteristic spectrum.

Figure 49.8 shows the appearance of a portion of the emission spectrum of the AlO molecule. Such spectra are often referred to as "band" spectra, for they usually consist of a sharp edge with many fine lines trailing off from that point. The fact that these molecular spectra also consist of sharp lines shows the necessity of quantizing all the motions in the molecule. That is, a molecule can vibrate or rotate not with any frequency but only with certain definite frequencies. A band represents a jump between two of these allowed vibrational energy states, the many fine lines resulting from simultaneous jumps between some of the molecular rotational energy states. A set of these bands, such as those shown in Fig. 49.8, corresponds to a certain electronic transition that may give rise to but one spectral line from an atom.

The application of the quantum theory brings order and interpretation to these complicated molecular spectra. The assignments of vibrational and rotational quantum numbers can now be made with certainty. Useful information on the sizes of the molecules, the nature of their energy states, how stable the molecules are, and the exact molecular fragment responsible for each spectrum results from these studies. Chemical analyses of gases may be made in this way, and these data aid in the perfection of all devices involving gaseous discharges (thyratrons, fluorescent lamps, high-pressure mercury-vapor lamps, etc.).

***49.10. Fluorescence.** There are substances that, when strongly illuminated by light they are capable of absorbing, reemit light of longer wavelength. This process is called *fluorescence*, and the fact that the fluorescent light is almost always of longer wavelength than that of the incident light is known as *Stokes' law*. This is to be expected since the energy *hf* of the photons of the fluorescent light may at the most be equal to the energy *hf* of the incident photons. There are cases where this law is violated, however, because energy stored in the atoms and molecules of the material is added to the energy of the incident waves.

Certain solids such as the sulfides of calcium, strontium, zinc, and barium show a persistence of the reemitted light so that it lasts for several seconds or even minutes after the exciting radiation or electron beam has been stopped. This is called *phosphorescence*. There is no essential difference between fluorescence and phosphorescence, except that the latter persists after the excitation has ceased, whereas in the former the light is reemitted within 10^{-7} or 10^{-8} sec.

A solid bombarded by charged particles also emits light and is also called fluorescent or phosphorescent. The emission process is the same, but the excitation of the atoms of the solid results from the electric field of the charged particle rather than from the electric field of the light wave.

Fluorescent paints are now widely used for highway markers, advertising signs, etc. The fluorescent screens of cathode-ray oscilloscopes and television tubes must not phosphoresce to any great extent, or the persistence of the images would interfere with any fast succession of screen patterns. As the result of war research, *phosphors* have been developed, which when irradiated with infrared light emit visible wavelengths—a striking violation of Stokes' law. Interesting fluorescent effects may be obtained by irradiating

Fig. 49.8. Emission band spectrum of a diatomic aluminum oxide molecule.

materials in the dark with ultraviolet light. For example, natural teeth then fluoresce brightly, but artificial teeth look perfectly black.

Another important application is the use of fluorescent screens to make visible with X rays the shadows of bones.

***49.11. Electron and Nuclear Spin.** If the Balmer series of hydrogen or the principal series of sodium is photographed with high dispersion and resolving power, each of the lines is found to be double. This *fine-structure* doublet interval amounts to 1.59A for the H_α line and to 6A for the yellow sodium line. The reason for this doubling is that the electron has a *spin* angular momentum $S = h/4\pi$. Now there is a magnetic moment associated with both the orbital motion of the electron and its spin motion (Sec. 31.4), and the only allowed relative orientations of two magnetic moments are parallel or "antiparallel" to each other. If as the electron jumps from one orbit to the other the orbital and spin angular momenta are parallel, the frequency of the emitted light quantum will be slightly different from that emitted by another atom with the two momenta antiparallel.

When each line of the sodium doublet is photographed under still higher resolution, it is found to be double. This so-called *hyperfine structure* results from the spinning of the nucleus of the atom. The orientation of this *nuclear spin* affects the frequency of the emitted light in very much the same way as does the orientation of the electron spin, but usually the nuclear spin produces the closer line structure.

Spectrum lines of many of the elements show electron-spin fine structure and nuclear-spin hyperfine structure to varying degrees. These fine structures often involve more than two components. Also, there is sometimes a closely spaced isotope structure in the lines if the nucleus of the atom has two or more isotopic forms. The study of these effects forms a fascinating branch of atomic physics, but it is beyond the scope of this text.

49.12. Relativistic Increase of Mass with Velocity. We have already mentioned (Sec. 29.5) that as an electron acquires high speed its mass changes. In the special theory of relativity published by Einstein in 1905, the mass m of an electron increases with its velocity v according to the relation

$$m = \frac{m_0}{\sqrt{1 - (v/c)^2}} \tag{49.10}$$

where m_0 is the mass of the electron at rest and c is the velocity of light. This equation has been thoroughly verified for particles of any mass.

The test of Eq. (49.10) was first made upon high-speed β particles or electrons (Sec. 50.3) ejected from the nuclei of naturally radioactive atoms. Some β particles are so energetic that they are moving with a velocity of 99.95 per cent that of light. Their velocity and charge-to-mass ratio may be determined by experiments on their deflection by electric and magnetic fields, similar to the method used originally by J. J. Thomson

(Sec. 29.3) with cathode rays. Since the β particles are identical with any other electrons, their rest mass is known. Such experiments yield perfect agreement with the relativity equation [Eq. (49.10)]. If $v = c/2$, the m is $1.15m_0$, while if $v = 0.99 \times c$ the ratio m/m_0 is 7.1. The formula for the kinetic energy of a fast-moving particle is no longer $E_{kin} = \frac{1}{2}mv^2$; it is derived in Sec. 50.2.

49.13. Electron Waves. As mentioned in Sec. 47.11 Davisson and Germer in 1925 performed an important experiment showing that moving electrons are wavelike in character. This result had already been predicted by the French theoretical physicist De Broglie. According to De Broglie a particle of mass m moving with a velocity v should have an associated wavelength λ given by

$$\lambda = \frac{h}{mv} \tag{49.11}$$

where h is Planck's constant. Thus a beam of electrons or atoms should behave in some experiments, such as diffraction by a crystal grating (Sec. 47.11), like a beam of photons or X rays.

According to Eq. (49.11), the greater the momentum mv of the particle, the shorter the wavelength λ. An electron with 1 ev of energy has a velocity of 0.02 times the velocity of light and has by Eq. (49.11) a λ of 12.23A. If an electron is accelerated by falling through a potential difference of 1 million volts, its velocity is 94 per cent that of light (the relativistic correction for the mass must be used), and from Eq. (49.11) its λ is 0.01A. The λ of the 1-ev electron is that of a "soft" X ray, that of the 1 Mev electron is equivalent to a "hard," or penetrating, γ ray.

49.14. Quantum Mechanics. Uncertainty Principle. Although the Bohr theory of the electron circulating about the atomic nucleus in fixed orbits (Sec. 49.2) worked perfectly for hydrogen, it was not so successful in explaining the details of the spectra of heavier atoms. In 1925, soon after De Broglie had suggested that the electron has a wavelike character, Schrödinger brought out his more acceptable *quantum mechanics*, while Heisenberg at the same time, followed by Dirac, also published an improved theory in a different mathematical form. In the quantum-mechanical treatment of the hydrogen atom the motion of the electron about the nucleus is described in terms of standing waves somewhat like the standing waves in a vibrating body emitting sound.

In the simplest form of this theory the length of the path $2\pi r$ of the electron moving about the proton must be an integral multiple of the wavelength $\lambda = h/mv$. That is,

$$n\frac{h}{mv} = 2\pi r \tag{49.12}$$

where $n = 1, 2, 3, \ldots$. This is exactly the Bohr relation [Eq. (11.23)], and therefore it is clear that this theory of Schrödinger's should also give correctly the frequencies of the hydrogen spectrum.

In the quantum mechanics it is more difficult to form a mental picture of what an atom is like. The modern theoretical physicist does not demand such a picture; rather, he works with the somewhat diffused density distributions of electron charge in the atom. Figure 49.6 is an example of such a charge density for the potassium atom. It is noteworthy that the peaks labeled K, L, M come at just about the same distances from the nucleus as do the shells of electrons calculated on the orbital model. Physicists find it convenient, therefore, to continue to speak of the motion of electrons in orbits, and of electron shells in discussing spectra and the applications of spectroscopic data to other atomic phenomena.

One of the spectroscopic facts that is explained by quantum mechanics but not by the older theory is that the minimum vibrational energy for a molecule is not zero. This residual energy, which cannot be subtracted from a molecule and would be present at the absolute zero of temperature, is called *zero-point energy*. In a complex polyatomic molecule in which each of the many modes of vibration retains half a quantum of vibrational energy the total zero-point energy may be a large fraction of an electron volt.

The famous *Heisenberg uncertainty principle* is closely related to the wave theory. In order to measure λ for a particle accurately its associated waves must be extended to a train of great length. Then, according to wave mechanics, the position at which the particle will be found is very uncertain since the wave extends over this large region. Conversely, to determine the position of a particle accurately the wave must be limited to a small region. By Fourier's theorem this can be done, but only by the overlapping of many waves of different wavelengths; *i.e.*, the exact wavelength of the particle is then uncertain. We thus have the following as the *uncertainty principle: The position and momentum of a particle cannot both be measured simultaneously with exactness.* According to Heisenberg's formulation of this principle the product of the uncertainties of the measurement of the position and momentum of the particle cannot be of a smaller order of magnitude than Planck's constant h. This is one of the basic laws of the physical world and is of great philosophical importance.

49.15. Spectra of Astronomical Sources.[1] The solar spectrum has already been mentioned in Sec. 45.1. The spectrum of the moon is

[1] Review Sec. 45.1.

identical with that of the sun, and this constitutes proof that the moon has no gaseous atmosphere to absorb any frequencies from the sunlight reflected by its surface. Some of the planets, on the other hand, produce absorption bands in the sunlight they reflect. Carbon dioxide has been thus shown to occur in the atmosphere of Venus and methane in the atmosphere of Mars, and from the intensities of the absorption bands the amounts of these compounds may be estimated.

The spectra of stars show the presence of elements found on the earth. Prominent are hydrogen, calcium, titanium, and iron. From the study of stellar spectra, astronomers may determine the distances, brightness, sizes, surface temperature, radial velocity, and mass of the stars, as well as their composition. The type of absorption spectrum always reveals the temperature of the absorbing material. For example, the spectrum of *Arcturus* shows absorption lines of iron, calcium, and hydrogen in such relative intensities as to indicate that the surface temperature of that very bright star is about 4200°.

If the relative velocity v of the stellar source and the earth along the line of sight, called the *radial velocity*, is appreciable in comparison with the velocity of light c, the frequency f of every spectral line is changed by the Doppler effect. Using this notation, Eqs. (38.1) and (38.1′) may be combined into one equation for the new frequency f'.

$$f' = \frac{c}{\lambda'} = \frac{c}{\lambda\left(\dfrac{c \pm v}{c}\right)} = \frac{f}{1 \pm \dfrac{v}{c}} \quad \text{(source in motion)} \qquad (49.13)$$

Similarly Eqs. (38.2) and (38.2′) may be combined into the one equation

$$f' = f\left(1 \pm \frac{v}{c}\right) \quad \text{(observer in motion)} \qquad (49.14)$$

For sound waves there is a definite physical difference bet[illegible] these two cases. When the source is in motion, there is a real chang[illegible] the wavelength λ, while if the observer is in motion λ is unchanged but there is an apparent change in frequency. Therefore, when v is not small compared with the velocity of sound, the two equations give different results for the same v. If the right-hand side of Eq. (49.13) is expanded by the binomial theorem, we have

$$f' = f\left(1 \mp \frac{v}{c} + \frac{v^2}{c^2} \mp \frac{v^3}{c^3} + \cdots\right)$$

which is the same as Eq. (49.14) except for the small terms v^2/c^2, etc. Owing to the extremely high velocity of light this difference is negligible.

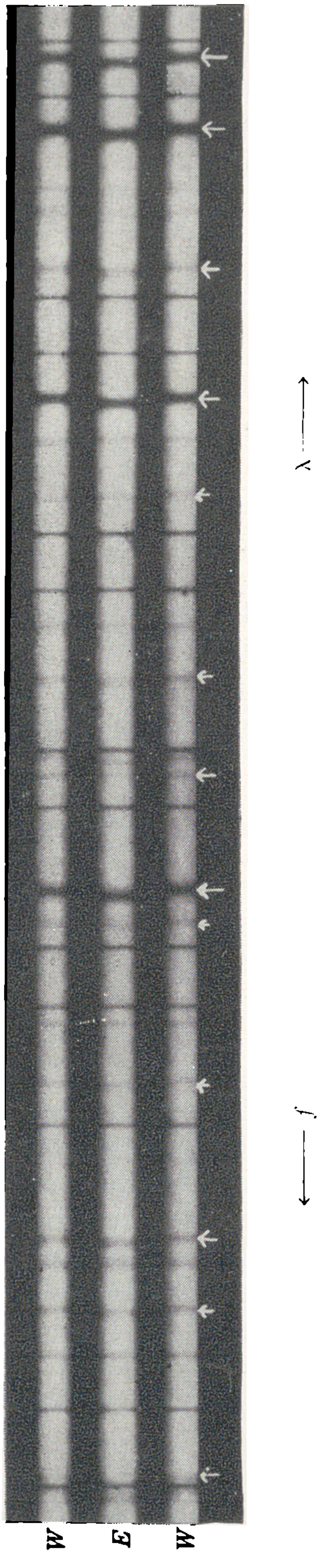


FIG. 49.9. Portion of solar spectrum in the red showing the Doppler effect arising from the sun's rotation. The spectra marked *W* are of light from the receding limb of the sun; that marked *E* is from the approaching limb. Those lines with no Doppler effect belong to a band spectrum from O_2 absorption in the earth's atmosphere. (*Courtesy of Mount Wilson Observatory.*)

Stars have radial velocities usually in the range 10 to 30 km/sec, but in some cases up to 300 km/sec. Even for the latter value, v^2/c^2 is only 1 millionth. Hence, for star light we always use Eq. (49.14).

When the spectrum of the west edge of the sun is compared with that of the east edge, it is found that all the lines due to absorption in the sun's atmosphere show a displacement corresponding to a velocity difference of 2.1 km/sec (Fig. 49.9). Comparison of spectra from the north and south edges of the sun shows no such line shifts. This shows that the sun rotates about a "north-south" axis. For certain double stars, known as *spectroscopic binaries*, the spectrum on one night may show all the lines as single, a few nights later these lines are double, and still later by the same time interval the lines are again single. This shows that the star is double, with the velocities of the rotating doublet having appreciable components in the line of sight.

A most important application of the Doppler effect is to measure the apparent motion of the distant spiral nebulae (the nearest one is 870,000 light years away). The spectra of these island universes outside our own galaxy (the Milky Way) always show the H and K lines of ionized calcium. But instead of falling at the wavelengths 3,968A and 3,933A as in laboratory sources, they show tremendous Doppler shifts toward the red, and *the more distant the nebula the greater the shift*. Recessional velocities as great as 14,300 miles/sec have been so determined. This startling result has given rise to the belief that our universe is expanding.

PROBLEMS

1. Calculate from the densities and atomic weights the volumes and linear dimensions of the atoms of the alkali metals. Review Sec. 20.8.

2. Calculate the wavelength in Angstrom units of the fifth line of the Balmer series of hydrogen.

3. Calculate the energy hf in ergs of a photon of wavelength (*a*) 6,000A and (*b*) 1A.

4. From the limit of the Balmer series (3,646A) calculate the minimum energy required to remove the electron in the H atom from the $n = 2$ orbit to infinity. Express your result in cm^{-1} and in electron volts.

★5. The ionization potential of the hydrogen atom is 13.58 volts. Compute the wavelength of the limit of the Lyman series in the vacuum ultraviolet.

6. Compute the wavelengths of the first lines of the Paschen, Brackett, and Pfund series of the hydrogen atom.

7. Calculate the wavelength separation between the H_β and D_β Balmer series lines.

★8. The limit of the principal absorption series of potassium lies at $\lambda = 2{,}858A$. Compute the ionization potential of potassium.

★9. What form would Eq. (49.9) take in practical mks units?

10. What is the energy in electron volts necessary to remove the single remaining electron from a helium ion?

11. If a positive and a negative electron form a rotating system similar to a hydrogen atom, what will be the energies of the system? (Remember the comments about reduced mass.)

12. A betatron produces electrons with a velocity 99.8 per cent that of light. Calculate the mass of these electrons.

13. Calculate the wave numbers and wavelengths of the $K\alpha$ lines for the lightest metal lithium ($Z = 3$) and the heaviest natural element uranium ($Z = 92$).

14. What should be the short-wavelength limit of the continuous X-ray spectrum produced by 30,000-volt electrons impinging on a target?

15. Compute the wavelengths observed for the H and K lines in the spectrum of a spiral nebula, which has a recessional radial velocity of 14,300 miles/sec.

CHAPTER 50
NUCLEAR PHYSICS

50.1. The Size of the Nucleus. In Chap. 49 the structure of the atom was discussed, and the binding forces holding the electrons in place were described. The nucleus of the atom, which contains most of the material, was seen to be only a small massive center about which the electrons move. The Coulomb force, arising from the charges on the nucleus and on the outer electrons, retains the electrons within the atom, and the magnitude of this force is independent of the shape and size of the nucleus because the nucleus is extremely small. An estimate of an upper limit of the size of the nucleus can be obtained from the experiments on the scattering of α particles described in Chap. 24.

If α particles bombard a material with heavy nuclei, such as a piece of gold foil, some of the α particles make "head-on" collisions with the nuclei and are scattered directly backward. As an α particle approaches the nucleus, its kinetic energy is transformed into potential energy in the electric field of the nuclear charge. The velocity of the α particle is reduced to zero, and the repulsive Coulomb force is a maximum when the α particle has approached the nucleus most closely. This distance is easily calculated. If V is the energy of the α particle in electron volts, then

$$\frac{1}{2} mv^2 = eV = \frac{2kZe^2}{r_0} \tag{50.1}$$

where r_0 is the minimum distance, Z the atomic number of the heavy material, and k the factor depending on the units employed. The quantity Ze is the nuclear charge, and $2e$ is the charge of the α particle. In esu, $k = 1$; in practical units, $k = 1/4\pi\epsilon_0$. For a 5-million-volt α particle approaching a gold nucleus for which $Z = 79$, the distance r_0 is

$$r_0 = \frac{2kZe}{V} = \frac{2 \times 79 \times 1.6 \times 10^{-19}}{4\pi \times 8.85 \times 10^{-12} \times 5 \times 10^6} = 4.56 \times 10^{-14} \text{ m}$$
$$= 4.56 \times 10^{-12} \text{ cm}$$

Experiment shows, however, that the scattering observed is exactly that computed from the assumption that the field is the same as the field of a point charge. It can be concluded, therefore, that the nuclear radius must be still smaller than the value of r_0 just calculated. If the scattering is observed from a light element instead of gold, the distance r_0 of closest approach is smaller. Equation (50.1), however, cannot be used in that case since some of the kinetic energy of the α particle is communicated to the scattering nucleus. Experiments with light elements show that the scattering is not quite what would be expected from Coulomb forces even when the motion of the target nucleus is taken into account. Although the deviations are small, they are definite and can be used to define what is meant by the radius of the nucleus. If the potential energy of an α particle at a distance r from the nucleus is plotted, a curve such as that shown in Fig. 50.1 results. For large values of r the curve falls off inversely with r; at small distances an *attractive* force exists. The radius R corresponding to the maximum of the curve may be defined as the nuclear radius. The curve of Fig. 50.1 represents equally well the potential energy of any positively charged particle in the field of the nucleus; only a change in the scale of ordinates is necessary. A proton has just half the potential energy of an α particle, since the charge is just one-half.

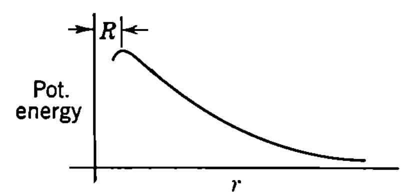


FIG. 50.1. The potential energy of a charged particle near a nucleus. Curve near origin unknown.

The nuclear radius can be determined more precisely and easily by other means to be discussed subsequently. It is found that the empirical equation

$$R = 1.4 \times 10^{-13} A^{1/3} \text{ cm} \tag{50.2}$$

represents the experimental data well. The quantity A is the atomic weight of the scattering nucleus. From this equation, the radius of a gold nucleus is found to be 8.1×10^{-13} cm; for a nucleus of carbon, $R = 3.2 \times 10^{-13}$ cm. It should be recalled that a linear dimension of an atom is about 10^5 times larger than R. For most purposes the nucleus can be approximated extremely well by a point charge. On the other hand the small size of the nucleus makes the investigation of its properties difficult.

50.2. The Mass-Energy Relation. In dealing with nuclear phenomena, one is nearly always concerned with particles that have very high energies. Although the energy of the 5-Mev α particle considered in the preceding section is only a small fraction of 1 erg, this energy is very much larger than the energies of the particles concerned in chemical or mechanical phenomena. If the energy of the particle in question is kinetic energy, then the velocity of the particle is correspondingly great. Particles with high velocities do not have a constant mass, as has already been mentioned, but the mass m increases with increasing velocity v according to the relation

$$m = m_0 \frac{1}{\sqrt{1 - (v/c)^2}} \tag{50.3}$$

where m_0 is the mass of the particle at rest, its *rest mass*, and c is the velocity of light. Equation (50.3) was first derived by H. A. Lorentz (1853–1928), a Dutch physicist, who constructed a theory of the electron. It was subsequently obtained by A. Einstein from his theory of relativity. The relation has received abundant experimental verification. If mass varies with velocity, the statement and application of Newton's second law of motion must be carefully made. We must say that the *rate of change of momentum* (not mass times acceleration) is equal to the force. In symbols,

$$F = \frac{d}{dt}(mv) \neq m\frac{dv}{dt} \tag{50.4}$$

We have already encountered a similar situation in the study of rockets in Sec. 8.14.

The kinetic energy of a particle of varying mass is not equal simply to $\frac{1}{2}mv^2$ but is equal to a more complex expression about to be derived. The kinetic energy of a particle is altered by the action of a force, and by definition the change dE in kinetic energy is equal to the work dW done by the force. Since

$$dE = dW = F\,dx$$

by Eq. (50.4) we have

$$dE = \frac{d}{dt}(mv)dx \tag{50.5}$$

If the value of m is substituted from Eq. (50.3), the change in kinetic energy becomes

$$dE = \frac{dx}{dt}\,d(mv) = vd\left[\frac{m_0 v}{\sqrt{1 - (v^2/c^2)}}\right]$$

Taking the differential we find

$$dE = \frac{m_0 v\,dv}{[1 - (v^2/c^2)]^{3/2}}$$

which can be easily verified to be equivalent to $c^2\,d(m_0/\sqrt{1 - (v^2/c^2)})$, and hence

$$dE = d(mc^2) \tag{50.6}$$

The change in kinetic energy is therefore equal to the change in the quantity mc^2. To find the total kinetic energy E of a particle with velocity v, Eq. (50.6) must be integrated, and the result is

$$E = mc^2 + C \tag{50.7}$$

where C is the constant of integration. To determine C it is necessary to know the value of E at some value of v. In the dynamics of low-velocity particles it is customary to choose $E = 0$ when $v = 0$. Under these conditions, $C = -m_0c^2$, and

$$E = mc^2 - m_0c^2 \tag{50.8}$$

The student should show that this reduces to $\frac{1}{2}\,mv^2$ when $v \ll c$.

In proposing his theory Einstein took the bold step of interpreting Eq. (50.8) somewhat differently. He postulated that, even when the velocity v of the particle is zero, some energy still remains in the particle and in fact that the total energy E_{tot} of the particle is obtained when $C = 0$ in Eq. (50.7), or

$$E_{\text{tot}} = mc^2 \tag{50.9}$$

From Einstein's point of view *mass thus becomes a form of energy.* If in some way the whole mass of a particle could be changed to some other form of energy by annihilation of the matter, Eq. (50.9) would give the amount of energy obtained. In terms of the kinetic energy E,

$$E_{\text{tot}} = E + m_0c^2$$

The quantity m_0c^2 is called the *rest energy* of the particle.

Einstein's hypothesis that energy is obtained by the annihilation of matter has been accurately verified by the study of nuclear processes, as we shall see. The complete annihilation of 1 kg of matter would produce 9×10^{16} joules, or 2.5×10^{10} kwhr, about one-tenth of the total electric energy produced a year in the United States. Complete annihilation of nuclear mass has not been accomplished, but changes in mass with the accompanying release of energy can be produced almost at will by nuclear processes.

50.3. Radioactivity. The first nuclear phenomenon to be studied in detail was the *radioactivity* of nuclei discovered in 1896 by Becquerel. Radioactivity is the spontaneous disintegration of a nucleus, with the emission of energy. The nuclei of a number of elements of high atomic number are observed to transform themselves in two principal ways: (1) A nucleus emits an α *particle* or a helium nucleus, and kinetic energy is released and divided between the α particle and the residual nucleus; (2) A negative electron of high energy is produced by the nucleus; *i.e.*, a β *ray* is emitted. In either mode of disintegration, some radiant energy may also be released, and this radiation is called γ *rays.* When a nucleus disintegrates, the atomic number of the residual, or daughter, nucleus must differ from that of the parent nucleus. If an α ray is shot out, the nuclear charge must diminish by $2e$ since the helium nucleus has $Z = 2$. At the same time the mass number must also diminish by 4, the mass number of helium. For example, the element polonium ($Z = 84$ and $M = 210$) emits α rays and changes to the element with $Z = 82$ and $M = 206$, or to an isotope of lead. This transmutation is often expressed by an equation similar to those used to describe chemical reactions, thus:

$$_{84}\text{Po}^{210} \longrightarrow {}_2\text{He}^4 + {}_{82}\text{Pb}^{206} \tag{50.10}$$

The atomic number of each element is written as a subscript on the left, the mass number as a superscript on the right. It is evident that the totals of both Z and M must have the same values on each side of the equation. If R is used as a general symbol for elements, then α disintegration is expressed by

$$_ZR^M \longrightarrow {}_2\text{He}^4 + {}_{Z-2}R^{M-4} \tag{50.11}$$

In a similar fashion an equation can be written for β decay. The parent nucleus loses a negative charge, and hence Z is increased by 1; the mass number remains unchanged. The mass number of an electron is zero since its mass is only $1/1840$ that of a hydrogen atom. Hence the equation is

$$_ZR^M \longrightarrow {}_{-1}\beta^0 + {}_{Z+1}R^M \tag{50.12}$$

where β is used to indicate the electron. The sub- and superscripts on β are usually omitted.

Radioactive decay is entirely a spontaneous process. The rate of decay is independent of all changes of temperature, pressure, electric fields, or any other factors that can be altered in the laboratory. The number dN of atoms that decay within a small interval dt is thus proportional to the number N of atoms present, and the constant of proportionality is characteristic of the parent atom. Hence

$$dN = -\lambda N\,dt \tag{50.13}$$

where the negative sign means that the number of atoms decreases with increasing t and λ is the constant of proportionality called the *decay constant*. Equation (50.13) can be integrated with the result

$$N = N_0e^{-\lambda t} \tag{50.14}$$

where N_0 is the number of parent nuclei present at $t = 0$. For many purposes it is convenient to write Eq. (50.14) in terms of another constant instead of λ. The time T when the number of atoms is reduced to half its initial value is given by

$$\frac{1}{2} = \frac{N}{N_0} = e^{-\lambda T}$$

If logarithms are taken of both sides,

$$\frac{\ln 2}{\lambda} = T$$

In terms of T, Eq. (50.14) becomes

$$N = N_0 e^{-(\ln 2\, t/T)} = N_0 e^{-\ln 2^{-t/T}} = N_0 2^{-t/T} \tag{50.15}$$

The number N is thus decreased by a factor of 2 if the disintegration proceeds for an interval T; hence T is called the *half-life* of the parent nucleus.

50.4. The Radioactive Series. The radioactive elements that occur in nature can be arranged in three radioactive series, each series being a chain of successive disintegrations. The series are named after the elements in the series with the longest lives. The elements are uranium ($_{92}U^{238}$), thorium ($_{90}Th^{232}$), and actinium ($_{89}Ac^{227}$). In Table 50.1 the members of the uranium and thorium series are given.

The individual members of the series are given names, although they may be chemically identical with elements having more common names. Thus MsThII is an isotope of Ac, and ThC is an isotope of Bi. It is seen that the half-lives extend over an enormous range of values.

The actinium series also starts with uranium and ends with lead. This circumstance makes possible a determination of the age of a mineral. If the amounts of uranium, thorium, and lead in a mineral are determined, it is possible to calculate from the known half-lives how long the disintegration process has proceeded since the mineral was formed. In this manner a time scale for the various geological periods has been established. The oldest rocks have an age of about 1.5×10^9 years.

It should be noted from Table 50.1 that RaC and ThC emit both α and β rays. The disintegrations take place in two possible ways, and a *branch* in each series is formed as indicated.

The atomic weights of members of a single radioactive series differ from one another by multiples of 4 since, when an α ray is emitted, the atomic weight decreases by 4, and the emission of an electron involves no change in atomic weight. The thorium series is composed of nuclei having atomic weights of the form $4n$ where n is an integer. Similarly the uranium and actinium series have mass numbers of the form $4n + 2$ and $4n + 3$, respectively. Recently a fourth radioactive series with atomic

Table 50.1. *Radioactive Series*

Element	Symbol	Z	M	Ray produced	Half-life
		Uranium Series			
Uranium I	UI	92	238	α	2×10^{9} years
Uranium X_1	UX_1	90	234	β	24.5 days
Uranium X_2	UX_2	91	234	β	1.14 min
Uranium II	UII	92	234	α	3×10^{5} years
Ionium	Io	90	230	α	8.3×10^{4} years
Radium	Ra	88	226	α	1,600 years
Radon	Rn	86	222	α	3.82 days
Radium A	RaA	84	218	α	3.05 min
Radium C	RaB	82	214	β	26.8 min
Radium C	RaC	83	214	α, β	19.7 min
Radium C'	RaC'	84	214	α	10^{-6} sec
Radium C''	RaC''	81	210	β	1.32 min
Radium D	RaD	82	210	β	22 years
Radium E	RaE	83	210	β	5 days
Polonium	Po	84	210	α	140 days
Lead	Pb	82	206	Stable	Very long
		Thorium Series			
Thorium	Th	90	232	α	1.8×10^{10} years
Mesothorium I	MsThI	88	228	β	6.7 years
Mesothorium II	MsThII	89	228	β	6.1 hr
Radiothorium	RaTh	90	228	α	1.90 years
Thorium X	ThX	88	224	α	3.64 days
Thoron	Tn	86	220	α	55 sec
Thorium A	ThA	84	216	α	0.14 sec
Thorium B	ThB	82	212	β	10.6 hr
Thorium C	ThC	83	212	α, β	61 min
Thorium C'	ThC'	84	212	α	3×10^{-7} sec
Thorium C''	ThC''	84	208	β	3.2 min
Lead	Pb	82	208	Stable	Very long

weights of the form $4n + 1$ has been discovered. None of the members of this series has been found to occur naturally, but they can be manufactured in the laboratory. The series has been called the *neptunium* series after the longest lived member $_{93}Np^{237}$, which has a half-life of 2.2×10^6 years.

***50.5. Ionization by α and β Rays.** When a charged particle moves through a gas with a high velocity, it produces ionization by collision as discussed in Sec. 36.4. In this way the kinetic energy of the charged particle is dissipated, and the particle itself can be detected by the ionization produced. Energy is also transferred without ionization to the atoms of the material through which the particle moves, and on the average, in air, a particle loses about 32 ev for each ion pair it forms. A heavy particle, such as an α ray, is not deflected when it collides with an atom unless it comes near the nucleus. An α particle traversing matter thus travels in a straight line until its energy is gone. The distance that the particle travels, or the *range* of the particle, depends on its initial energy. From measurements of the range the energy can be determined, and this technique is often employed for nuclear particles. The relation between the range in air and the energy is shown in Fig. 50.2 for fast-moving α particles and for protons. The range of an α particle of 5 Mev is seen to be 3.5 cm, whereas the range of a proton of the same energy is 10 times as large.

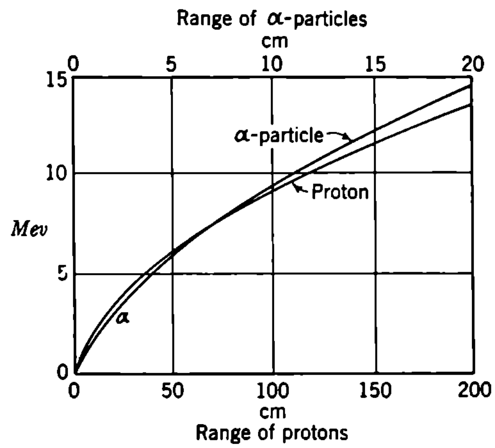
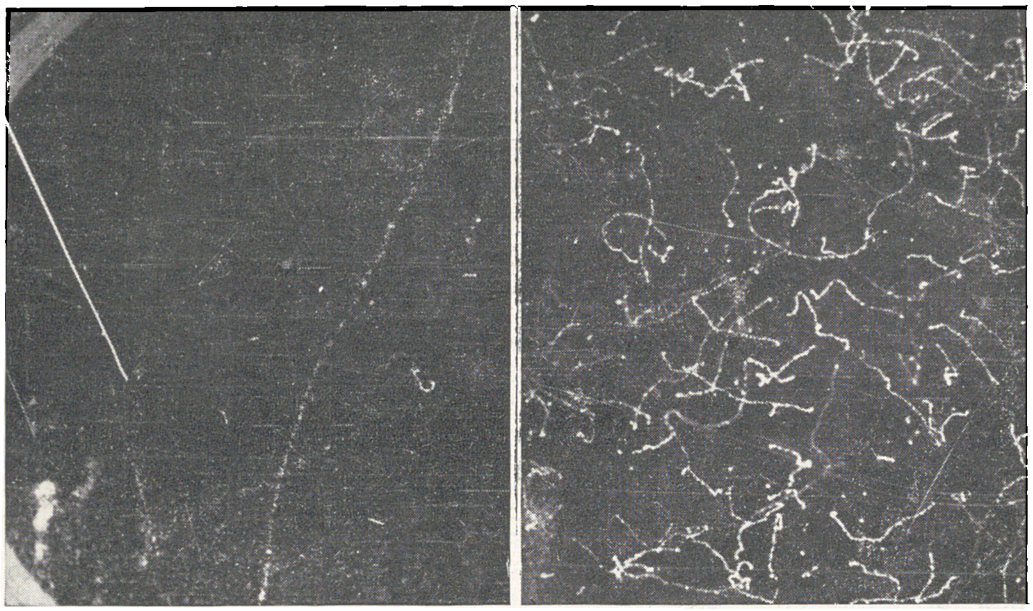


FIG. 50.2. Range vs. energy for proton and α-particles in air at normal density.

The distance that a β ray travels before it loses its kinetic energy is not a definite quantity. The momentum of a β ray is relatively small, and it experiences large deflections in its path in nearly every collision with an atom. The path of an electron is thus a tortuous one, and its length is impossible to determine. On the average a β particle produces only about 0.01 as many ion pairs per centimeter of path as are produced by an α ray. The range is consequently of the order of 100 times larger than the range of an α particle, or 10 times the range of a proton.

C. T. R. Wilson devised an instrument, called a *cloud chamber*, whereby the ionization produced by α and β rays can be made visible. If an enclosure contains a gas saturated with water vapor and the gas is suddenly expanded, the gas becomes supersaturated and the water vapor condenses to form a cloud. If the amount of expansion is properly adjusted, drops of water form only on the ions that are present. If an α particle traverses the chamber just before the expansion, the condensing water forms a dense line of drops along the path and the track is made visible and can be photographed. Beta rays form tracks with fewer drops per unit path length, and the number of ion pairs formed can be found by counting these drops. Figure 50.3*a* shows the tracks obtained with a cloud chamber of an α and a β particle.

The tracks of α-particles can also be made visible by allowing the particles to penetrate a photographic emulsion. The ionization produced renders the silver bromide grains along the path developable.



(*a*) (*b*)

FIG. 50.3. (*a*) Tracks of an alpha particle (*left*) and a fast electron (*right*) in a Wilson cloud chamber. The faint sharp line to the extreme left is a reflection of the α-particle track in the glass cover of the chamber. (*b*) Tracks produced when helium gas is irradiated by X rays. The long tortuous tracks are produced by photoelectrons ejected from the gas; the small blobs are very short tracks from Compton electrons. (*Courtesy of W. E. Hazen.*)

***50.6. Ionization Produced by γ Rays.** The γ rays that are sometimes formed in radioactive disintegrations also produce ionization, but only sparsely. The number of ion pairs formed per centimeter of path is only about $\frac{1}{100}$ the number formed by a β ray or only 1/10,000 the number that an α ray produces. Gamma rays are consequently more penetrating than α rays by a factor of roughly 10,000. The ionization is not produced directly; the γ ray must first produce an electron, and then the electron ionizes. One mechanism by which an electron is produced is the photoelectric effect, which we have already discussed in Chap. 36. The whole energy of the γ-ray photon disappears, and a single electron is produced.

It is also possible for a γ-ray photon to give up only a part of its energy to an electron. This is the *Compton effect* (A. H. Compton, 1921). The Compton effect is best regarded as a collision between an electron at rest and a high-energy photon. A photon possesses energy of an amount hf, as we know, but experiments show that it also possesses *momentum* of an amount hf/c. In the collision between a photon and an electron, both energy and momentum must be conserved. The electron gains some kinetic energy, which is given by Eq. (50.8), and the photon loses an equal amount. The frequency f of the photon is thus changed to the value of f' given by

$$hf = hf' + mc^2 - m_0c^2 \tag{50.16}$$

The momentum of the photon is also changed in the collision, and conservation of the total momentum of the system leads to the equations

$$\left.\begin{aligned} \frac{hf}{c} &= \frac{hf'}{c}\cos\theta + mv\cos\phi \\ \frac{hf'}{c}\sin\theta &= mv\sin\phi \end{aligned}\right\} \tag{50.17}$$

where the angles θ and ϕ are indicated in Fig. 50.4. The three equations (50.16) and (50.17) can be used to eliminate v and ϕ, for example, and the change in frequency of the photon is obtained as a function of the angle of scattering θ. To do this it must be remembered that m is a function of v as given by Eq. (50.3). The result is

$$\frac{1}{f'} - \frac{1}{f} = \frac{h}{m_0c^2}(1 - \cos\theta) \tag{50.18}$$

In terms of the wavelengths λ and λ' of the photons this equation can be written

$$\lambda' - \lambda = \frac{h}{m_0c}(1 - \cos\theta) \tag{50.19}$$

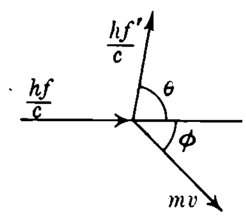


FIG. 50.4. Diagram of the Compton effect.

The change in wavelength of the photon on collision thus depends only on the scattering angle. Substitution of the values for the constants leads to the value $h/m_0c = 2.43 \times 10^{-10}$ cm. If the wavelength λ of the photon before collision is small compared with this length, practically all the energy of the photon is transferred to the electron.

The Compton effect can be demonstrated easily by passing an intense beam of γ rays through a Wilson cloud chamber. Many low-energy electrons whose paths make all angles with the direction of the beam are visible. Fig. 50.3*b* shows the tracks in a cloud chamber of electrons produced by the photoelectric and Compton effects.

50.7. Artificial Transmutation. The transformation of one element into another not only occurs spontaneously in radioactive disintegrations but also can be accomplished by bombarding an element with high-energy particles. The first artificial transmutation was performed in 1919 by Rutherford, who bombarded nitrogen gas with α particles from RaC′ and found that protons were produced. A nuclear reaction took place according to the equation

$$_2He^4 + {}_7N^{14} \rightarrow [_9F^{18}] \rightarrow {}_8O^{17} + {}_1H^1 \tag{50.20}$$

The reaction is shown here to take place in two steps: (1) An unstable "compound" nucleus $_9F^{18}$ is formed. (2) The unstable nucleus disintegrates to form oxygen and a proton. The formation of the compound nucleus is suggested by recent theories of nuclear processes; it disintegrates extremely rapidly, and only the end products are observed.

Following Rutherford's work, many other nuclear reactions have been produced. At first the bombarding particles were always α particles from radioactive disintegrations. Later, protons, α particles, and deuterons accelerated in high-voltage tubes or by cyclotrons were used. All the elements in the periodic table can be disintegrated, and the number and complexity of the reactions observed are very great. Some examples are

$$_1H^1 + {_9F^{19}} \rightarrow [_{10}Ne^{20}] \rightarrow {_8O^{16}} + {_2He^4} \qquad (50.21a)$$

$$_1H^1 + {_5B^{11}} \rightarrow [_6C^{12}] \rightarrow {_6C^{12}} + \gamma \text{ ray} \qquad (50.21b)$$

$$_1H^2 + {_8O^{16}} \rightarrow [_9F^{18}] \rightarrow {_7N^{14}} + {_2He^4} \qquad (50.21c)$$

$$_1H^1 + {_3Li^7} \rightarrow [_4Be^8] \rightarrow {_3Li^7} + {_1H^1} \qquad (50.21d)$$

$$_1H^1 + {_5B^{11}} \rightarrow [_6C^{12}] \rightarrow [_4Be^8] + {_2He^4} \rightarrow {_2He^4} + {_2He^4} + {_2He^4} \qquad (50.21e)$$

In the reaction described by Eq. (50.21a) the compound nucleus $_{10}Ne^{20}$ is formed, which is ordinarily a stable isotope. It possesses, however, too large an energy to remain stable, and it breaks up again. In the reactions given by Eqs. (50.21b) and (50.21e) the same compound nucleus is formed, but it breaks up in two different ways to form stable $_6C^{12}$ or to form three α particles by a two-stage process. The reactions of Eqs. (50.20) and (50.21c) also form the same compound nucleus but proceed in opposite directions.

In all nuclear reactions, just as in all chemical processes, the conservation of energy is maintained. The reaction equation must balance not only for charge and mass number but also for energy. It is necessary, however, to take into account the rest energy m_0c^2 of each particle in the reaction as well as the kinetic energies. To perform such calculations it is necessary to convert frequently from mass, expressed in atomic weight, to energy, expressed in electron volts, by the equation

$$m_0c^2 = eV \qquad (50.22)$$

It is easily found that the *energy of unit atomic weight, or* 1 *mass unit, is equal to* 931 *Mev*. In a nuclear reaction the bombarded nucleus is at rest, and only the bombarding particle has kinetic energy. In the products, however, momentum must be conserved, and all the products have kinetic energy. When only two products of momentum mv and $m'v'$ result, then

$$mv = m'v' \qquad (50.23)$$

if the initial momentum of the bombarding particle is neglected. If the velocities v and v' are small compared with c, then from Eq (50.23)

$$\frac{E}{E'} = \frac{\frac{1}{2}mv^2}{\frac{1}{2}m'v'^2} = \frac{(mv)^2m'}{(m'v')^2m} = \frac{m'}{m} \tag{50.24}$$

where E and E' are the kinetic energies of the two particles.

As an example of energy balance, consider first the simple reaction of Eq. (50.21*b*). The momentum of the γ ray can be neglected, and hence also the kinetic energy of the carbon nucleus. From the values of the atomic masses in Table 29.2 the energy E of the γ ray must be

$$\begin{aligned} E &= 1.0081 + 11.0129 - 12.0040 = 0.0170 \text{ mass unit} \\ &= 15.8 \text{ Mev} \end{aligned}$$

Thus mass energy disappears and is transformed to radiant energy.

A second example is afforded by the reaction discovered by Rutherford [Eq. (50.20)]. The sum of the masses on the left-hand side is 18.0114, and the total mass on the right-hand side is 18.0126. A mass of 0.0012 unit is thus created at the expense of the kinetic energy of the bombarding particle. The α particle,[1] let us say, has a kinetic energy of 7.7 Mev, and the created mass is equivalent to 1.12 Mev. There is thus 6.6 Mev of kinetic energy left over to be distributed between the oxygen nucleus and the proton. From Eq. (50.24), the proton receives $^{17}/_{18}$ times 6.6 Mev, or 6.2 Mev. The range of the proton of this energy can be derived from Fig. 50.2 and is found to be 49 cm of air. The latest experimental determinations of this range yield a value of 48 cm, in excellent agreement with the calculations.

50.8. The Neutron. When an intense beam of α particles is used to bombard a target of beryllium, it is found that rays of great penetrating power are given off. At first it was thought that these rays were penetrating γ rays of the usual sort. If a layer of paraffin or other substance containing hydrogen is placed in the beam of penetrating rays, fast protons in large numbers are ejected. These protons were thought to be produced by a sort of Compton effect, just as fast electrons would be formed by a beam of γ rays. J. Chadwick, a British physicist, suggested the correct explanation. He established in 1932 by careful experiments and calculations that the penetrating rays consist of particles with no charge and a mass approximately equal to that of the proton, and he called these particles *neutrons*. Since neutrons have no charge, they produce no ionization and can penetrate matter easily. If a neutron

[1] Alpha particles coming from RaC′, frequently used in this experiment, happen to have this energy.

collides with a hydrogen nucleus, however, it can transfer its kinetic energy to the proton. If the collision is "head on," all the kinetic energy is given to the proton and the neutron is stopped. The reaction with beryllium is, therefore,

$$_2He^4 + {}_4Be^9 \rightarrow {}_6C^{12} + {}_0n^1 \tag{50.25}$$

when $_0n^1$ is the symbol for a neutron. The kinetic energy of the neutron can be found from the maximum range of the recoil protons. Since the masses of the other nuclei in Eq. (50.25) are known, the mass of the neutron can be determined. The atomic weight derived in this way is 1.0090. The atomic number of the neutron is zero.

Since the charge of a neutron is zero, the potential energy of a neutron in the field of a nucleus is zero at large distances. Close to the nucleus, it is found that an attractive force exists. The potential-energy curve of a neutron is represented schematically in Fig. 50.5. This curve should be compared with that of an α particle or proton shown in Fig. 50.1. A fast neutron makes collisions with nuclei if the distance of approach is less than the nuclear radius R where the potential-energy curve becomes negative. In a collision the neutron is scattered, and measurements of the number of neutrons scattered are used to derive the values of the nuclear radii, which are expressed by Eq. (50.2).

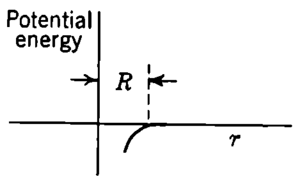


FIG. 50.5. The potential energy of a neutron at a distance r from a nucleus.

***50.9. The Binding Energies of Nuclei.** Neutrons and protons form the building blocks out of which all nuclei are constructed. For example, an α particle of mass number $M = 4$ and atomic number $Z = 2$ is composed of two neutrons and two protons. If N is the number of neutrons in a nucleus, then

$$M = Z + N \tag{50.26}$$

In elements of low mass number, Z is approximately equal to N. At the upper end of the periodic table N becomes larger than Z. In gold, for example, $Z = 79$, and only one isotope exists for which $M = 197$. The number N of neutrons is therefore 118, or nearly 1.5 times the number of protons.

Neutrons and protons are held together in the nucleus by forces whose nature is little understood at present. The nuclear forces must be attractive and vary more rapidly with distance than Coulomb forces. This is evident from the potential-energy curve of Fig. 50.1, where deviations from Coulomb repulsion for a proton occur only at small distances. At these small distances the nuclear force of attraction exceeds the Coulomb force of repulsion. It is the mutual repulsion of the protons in a nucleus that requires the presence of more neutrons than protons in heavy nuclei if a stable structure is to result.

Although our knowledge of the nature of nuclear forces themselves is still incomplete, it is easy to find experimentally the *binding energy* of a particle in a nucleus. The binding energy is the energy that must be supplied to remove the particle and is therefore exactly analogous to the ionization energy of an electron in an atom. It is necessary to know only the atomic weight of the nucleus and the weights of the neutrons and protons that compose it. The difference in atomic weight represents, by the mass-energy relation, the energy that would have to be supplied to separate the particles. Consider, for example, the α particle of atomic weight 4.0039. If an α particle is formed from ${}_2He^3$ of atomic weight 3.0171 by adding a neutron of atomic weight 1.0090, the total weight of the separated components is 4.0261. The binding energy of a neutron in ${}_2He^4$ is therefore $4.0261 - 4.0039 = 0.0222$, or 20.7 Mev. Since the mass of ${}_1H^3$ is almost exactly the same as that of ${}_2He^3$, the binding energy of a proton in an α particle is almost exactly equal to that of a neutron. The mass of two protons plus that of two neutrons is $2 \times 1.0081 + 2 \times 1.0090 = 4.0342$, which is greater than the atomic weight of ${}_2He^4$ by 0.0303 unit, or 28.2 Mev.

The total binding energy of a nucleus is called the *mass defect* ΔA. If A is the atomic weight, then

$$\Delta A = 1.0081Z + 1.0090N - A \tag{50.27}$$

For light elements, $Z \cong N \cong M/2$, and $A \cong M$. Hence

$$\Delta A \cong 0.0085M$$

The binding energy per particle in the nucleus is therefore 0.0085 mass unit, or about 8 Mev. The binding energy per particle in a nucleus increases with increasing atomic number until it reaches a broad maximum somewhere near the atomic number of iron or nickel; it then decreases again up to the highest atomic number. The nuclei in the center of the periodic table are thus the most stable ones.

50.10. Artificial Radioactivity. Nuclear reactions often lead to nuclei that are not present in nature. If sodium is bombarded by deuterons, for example, the reaction

$${}_{11}Na^{23} + {}_1H^2 \rightarrow {}_{11}Na^{24} + {}_1H^1 \tag{50.28}$$

takes place. The sodium isotope of mass number 24 is not ordinarily present in sodium. The reason is not hard to find since ${}_{11}Na^{24}$ is radioactive. The half-life is short, 14.8 hr, and any of this isotope that might have been formed once would have decayed long ago. The isotope ${}_{11}Na^{24}$, or radiosodium, is β-active and decays to ${}_{12}Mg^{24}$, a stable isotope. In a similar manner, radiophosphorus can be produced by the reaction

$${}_{15}P^{31} + {}_1H^2 \rightarrow {}_{15}P^{32} + {}_1H^1 \tag{50.29}$$

Radiophosphorus is also β-active and has a half-life of 15 days.

A radioactive form of nearly every element can be made by some nuclear process. Neutrons are particularly suitable for "inducing" radioactivity since they are not repelled away from the nucleus by Coulomb forces. Examples of neutron reactions are

$$\left.\begin{array}{l} {}_{13}Al^{27} + {}_{0}n^{1} \rightarrow {}_{11}Na^{24} + {}_{2}He^{4} \\ {}_{20}Ca^{42} + {}_{0}n^{1} \rightarrow {}_{19}K^{42} + {}_{1}H^{1} \end{array}\right\} \qquad (50.30)$$

The radioactive potassium produced by the second of these reactions has a half-life of 12.4 hr. Radioactive forms of common elements are extremely useful. Only a minute amount of the radioactive isotope is necessary for detection. Small amounts of a radioisotope can be added to an element and the presence of that element can be detected by following the radioactive *tracer* as the element undergoes complex chemical or biological changes. If radiosodium chloride dissolved in water is swallowed, the amount of radioactivity in the blood stream can be used to determine the rate of assimilation of salt.

50.11. The Positive Electron. In Sec. 50.6 two processes by which a γ ray can produce electrons were described—the photoelectric effect and the Compton effect. If the γ ray has a high enough energy, a third process occurs. In the field about a nucleus the energy of a γ ray can be transformed into the rest energy of *two* electrons, one negatively charged, the other positively charged. Radiation is thus directly transformed to mass. The rest energy of an electron is 0.5 Mev, and hence the energy of the γ ray must be larger than 1.0 Mev if pairs of electrons are to be produced. At very high energies, pair production is the principal means of absorption of γ rays. The positive electron was discovered in 1932 by C. D. Anderson in experiments on cosmic radiation with a Wilson cloud chamber.

The negative electron of an electron pair is identical with all other electrons. The positive electron has, however, only a transitory existence. Once a positive electron, or positron, has lost its kinetic energy by ionization, it is attracted to one of the numerous negative electrons that are present in all matter. The mass of two electrons disappears and is transformed back to radiation. In order to conserve momentum in this annihilation process, two quanta are produced, which travel in opposite directions, each with 0.5. Mev of energy.

Many of the artificially prepared radioactive elements emit positive electrons instead of negative ones. In fact the first radioactive element to be made, ${}_{15}P^{30}$, decays by positron emission with a half-life of 2.5 min.

***50.12. The Neutrino.** If a radioactive element emits α particles, the energy released has a definite value that can be accounted for by the difference in mass of the parent element and the mass of the daughter element plus the α particle. The α particle "comes out" of the nucleus. Whether it existed as such within the nucleus or whether it was formed at the moment of emission, from neutrons and protons, is not certain. If a radioactive element emits electrons, either positive or negative, the situation is much more complicated. Electrons do not exist as such within a nucleus, nor can they be constructed of neutrons and protons. Moreover the emitted electrons do not all have the same energy. Rather, they have a distribution in energy ranging from zero to some definite upper limit. Such an energy distribution is shown in Fig. 50.6 for P^{32}.

Careful experiments seem to show, however, that the same total amount of energy is lost by the nucleus at each decay process and that this energy corresponds to the upper limit of the β-ray energies. Where, then, does the rest of the energy go? To account for the apparent violation of a fundamental law of physics the existence of a new particle, the *neutrino*, has been suggested. The neutrino has a very small rest mass (possibly zero) and no charge. It can, however, possess energy and momentum, just as a photon without rest mass has energy and momentum. When a β ray is emitted with less than the maximum amount of energy, a neutrino is also supposed to be emitted and to carry away the extra energy. The neutrino has not yet been observed, and hence these suggestions cannot be regarded as proved.

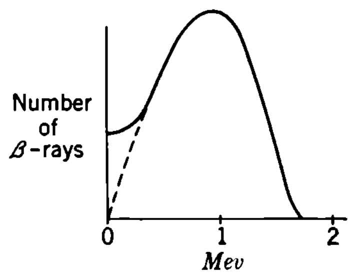


FIG. 50.6. Energy distribution of emitted electrons. The solid curve gives the experimental results for P^{32}. Theory predicts that the curve should pass through the origin as indicated by the dashed line. The difference is the result of scattering and absorption in the source.

50.13. Nuclear Fission and the Atomic Bomb. A nuclear reaction of an entirely different type is found to take place in various of the heaviest elements. If ${}_{92}U^{235}$ (actinouranium) absorbs a slow-moving neutron, a very violent reaction takes place. The compound nucleus ${}_{92}U^{236}$ is found to split into two parts of approximately equal mass, and several neutrons are emitted. This splitting is called *nuclear fission* and can take place in many ways. One reaction is, for example,

$$ {}_{92}U^{235} + {}_{0}n^{1} \longrightarrow [{}_{92}U^{236}] \longrightarrow {}_{54}X^{139} + {}_{38}Sr^{97-x} + x\,{}_{0}n^{1} \qquad (50.31)$$

where x is perhaps 2 or 3. The xenon and strontium formed are not stable elements, but each decays by radioactive processes until stable forms are produced. Thus

$$\left.\begin{aligned} {}_{54}X^{139} &\xrightarrow{0.5\text{ min}} {}_{55}Cs^{139} + \beta \\ {}_{55}Cs^{139} &\xrightarrow{6\text{ min}} {}_{56}Ba^{139} + \beta \\ {}_{56}Ba^{139} &\xrightarrow{86\text{ min}} {}_{57}La^{139} + \beta \end{aligned}\right\} \qquad (50.32)$$

where the half lives of the successive β transformations are indicated. Every element between selenium ($Z = 37$) and praseodymium ($Z = 59$)

has been found as a product of uranium fission. The end products of the fission reaction lie in the region of the periodic table where the elements are most stable. Consequently the energy released in the reaction is large; roughly 200 Mev is produced in each fission.

It should be noted that several neutrons are produced in the fission process. This circumstance makes possible a *chain reaction*. The neutrons produced in each fission of ${}_{92}U^{235}$ can be utilized to produce other fissions. Thus, if the neutrons are used efficiently, the process becomes a continuing one. The neutrons produced by fissions are fast neutrons, whereas the fission reaction requires slow ones. The neutrons must be slowed down by collisions with nuclei before they can be used again, and not too many neutrons must be lost by capture or by diffusion away from the uranium. If a large quantity of uranium is mixed with graphite, which serves to slow down the neutrons without capturing them, then the chain reaction proceeds and nuclear energy is continuously liberated. A large volume is necessary to minimize the escape of neutrons from the surface. Such a nuclear reactor has been called a *pile*.

Uranium metal is a mixture of isotopes of mass numbers 235 and 238, and only the first of these is fissionable by slow neutrons. The presence of the 238 isotope retards the chain reaction since U^{238} captures neutrons. If pure U^{235} were used, the chain reaction would proceed with explosive violence and an "atomic" bomb would be the result. It is extremely difficult, however, to separate this isotope from uranium metal.

The capture of a neutron by ${}_{92}U^{238}$ leads to the isotope ${}_{92}U^{239}$, which decays by two β-ray disintegrations, first to neptunium, ${}_{93}Np^{239}$, and then to plutonium, ${}_{94}Pu^{239}$. Plutonium is also fissionable by slow neutrons, and it can be chemically separated from uranium. Thus a uranium pile can be used to manufacture plutonium, and the plutonium used for an atomic bomb.

***50.14. Energy Production in Stars.** The transformation of nuclear mass into energy is the means of energy production in stars and in our sun. The ultimate source of all the energy that we use is thus the rest energy of nuclei. Within the interior of a star the temperature is exceedingly large, of the order of tens of millions of degrees. The thermal velocities that particles have at these temperatures are correspondingly high, and nuclear reactions occur. In our sun and many other stars a chain of reactions occurs that is represented by the following equations:

$$\left.\begin{aligned} {}_1H^1 + {}_6C^{12} &\rightarrow {}_7N^{13} \\ {}_7N^{13} &\rightarrow {}_6C^{13} + \beta^+ \\ {}_1H^1 + {}_6C^{13} &\rightarrow {}_7N^{14} \\ {}_1H^1 + {}_7N^{14} &\rightarrow {}_8O^{15} \\ {}_8O^{15} &\rightarrow {}_7N^{15} + \beta^+ \\ {}_1H^1 + {}_7N^{15} &\rightarrow {}_6C^{12} + {}_2He^4 \end{aligned}\right\} \tag{50.33}$$

The symbol β^+ indicates a positron emitted radioactively. The net result of this chain of reactions is the transformation of four protons into an α particle and two positrons. The carbon starting the reaction is restored at the end. The energy release can be calculated from the masses and is found to be 27.7 Mev for each α particle formed. The sun is therefore "burning" its hydrogen into helium to produce energy. From the amount of hydrogen present in the sun, it can be calculated that this transformation can continue for at least 10^7 years.

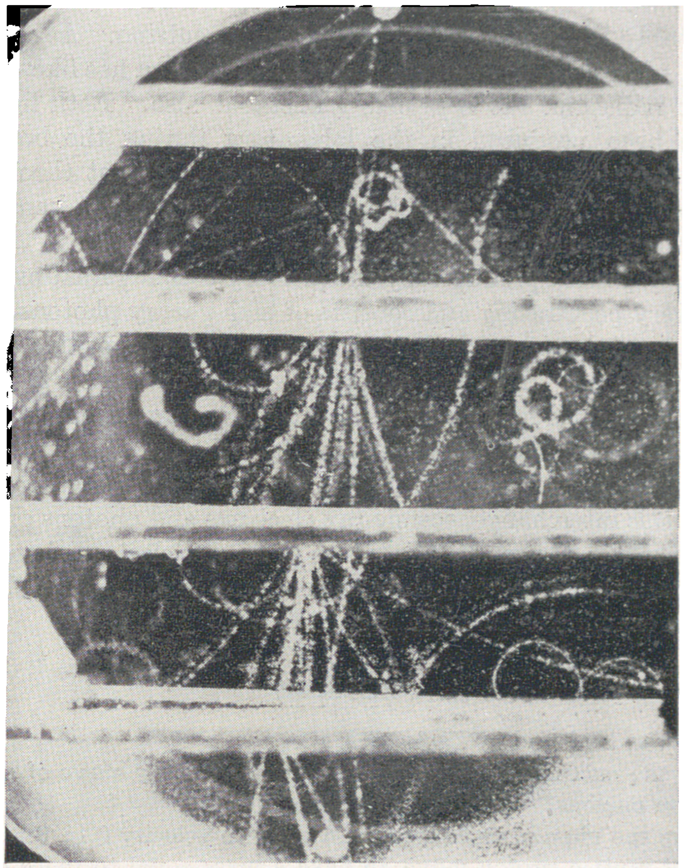


FIG. 50.7. A cloud-chamber photograph of a cosmic-ray shower. An electron enters at the top and produces many secondary particles in the lead plates. (*Courtesy of W. E. Hazen.*)

★50.15. Cosmic Radiation and the Mesotron. Attempts to explain the conductivity of the air and the production of ion pairs have led to the conclusion that the earth is continually bombarded by a *cosmic radiation* consisting of charged particles of high energy coming from unknown sources in outer space. The number and energy of these particles can be estimated from their deflection by the magnetic field of the earth, which results in a variation of the intensity of cosmic-ray ionization with latitude. It is found that about 1.5 particles per square centimeter per second are incident and that these particles are positively charged. It seems likely that they are protons. The energy of each particle is much larger than can be produced in the laboratory and ranges from 3×10^9 ev up to perhaps 10^{16} ev, with a mean value of about 10^{10} ev. The interaction of particles of such high energy with matter is a subject of much interest.

Although the details of the processes involved are completely unknown, the interaction of the primary cosmic radiation with the atmosphere leads to the production of radiation of two sorts. The more penetrating, or *hard*, component consists of particles each having a single electronic charge, either positive or negative, and a mass 200 times that of an electron. The particles are called *mesotrons*, or *mesons*. Mesotrons are radioactive, with a half-life of about 2 μsec. They presumably decay into an electron and a neutrino. Another particle, a heavy meson, of about 300 times the mass of the electron has likewise been found in the cosmic radiation. It is probably unstable, also. Both light and heavy mesons have been produced in the laboratory during the bombardment of nuclei with α particles of extremely high energy.[1] Recent theories of nuclear structure explain nuclear forces in terms of mesons, but our knowledge is very imperfect at the present time.

The other component produced by the primary cosmic radiation in the atmosphere is not very penetrating and consists of high-energy photons and electrons. The photons are absorbed by pair production in which a positive and a negative electron are produced. The high-energy electrons, on the other hand, produce photons when they encounter nuclei just as electrons produce X rays in an X-ray tube. There is consequently a continual interchange of energy between the electrons and photons in the cosmic radiation passing through the atmosphere. As a result of this interchange, cosmic-ray electrons and photons arrive in groups of rays, or *cosmic-ray showers*. The tracks of shower electrons in a Wilson cloud chamber are easily observed and show the extreme complexity of the phenomena involved. An example is shown in Fig. 50.7.

PROBLEMS

Mass values can be found in Table 29.2.

1. What is the maximum height of the potential-energy curve of a proton in the field of a carbon nucleus? A bismuth nucleus?

2. Calculate the mass of an electron moving with velocity $0.1c$, $0.5c$, $0.99c$, $0.999c$.

3. How fast must a proton move to have a mass equal to twice the rest mass? What is its energy in electron volts?

4. What is the mass equivalent of 1 ev? Express the answer in grams and on the scale of atomic weights.

5. How long does it take a polonium source of α particles to decay to one-tenth the original value?

6. A mineral contains equal numbers of atoms of U^{238} and Th^{232}. How long a time must elapse before there are 10 times as many Th as U atoms?

7. A polonium α particle has an energy of 5.3 Mev. What is the energy of recoil of the residual lead nucleus?

***8.** An X-ray photon of 50,000 ev energy is scattered by the Compton effect at an angle of 90°. What is the energy of the scattered photon?

[1] The α particles used for these experiments were accelerated in the cyclotron pictured in Fig. 29.9.

9. Write the equations for three reactions that might occur when deuterons bombard $_6C^{13}$.

10. How much mass energy is released or absorbed in the reaction when protons bombard $_6C^{13}$ to produce $_7N^{14}$?

***11.** Calculate the binding energy of the deuteron.

12. Photons can eject protons from a nucleus by an action similar to the photoelectric effect. How much energy must a photon have to disintegrate $_2He^3$ by this process?

***13.** What is the energy in calories per square centimeter per second incident on the earth in the form of cosmic radiation? Compare this with the energy received from the sun. See page 684 for solar constant.

14. How much mass is radiated by the sun in 1 sec? Use the data given in Prob. 14, Chap. 48.

APPENDIX 1
UNITS

Length:

1 meter (m)	= 100 centimeters (cm)
1 micron (μ)	= 10^{-4} cm
1 millimicron (mμ)	= 10^{-7} cm
1 Angstrom unit (A)	= 10^{-8} cm
1 inch (in.)	= 2.540 cm
1 foot (ft)	= 30.48 cm
1 mile	= 1.609 km
1 mile	= 5,280 ft
1 cm	= 0.3937 in.
1 m	= 39.37 in.
1 kilometer (km)	= 0.6214 mile
1 in.	= 1,000 mils

Area:

1 cm^2	= 0.155 $in.^2$
1 $in.^2$	= 6.452 cm^2
1 m^2	= 10.76 ft^2
1 ft^2	= 0.0929 m^2

Volume:

1 liter	= 1,000 cm^3
1 liter	= 1.057 quarts (qt)
1 quart (qt)	= 0.9463 liter
1 gallon (gal)	= 231 $in.^3$
1 liter	= 0.0351 ft^3
1 ft^3	= 28.32 liters
1 ft^3	= 7.481 gal
1 $in.^3$	= 16.39 cm^3
1 cm^3	= 0.061 $in.^3$

Mass:

1 gram (gm)	= 1,000 milligrams (mg)
1 kilogram (kg)	= 2.205 pounds (lb)
1 gm	= 0.03527 ounce (oz)
1 ounce (oz)	= 28.35 gm
1 pound (lb)	= 453.6 gm
1 lb	= 7,000 grains
1 U.S. short ton	= 2,000 lb
1 metric ton	= 2,205 lb
1 slug	= 32.2 lb

Density:

1 gm/cm^3	= 62.43 lb/ft^3 = 8.345 lb/gal
1 lb/ft^3	= 0.01602 gm/cm^3

Velocity:

1 cm/sec	= 0.03281 ft/sec
1 cm/sec	= 0.02237 mile/hr
1 ft/sec	= 30.48 cm/sec
1 mile/hr	= 44.7 cm/sec
1 mile/min	= 60 miles/hr = 88 ft/sec

Acceleration:

1 cm/sec^2	= 0.03281 ft/sec^2
g	= 981 cm/sec^2 = 32.2 ft/sec^2

Force:

1 dyne	= 7.233×10^{-5} poundal
1 poundal	= 13,825 dynes
1 gm-force	= 981 dynes
1 dyne	= 2.247×10^{-6} lb-force
1 newton	= 10^5 dynes
1 lb-force	= 4.45×10^5 dynes
1 lb-force	= 32.2 poundals

Pressure:

1 atmosphere (atm)	= 14.7 lb/in^2 = 1,033 gm/cm^2
1 atm	= 760 mm of mercury at 0°C
1 atm	= 1,013,000 $dynes/cm^2$
1 atm	= 1.013 bars = 1,013 millibars
1 $dyne/cm^2$	= 1.45×10^{-5} lb/in^2
1 lb/in^2	= 69,870 $dynes/cm^2$
1 gm/cm^2	= 0.01422 lb/in^2
1 lb/in^2	= 70.31 gm/cm^2

Energy:

1 joule	= 10^7 ergs = 0.239 cal = 0.738 ft lb = 0.1020 kg m
1 calorie (cal)	= 4.186 joules = 3.087 ft lb = 0.427 kg m
1 British thermal unit (Btu)	= 1,055 joules = 252 cal = 778 ft lb = 0.293 watt hr
1 ft lb	= 1.356 joules = 0.3239 cal = 0.001285 Btu
1 erg	= 0.7373×10^{-7} ft lb
1 ft lb	= 1.356×10^7 ergs
1 kilowatt hour (kwhr)	= 36×10^{12} ergs
1 kg m	= 7.233 ft lb

Power:

1 horsepower (hp)	= 33,000 ft lb/min = 550 ft lb/sec = 746 watts
1 watt	= 1 joule/sec = 10^7 ergs/sec = 0.239 cal/sec
1 kilowatt	= 1.341 hp = 738 ft lb/sec = 0.948 Btu/sec

APPENDIX 2

VALUES OF IMPORTANT GENERAL PHYSICAL CONSTANTS

Acceleration of gravity (normal) $g = 980.665$ cm sec^{-2}
Liter = 1,000.028 ± 0.002 cm^3
Maximum density of water $\rho_m(H_2O) = 0.999972 \pm 0.000002$ gm cm^{-3}
Velocity of light $c = (2.99776 \pm 0.00004) \times 10^{10}$ cm sec^{-1}
Gravitation constant $G = (6.670 \pm 0.005) \times 10^{-8}$ dyne cm^{-2} gm^{-2}
Normal atmospheric pressure $p_0 = (1.013246 \pm .000004) \times 10^6$ dyne cm^{-2}
Volume of perfect gas (0°C, p_0, *i.e.*, standard atmospheric pressure)

$$V_0 = 22.4146 \pm 0.0006 \text{ liter mole}^{-1}$$

Avogadro's number $N_0 = (6.0228 \pm 0.0011) \times 10^{23}$ mole^{-1}
Gas constant per mole $R_0 = V_0 p_0 / T_0 = (8.31436 \pm 0.00038) \times 10^7$ erg deg^{-1} mole^{-1}
Boltzmann constant $k = R_0/N_0 = (1.38047 \pm 0.00026) \times 10^{-16}$ erg deg^{-1}
Ice point (absolute scale) $T_0 = 273.16 \pm 0.01$°K
Mechanical equivalent of heat $J = 4.1855 \pm 0.0004$ joule cal^{-1}
Mass of an electron $m_0 = (9.1066 \pm 0.0032) \times 10^{-28}$ gm
Mass of hydrogen atom $M_H = (1.67339 \pm 0.00031) \times 10^{-24}$ gm
Mass of proton $M_p = (1.67248 \pm 0.00031) \times 10^{-24}$ gm
Ratio, mass H atom to mass electron $(e/m)/(e/M_H) = 1{,}837.5 \pm 0.5$
Mass of α particle $M_\alpha = (6.6442 \pm 0.0012) \times 10^{-24}$ gm
Faraday constant $F = 96{,}501 \pm 10$ coulombs per gm equiv.
Electronic charge $e = (4.8025 \pm 0.001) \times 10^{-10}$ esu
Planck constant $h = (6.624 \pm 0.002) \times 10^{-27}$ erg sec

APPENDIX 3
CONVERSION TABLE OF ELECTRICAL UNITS

The entries in the last two columns are the equivalents of the quantities in the third column in electrostatic and electromagnetic units respectively.

Quantity	Symbol	Practical unit (mks)	Electrostatic unit (esu)	Electromagnetic unit (emu)
Mass	m	1 kilogram	1,000 grams	1,000 grams
Length	l	1 meter	100 centimeters	100 centimeters
Time	t	1 second	1 second	1 second
Force	F	1 newton	10^5 dynes	10^5 dynes
Energy	W	1 joule	10^7 ergs	10^7 ergs
Power	P	1 watt	10^7 ergs/second	10^7 ergs/second
Charge	Q	1 coulomb	3×10^9 statcoulombs	0.1 abcoulomb
Current	I	1 ampere	3×10^9 statamperes	0.1 abampere
Electric potential	V	1 volt	$\frac{1}{3} \times 10^{-2}$ statvolt	10^8 abvolts
Electric field strength	E	1 volt/meter	$\frac{1}{3} \times 10^{-4}$ statvolt/centimeter	10^6 abvolts/centimeter
Resistance	R	1 ohm	$\frac{1}{9} \times 10^{-11}$ statohm	10^9 abohms
Inductance	L	1 henry	$\frac{1}{9} \times 10^{-11}$ stathenry	10^9 centimeters
Capacitance	C	1 farad	9×10^{11} centimeters	10^{-9} abfarad
Flux of magnetic induction	Φ	1 weber		10^8 maxwells
Magnetic induction	B	1 weber/square meter		10^4 gausses
Magnetic intensity	H	1 ampere/meter		$4\pi \times 10^{-3}$ oersted
Permittivity	ϵ	1 farad/meter	$\frac{1}{8.85 \times 10^{-12}}$*	
Permittivity of free space	ϵ_0	8.85×10^{-12} farad/meter	1*	
Permeability	μ	1 henry/meter		$\frac{1}{1.257 \times 10^{-6}}$*
Permeability of free space	μ_0	1.257×10^{-6} henry/meter		1*

* These units are pure numbers.

APPENDIX 4

GREEK ALPHABET

Alpha (a)	A α	Nu (n)	N ν
Beta (b)	B β	Xi (x)	Ξ ξ
Gamma (g)	Γ γ	Omicron (o)	O o
Delta (d)	Δ δ or ∂	Pi (p)	Π π
Epsilon (e)	E ϵ	Rho (r)	P ρ
Zeta (z)	Z ζ	Sigma (s)	Σ σ or ς
Eta (h)	H η	Tau (t)	T τ
Theta (th)	Θ θ	Upsilon (u)	Υ υ
Iota (i)	I ι	Phi (ph)	Φ φ or ϕ
Kappa (k)	K κ	Chi (ch)	X χ
Lambda (l)	Λ λ	Psi (ps)	Ψ ψ
Mu (m)	M μ	Omega (o)	Ω ω

APPENDIX 5
USEFUL MATHEMATICS

$$\pi = 3.1416$$
$$e = 2.7183$$

Binomial Theorem:

$$(1 \pm x)^n = 1 \pm nx + \frac{n(n-1)}{2}x^2 \pm \frac{n(n-1)(n-2)}{2 \cdot 3}x^3 + \cdots \quad \text{provided } x^2 < 1$$

Special cases:

$$(1 \pm x)^{-1} = 1 \mp x + x^2 \mp x^3 + \cdots$$
$$(1 \pm x)^{1/2} = 1 \pm \frac{x}{2} - \frac{x^2}{8} \pm \frac{3x^3}{48} - \cdots$$
$$(1 \pm x)^{-1/2} = 1 \mp \frac{x}{2} + \frac{3x^2}{8} \mp \frac{15x^3}{48} + \cdots$$

Series:

$$e^x = 1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \cdots$$
$$\sin x = x - \frac{x^3}{6} + \cdots$$
$$\cos x = 1 - \frac{x^2}{2} + \cdots$$
$$\tan x = x + \frac{x^3}{3} + \cdots$$

Integration and Differentiation:

$\frac{dy}{dx}$	$y = \int \left(\frac{dy}{dx}\right) dx$*
x^n	$\frac{x^{n+1}}{n+1}$
nx^{n-1}	x^n
$\sin x$	$-\cos x$
$\cos x$	$\sin x$
$\tan x$	$-\ln \cos x$
$\cot x$	$\ln \sin x$
$\sec^2 x$	$\tan x$
$-\operatorname{cosec}^2 x$	$\cot x$
$\sin^2 x$	$\frac{1}{2}x - \frac{1}{2}\sin x \cos x$
$\cos^2 x$	$\frac{1}{2}x + \frac{1}{2}\sin x \cos x$
$\frac{1}{\sqrt{A^2 + x^2}}$	$\sin^{-1}\frac{x}{A}$
$\frac{A}{x^2 + A^2}$	$\tan^{-1}\frac{x}{A}$
e^{ax}	$\frac{e^{ax}}{a}$

* The constant of integration if omitted in all cases.

APPENDIX 6

NATURAL TRIGONOMETRIC FUNCTIONS

Angle		Sine	Cosine	Tan-gent	Angle		Sine	Cosine	Tan-gent
°	radians				°	radians			
0	0.0000	0.000	1.000	0.000					
1	0.0175	.018	1.000	.018	46	0.8029	.719	.695	1.036
2	0.0349	.035	0.999	.035	47	0.8203	.731	.682	1.072
3	0.0524	.052	.999	.052	48	0.8378	.743	.669	1.111
4	0.0698	.070	.998	.070	49	0.8552	.755	.656	1.150
5	0.0873	.087	.996	.088	50	0.8727	.766	.643	1.192
6	0.1047	.105	.995	.105	51	0.8901	.777	.629	1.235
7	0.1222	.122	.993	.123	52	0.9076	.788	.616	1.280
8	0.1396	.139	.990	.141	53	0.9250	.799	.602	1.327
9	0.1571	.156	.988	.158	54	0.9425	.809	.588	1.376
10	0.1745	.174	.985	.176	55	0.9599	.819	.574	1.428
11	1.1920	.191	.982	.194	56	0.9774	.829	.559	1.483
12	0.2094	.208	.978	.213	57	0.9948	.839	.545	1.540
13	0.2269	.225	.974	.231	58	1.0123	.848	.530	1.600
14	0.2443	.242	.970	.249	59	1.0297	.857	.515	1.664
15	0.2618	.259	.966	.268	60	1.0472	.866	.500	1.732
16	0.2793	.276	.961	.287	61	1.0647	.875	.485	1.804
17	0.2967	.292	.956	.306	62	1.0821	.883	.470	1.881
18	0.3142	.309	.951	.325	63	1.0996	.891	.454	1.963
19	0.3316	.326	.946	.344	64	1.1170	.899	.438	2.050
20	0.3491	.342	.940	.364	65	1.1345	.906	.423	2.145
21	0.3665	.358	.934	.384	66	1.1519	.914	.407	2.246
22	0.3840	.375	.927	.404	67	1.1694	.921	.391	2.356
23	0.4014	.391	.921	.425	68	1.1868	.927	.375	2.475
24	0.4189	.407	.914	.445	69	1.2043	.934	.358	2.605
25	0.4363	.423	.906	.466	70	1.2217	.940	.342	2.747
26	0.4538	.438	.899	.488	71	1.2392	.946	.326	2.904
27	0.4712	.454	.891	.510	72	1.2566	.951	.309	3.078
28	0.4887	.470	.883	.532	73	1.2741	.956	.292	3.271
29	0.5061	.485	.875	.554	74	1.2915	.951	.276	3.487
30	0.5236	.500	.866	.577	75	1.3090	.966	.259	3.732
31	0.5411	.515	.857	.601	76	1.3265	.970	.242	4.011
32	0.5585	.530	.848	.625	77	1.3439	.974	.225	4.331
33	0.5760	.545	.839	.649	78	1.3614	.978	.208	4.705
34	0.5934	.559	.829	.675	79	1.3788	.982	.191	5.145
35	0.6109	.574	.819	.700	80	1.3963	.985	.174	5.671
36	0.6283	.588	.809	.727	81	1.4137	.988	.156	6.314
37	0.6458	.602	.799	.754	82	1.4312	.990	.139	7.115
38	0.6632	.616	.788	.781	83	1.4486	.993	.122	8.144
39	0.6807	.629	.777	.810	84	1.4661	.995	.105	9.514
40	0.6981	.643	.766	.839	85	1.4835	.996	.087	11.43
41	0.7156	.656	.755	.869	86	1.5010	.998	.070	14.30
42	0.7330	.669	.743	.900	87	1.5184	.999	.052	19.08
43	0.7505	.682	.731	.933	88	1.5359	.999	.035	28.64
44	0.7679	.695	.719	.966	89	1.5533	1.000	.018	57.29
45	0.7854	.707	.707	1.000	90	1.5708	1.000	.000	∞

APPENDIX 7

	Logarithms										Proportional parts								
	0	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9
10	0000	0043	0086	0128	0170	0212	0253	0294	0334	0374	4	8	12	17	21	25	29	33	37
11	0414	0453	0492	0531	0569	0607	0645	0682	0719	0755	4	8	11	15	19	23	26	30	34
12	0792	0828	0864	0899	0934	0969	1004	1038	1072	1106	3	7	10	14	17	21	24	28	31
13	1139	1173	1206	1239	1271	1303	1335	1367	1399	1430	3	6	10	13	16	19	23	26	29
14	1461	1492	1523	1553	1584	1614	1644	1673	1703	1732	3	6	9	12	15	18	21	24	27
15	1761	1790	1818	1847	1875	1903	1931	1959	1987	2014	3	6	8	11	14	17	20	22	25
16	2041	2068	2095	2122	2148	2175	2201	2227	2253	2279	3	5	8	11	13	16	18	21	24
17	2304	2330	2355	2380	2405	2430	2455	2480	2504	2529	2	5	7	10	12	15	17	20	22
18	2553	2577	2601	2625	2648	2672	2695	2718	2742	2765	2	5	7	9	12	14	16	19	21
19	2788	2810	2833	2856	2878	2900	2923	2945	2967	2989	2	4	7	9	11	13	16	18	20
20	3010	3032	3054	3075	3096	3118	3139	3160	3181	3201	2	4	6	8	11	13	15	17	19
21	3222	3243	3263	3284	3304	3324	3345	3365	3385	3404	2	4	6	8	10	12	14	16	18
22	3424	3444	3464	3483	3502	3522	3541	3560	3579	3598	2	4	6	8	10	12	14	15	17
23	3617	3636	3655	3674	3692	3711	3729	3747	3766	3784	2	4	6	7	9	11	13	15	17
24	3802	3820	3838	3856	3874	3892	3909	3927	3945	3962	2	4	5	7	9	11	12	14	16
25	3979	3997	4014	4031	4048	4065	4082	4099	4116	4133	2	3	5	7	9	10	12	14	15
26	4150	4166	4183	4200	4216	4232	4249	4265	4281	4298	2	3	5	7	8	10	11	13	15
27	4314	4330	4346	4362	4378	4393	4409	4425	4440	4456	2	3	5	6	8	9	11	13	14
28	4472	4487	4502	4518	4533	4548	4564	4579	4594	4609	2	3	5	6	8	9	11	12	14
29	4624	4639	4654	4669	4683	4698	4713	4728	4742	4757	1	3	4	6	7	9	10	12	13
30	4771	4786	4800	4814	4829	4843	4857	4871	4886	4900	1	3	4	6	7	9	10	11	13
31	4914	4928	4942	4955	4969	4983	4997	5011	5024	5038	1	3	4	6	7	8	10	11	12
32	5051	5065	5079	5092	5105	5119	5132	5145	5159	5172	1	3	4	5	7	8	9	11	12
33	5185	5198	5211	5224	5237	5250	5263	5276	5289	5302	1	3	4	5	6	8	9	10	12
34	5315	5328	5340	5353	5366	5378	5391	5403	5416	5428	1	3	4	5	6	8	9	10	11
35	5441	5453	5465	5478	5490	5502	5514	5527	5539	5551	1	2	4	5	6	7	9	10	11
36	5563	5575	5587	5599	5611	5623	5635	5647	5658	5670	1	2	4	5	6	7	8	10	11
37	5682	5694	5705	5717	5729	5740	5752	5763	5775	5786	1	2	3	5	6	7	8	9	10
38	5798	5809	5821	5832	5843	5855	5866	5877	5888	5899	1	2	3	5	6	7	8	9	10
39	5911	5922	5933	5944	5955	5966	5977	5988	5999	6010	1	2	3	4	5	7	8	9	10
40	6021	6031	6042	6053	6064	6075	6085	6096	6107	6117	1	2	3	4	5	6	8	9	10
41	6128	6138	6149	6160	6170	6180	6191	6201	6212	6222	1	2	3	4	5	6	7	8	9
42	6232	6243	6253	6263	6274	6284	6294	6304	6314	6325	1	2	3	4	5	6	7	8	9
43	6335	6345	6355	6365	6375	6385	6395	6405	6415	6425	1	2	3	4	5	6	7	8	9
44	6435	6444	6454	6464	6474	6484	6493	6503	6513	6522	1	2	3	4	5	6	7	8	9
45	6532	6542	6551	6561	6571	6580	6590	6599	6609	6618	1	2	3	4	5	6	7	8	9
46	6628	6637	6646	6656	6665	6675	6684	6693	6702	6712	1	2	3	4	5	6	7	7	8
47	6721	6730	6739	6749	6758	6767	6776	6785	6794	6803	1	2	3	4	5	5	6	7	8
48	6812	6821	6830	6839	6848	6857	6866	6875	6884	6893	1	2	3	4	4	5	6	7	8
49	6902	6911	6920	6928	6937	6946	6955	6964	6972	6981	1	2	3	4	4	5	6	7	8
50	6990	6998	7007	7016	7024	7033	7042	7050	7059	7067	1	2	3	3	4	5	6	7	8
51	7076	7084	7093	7101	7110	7118	7126	7135	7143	7152	1	2	3	3	4	5	6	7	8
52	7160	7168	7177	7185	7193	7202	7210	7218	7226	7235	1	2	2	3	4	5	6	7	7
53	7243	7251	7259	7267	7275	7284	7292	7300	7308	7316	1	2	2	3	4	5	6	6	7
54	7324	7332	7340	7348	7356	7364	7372	7380	7388	7396	1	2	2	3	4	5	6	6	7
	0	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9

	Logarithms										Proportional parts								
	0	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9
55	7404	7412	7419	7427	7435	7443	7451	7459	7466	7474	1	2	2	3	4	5	5	6	7
56	7482	7490	7497	7505	7513	7520	7528	7536	7543	7551	1	2	2	3	4	5	5	6	7
57	7559	7566	7574	7582	7589	7597	7604	7612	7619	7627	1	2	2	3	4	5	5	6	7
58	7634	7642	7649	7657	7664	7672	7679	7686	7694	7701	1	1	2	3	4	4	5	6	7
59	7709	7716	7723	7731	7738	7745	7752	7760	7767	7774	1	1	2	3	4	4	5	6	7
60	7782	7789	7796	7803	7810	7818	7825	7832	7839	7846	1	1	2	3	4	4	5	6	6
61	7853	7860	7868	7875	7882	7889	7896	7903	7910	7917	1	1	2	3	4	4	5	6	6
62	7924	7931	7938	7945	7952	7959	7966	7973	7980	7987	1	1	2	3	3	4	5	6	6
63	7993	8000	8007	8014	8021	8028	8035	8041	8048	8055	1	1	2	3	3	4	5	5	6
64	8062	8069	8075	8082	8089	8096	8102	8109	8116	8122	1	1	2	3	3	4	5	5	6
65	8129	8136	8142	8149	8156	8162	8169	8176	8182	8189	1	1	2	3	3	4	5	5	6
66	8195	8202	8209	8215	8222	8228	8235	8241	8248	8254	1	1	2	3	3	4	5	5	6
67	8261	8267	8274	8280	8287	8293	8299	8306	8312	8319	1	1	2	3	3	4	5	5	6
68	8325	8331	8338	8344	8351	8357	8363	8370	8376	8382	1	1	2	3	3	4	4	5	6
69	8388	8395	8401	8407	8414	8420	8426	8432	8439	8445	1	1	2	2	3	4	4	5	6
70	8451	8457	8463	8470	8476	8482	8488	8494	8500	8506	1	1	2	2	3	4	4	5	6
71	8513	8519	8525	8531	8537	8543	8549	8555	8561	8567	1	1	2	2	3	4	4	5	5
72	8573	8579	8585	8591	8597	8603	8609	8615	8621	8627	1	1	2	2	3	4	4	5	5
73	8633	8639	8645	8651	8657	8663	8669	8675	8681	8686	1	1	2	2	3	4	4	5	5
74	8692	8698	8704	8710	8716	8722	8727	8733	8739	8745	1	1	2	2	3	4	4	5	5
75	8751	8756	8762	8768	8774	8779	8785	8791	8797	8802	1	1	2	2	3	3	4	5	5
76	8808	8814	8820	8825	8831	8837	8842	8848	8854	8859	1	1	2	2	3	3	4	5	5
77	8865	8871	8876	8882	8887	8893	8899	8904	8910	8915	1	1	2	2	3	3	4	4	5
78	8921	8927	8932	8938	8943	8949	8954	8960	8965	8971	1	1	2	2	3	3	4	4	5
79	8976	8982	8987	8993	8998	9004	9009	9015	9020	9025	1	1	2	2	3	3	4	4	5
80	9031	9036	9042	9047	9053	8058	9063	9069	9074	9079	1	1	2	2	3	3	4	4	5
81	9085	9090	9096	9101	9106	9112	9117	9122	9128	9133	1	1	2	2	3	3	4	4	5
82	9138	9143	9149	9154	9159	9165	9170	9175	9180	9186	1	1	2	2	3	3	4	4	5
83	9191	9196	9201	9206	9212	9217	9222	9227	9232	9238	1	1	2	2	3	3	4	4	5
84	9243	9248	9253	9258	9263	9269	9274	9279	9284	9289	1	1	2	2	3	3	4	4	5
85	9294	9299	9304	9309	9315	9320	9325	9330	9335	9340	1	1	2	2	3	3	4	4	5
86	9345	9350	9355	9360	9365	9370	9375	9380	9385	9390	1	1	2	2	3	3	4	4	5
87	9395	9400	9405	9410	9415	9420	9425	9430	9435	9440	0	1	1	2	2	3	3	4	4
88	9445	9450	9455	9460	9465	9469	9474	9479	9484	9489	0	1	1	2	2	3	3	4	4
89	9494	9499	9504	9509	9513	9518	9523	9528	9533	9538	0	1	1	2	2	3	3	4	4
90	9542	9547	9552	9557	9562	9566	9571	9576	9581	9586	0	1	1	2	2	3	3	4	4
91	9590	9595	9600	9605	9609	9614	9619	9624	9628	9633	0	1	1	2	2	3	3	4	4
92	9638	9643	9647	9652	9657	9661	9666	9671	9675	9680	0	1	1	2	2	3	3	4	4
93	9685	9689	9694	9699	9703	9708	9713	9717	9722	9727	0	1	1	2	2	3	3	4	4
94	9731	9736	9741	9745	9750	9754	9759	9763	9768	9773	0	1	1	2	2	3	3	4	4
95	9777	9782	9786	9791	9795	9800	9805	9809	9814	9818	0	1	1	2	2	3	3	4	4
96	9823	9827	9832	9836	9841	9845	9850	9854	9859	9863	0	1	1	2	2	3	3	4	4
97	9868	9872	9877	9881	9886	9890	9894	9899	9903	9908	0	1	1	2	2	3	3	4	4
98	9912	9917	9921	9926	9930	9934	9939	9943	9948	9952	0	1	1	2	2	3	3	4	4
99	9956	9961	9965	9969	9974	9978	9983	9987	9991	9996	0	1	1	2	2	3	3	3	4
	0	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9

INDEX

A

Abcoulomb, 391
Aberration, 598
  chromatic, 598, 602
  in lenses, 597
  spherical, 573, 584, 598
Absolute humidity, 315
Absolute temperature, 224, 260, 298
Absolute units of force, 81
Absolute zero, 262
Absorbing power, acoustic, 548
Absorptance, 678, 685
Absorption, selective, 675
Absorption coefficient, 547
Absorption spectra, 628
Abvolt, 432
Acceleration, 62, 64
  centripetal, 74, 97
  constant, 68
  instantaneous, 72
Accommodation of eye, 606
Accuracy, 7
Achromat, 604
Achromatic lens, 604
Achromatic prism, 578
Acoustics, architectural, 546
Action, 105
  Planck's constant of, 170, 237, 489, 555, 690
Addition, of colors, 632
  of many forces, 14, 18
  of two forces, 13
  rule for forces, 14
Adhesion, 214
Adiabatic bulk modulus of elasticity, 267
Adiabatic process, 266
Age of minerals, 713
Air-mass analysis, 321
Airplane, in curved flight, 101
  lift on wing, 209
Air-speed indicator, 207
Alpha particles, 329, 711
  range of, 715
  scattering of, 708
Alternating emf, 435
Ammeter, 402
Ampère, André, 366, 408
Ampere, the, 366
Ampère's law, 409
Amplification, gas, 492
Amplification factor, 476
Amplifier, gain of, 477
  triode, 476
Amplitude, 173, 502
  displacement, 187
  pressure, 529
  velocity, 186
Analyzer, 674
Anaxagoras, 10
Andrews, Thomas, 268
Anemometer, 314
Angle, of contact, 217
  of dip, 427
  glancing, 667
  limiting, of repose, 87
  polarizing, 672
  of resolution, 660
Angstrom unit, 563, 729
Angular dispersion, 577
Angular impulse, 152
Angular momentum, 152
Angular velocity, 75
Anode, 380, 471
Antinode, 514
Aperture, 598
  diffraction by circular, 658
  Fresnel diffraction by circular, 653
  relative, 614
Aqueous humor, 606
Archimedes' principle, 196
Arcturus, spectrum of, 705
Aristotle, 10
Armature, 406, 435
Artificial radioactivity, 721
Artificial transmutation, 717
Aspirator, 207
Astigmatic difference, 600
Astigmatism, 585, 598, 600, 607
Astronomical spectra, 704
Atmosphere, 198, 312
Atmospheric humidity, 315

Atmospheric pressure, 8, 314
Atmospheric refraction, 589
Atom, ionized, 697
  mass of, 284
  Rutherford model, 690
  shape of, 285
  size of, 285
Atomic bomb, 723
Atomic energy, 397
Atomic heat capacity, 237
Atomic number, 686, 694
Atomic volume, 686
Attraction, inverse-square, 169
Atwood machine, 85, 91
Audibility, lower limit of, 542
  threshold of, 544
Available power, 383
Avogadro's law, 263, 275
Avogadro's number, 278
Axes, theorem of parallel, 144
Axis, crystallographic, 674
  fixed, 140
  principal, 581
  rigidity of gyroscopic, 154

B

Back emf, 382, 406
Balance, beam, 39
  sensitivity of, 39
Ballistic galvanometer, 405
Ballistic pendulum, 127
Ballistics, 96
Balloons, 198
Balmer series, 689
Band spectra, 700
  of O_2, 706
Banked road, 99
Bar, 192
Bar photometer, 623
Barkhausen effect, 451
Barometer, aneroid, 196
  mercury, 195
Barrel distortion, 602
Baseball, pitched, 131
Basilar membrane, 546
Battery, 383
Beam balance, 39
Beats, 512
Bel, the, 543
Bell, 539
Bell, A. G., 543
Belt, 158
Bernoulli, Daniel, 205
Bernoulli's equation, 205
Bessel, Friedrich, 84
Beta particle, 702, 711
  ionization by, 715
Beta rays, 702, 711
Betatron, 432
Billiards, 129
Bimetallic strip, 227
Binding energy of nuclei, 720
Binocular, prism, 613
Biprism, 642
Birge, R. T., 561
Bjerknes, Vilhelm and Jacob, 320
Black, Joseph, 233
Black body, 310
  ideal, 679
Black-body radiation, 678
Black-body radiation laws, 681
Blind spot, 606
Block and tackle, 43
Bohr, Niels, 170, 690, 693
Bohr atomic theory, 169
Bohr orbit, 691
Bohr postulate, 170, 690
Bohr radius, 170, 691
Boiling point, 252
Boltzmann, L., 275, 310
Boltzmann constant, 278
Bomb, bursting, 124
Bomb calorimeter, 243
Bombsight, 95
Bourdon gauge, 196
Bow wave, 527
Bowl, hemispherical rotating, 104
Boyle's law, 258
  deviations from, 268
Bracket, 46
Brackett series for hydrogen, 692
Bradley, J., 559
Bragg's law, 667
Brewster, Sir David, 672
Brewster's law, 673
Brickwedde, F. G., 692
Brightness of color, 630
British system of units, 89
British thermal unit (Btu), 235

Brown, Robert, 280
Brownian movement, 280
Bulk modulus, 53
Bullet, velocity of, 129
Bunsen (grease-spot) photometer, 623
Buoyancy, 197
  center of, 201

C

Calcite crystal, 674
Calorie, 234
Calorimeter, 239
Camera, 615
Canada balsam, 675
Candle, 620
  new international, 620
  standard, 620
Candle power, 621
Cannon, 131
Capacitance, of isolated conductor, 359
  of sphere, 359
Capacitor, 360
  in circuits, 362
  electrolytic, 362
Capillarity, 214, 218
Carnot, Sadi, 298
Carnot cycle, 296
Carrier wave, 481
Cassegrain telescope, 614
Cathode, 380
Cathode-ray tube, 485
Caustic curve, 574
Cavendish, H., 160
Cavendish experiment, 160
Ceiling, 317
Ceiling balloons, 317
Cell, 381
  Daniell, 383
  dry, 383
  electrolytic, 382
  lead storage, 384
  Weston standard, 384
Center, of buoyancy, 201
  of circular wire, 139
  of gravity, 35, 136
  of mass, 132, 135
  optical, 595
Centigrade scale, 224
Central force, 133
Centrifugal force, 97
Centripetal acceleration, 74, 97
Centripetal force, 97
Centroid, 140
Cgs system, 5, 89
Chain reaction, 724
Charge, conservation of, 323
  dimensions of, 327
  distribution in potassium, 697
  on electron, 326
  electronic, 350
  induced, 330
  surface, density, 339
  units of, 325
Chinook, 320
Chromatic aberration, 598, 602
Chromaticity diagram, 634
Ciliary muscle, 606
Circle of least confusion, 598
Circular motion, 73
Circular wire, center of mass of, 139
Clausius, Rudolf, 291
Cloud chamber, 715
Cochlea, 545
Coercive force, 449
Cohesion, 214
Coil, search, 436
Cold front, 321
Collisions, 128
  ionization by, 492
Color, 630
  addition of, 632
  mixture curves, 633
  primary, 632
  purity of, 630, 635
  saturated, 635
  by subtraction, 636
  unsaturated, 635
Colorimetry, 630
Coma, 598
Combustion, heat of, 242
Commutator, 435
Compass needle, 427
Complementary colors, 635
Components, rectangular, 17
Compound motor, 406
Compressibility, 53, 687
Compression, 50
Compressional wave, 509
Compton, A. H., 716

Compton effect, 555, 716
Concave grating, 666
Concave mirror, 581
Concurrent forces, 27
Condensation, 246
Condenser, 360
  optical, 611
Conductance, 369
  mutual, 476
Conduction, 303, 305
  in electrolytes, 378
  through pipe wall, 308
Conductivity, electric, 367
  in meteorology, 312
  thermal, 306
Conductor, 324
  capacitance of isolated, 359
  field near, 339
  field within, 340
  force between two, 410
Confusion, circle of, 598
Conjugate foci, 582
Conjugate rays, 596
Consequent poles, 421
Conservation, of charge, 323
  of mechanical energy, 114, 117
  of momentum, 123
Consonance, 541
Constriction, flow through, 207
Constructive interference, 631
Contact, angle of, 217
Continuity, of flow, 205
  of state, 270
Continuous emission spectrum, 627
Continuous-flow method, 242
Convection, 303
Converging lens, 593
Convex mirror, 583
Cooling, Newton's law of, 241
Cooling curve, 248
Coplanar forces, 30
Coriolis force, 317
Cornea, 606
Corner reflector, 572
Corpuscular theory of light, 551
Cosmic radiation, 725
Cosmic-ray showers, 725
Coulomb, C. A., 325
Coulomb, the, 327
Coulomb's law, 325
  for dielectric materials, 358
Counter, Geiger, 493
Counter emf, 382, 406
Counterforce, 105
Couple, 35
Critical angle, 569
Critical damping, 183
Critical isothermal, 269
Critical point, 269
Critical pressure, 270
Critical temperature, 270
Critical volume, 270
Crooke's dark space, 494
Cross-wind lift, 96
Cryohydrate, 249
Crystal rectifier, 495
Crystal spectrograph, 667
Crystallographic axis, 674
Curie temperature, 449
Current, density of, 366
  eddy, 465
  effective, 455
  electric, 366
  growth of, in inductances, 439
  loop, magnetic field of, 413
  magnetic field of, 408
  magnetic force on, 400
  magnetizing, 464
  ocean, 303
  rms value of, 456
  three-phase, 466
  torque on, 401
Curvature, of image field, 598, 600
  radius of, 501, 573, 586
Cycle, Carnot, 296
  Otto, 300
  Rankine, 299
  thermodynamic, 294
Cyclone, 318
Cyclotron, 397, 693

D

Dalton, John, 275
Dam, 199
Damped oscillation, 181
Damping, electrodynamic, 436
  of sound, 528
Daniell cell, 383

Davisson, C. J., 666, 703
Davy, Sir Humphry, 233
De Broglie, Louis, 661, 703
De Broglie wavelength, 662
Dead beat motion, 183
Decay constant, 712
Decibel, 542
Declination, magnetic, 427
Deformation, elastic, 51
Degree of freedom, 237, 280
Delta connection, 466
Demodulation, 480
Density, 1, 191
  of electric energy, 364
  of magnetic energy, 452
  of surface charge, 339
  of surface current, 417
Derrick, 48
Destructive interference, 639
Deuterium, 692
Deuteron, 692
Deviation, minimum, 575
Dew point, 316
Diamagnetism, 446
Diameter of molecule, 213
Dielectric constant, 355
Dielectric material, 353
  Gauss's law for, 358
Differential pulley, 44
Differentiation, 62
Diffraction, 517, 556
  by circular obstacle, 654
  of electrons, 666
  Fraunhofer, 651
  Fresnel, 651
    by circular aperture, 653
  grating, 663
  patterns, light, 651
    X-ray, 668
  by single slit, 655–658
  of X rays, 666
Diffuse reflection, 566
Diffusion, 280
Dimensional analysis, 87
Dimensions, physical, 7
Diode, 473
Diopter, 608
Dipole, electric, 335, 353
  magnetic, 421, 427
    moment, 415, 421
Dirac, P. A. M., 203
Direct-vision prism, 578, 617
Disk dynamo, 433
Dispersion, 575
  angular, 577
  normal, 577
Dispersive power, 577, 603
Displacement, 71, 173
  amplitude, 187
  electric, 358
Dissonance, 541
Distance of distinct vision, 607
Distortion, barrel, 602
  of image, 598
  pincushion, 602
Diverging lens, 593
Dominant wavelength, 630
Doppler effect, 525, 705
Double refraction, 674
Double-slit experiment, 640
Double weighing, 40
Drag, 96
Dry cell, 383
Dulong-Petit law, 237
Dulong-Petit method, 230
Dynamometer, 418
Dyne, 82

E

Ear, 545
Earth potential, 344
Eddy currents, 465
Efficiency, of engines, 297
  luminous, 621
  of machines, 119
Efflux, speed of, 209
Einstein, A., 702, 710
Elastic limit, 55
Elastic moduli, 54
Elasticity, 128
  rubberlike, 56
Electric current, 366
Electric dipole, 336, 353
Electric displacement, 358
Electric energy density, 364
Electric field, 333
Electric generator, 434
Electric noise, 496
Electric oscillations, 468

Electric potential, 341
Electric susceptibility, 355
Electrochemical equivalent, 391
Electrodes, 380
Electrodynamics, 323
Electrolysis, 379
  Faraday law of, 381
Electrolytes, conduction in, 378
Electrolytic capacitor, 362
Electrolytic cell, 382
Electromagnetic radiation, 442, 483
Electromagnetic spectrum, 564
Electromagnetic theory of light, 552
Electromagnetic waves, 553, 562
  length of, 562
  origin of, 562
Electromagnetism, 323
  fundamental equations of, 441
Electromotive force (emf), 369
  alternating, 435
  counter, 382, 406
  induced, 429
  Peltier, 386
  Seebeck, 386
  Thompson, 385
Electron, charge on, 326, 350
  diffraction of, 66
  lens, 496
  mass of, 330
  micrograph, 663
  microscope, 496, 661
  multiplier tube, 494
  shells, 694–697
  spin, 448, 702
  valence, 694
  volt, 349
  wave, 703
Electroplating, 380
Electroscope, 324
Electrostatic induction, 331
Electrostatic volt, 342
Electrostatics, 323
Elements, 687
Elevator, 90
Emf (*see* Electromotive force)
Emission, secondary electron, 494
  thermionic, 471
Emissivity of surface, 311, 679, 685
Emittance, radiant, 684
Emittance, spectral, 680, 685
Emittance curve, 680
End correction, 536
Energy, 110, 118
  atomic, 397
  of charged capacitor, 363
  equipartition of, 280
  internal, 237, 264, 290
  ionization, 694, 698
  kinetic, 109, 115
  in magnetic field, 441
  potential, 115
  production in stars, 724
  radiant, 677
  of wave motion, 508
  zero-point, 704
Engines, efficiency of, 297
  internal-combustion, 300
Entrance pupil, 612
Entropy, 292
Equation of state, 262
  derivation of, 276
Equilibrant, 33
Equilibrium, extended body, 32
  force condition for, 28
  of particle, 27
  torque condition for, 32
Equipotential surfaces, 344
Equivalent, electrochemical, 381
Equivalent surface current, 421
Erecting prism, 570
Escape, speed of, 169
Eutectic, 250
Exit pupil, 612
Expanding universe, 707
Expansion, apparent coefficient of, 230
  of gas, 259
  linear, coefficient of, 226, 687
  of liquids, 229
  surface, 229
  volume, 229
  volume coefficient of gas, 261
  of water, 230
  work done by, 252
Exploding shell, 125
Exterior ballistics, 96
External forces, 107, 132
Extraordinary ray, 675
Eye, 606
Eyepiece, 609

F

f number, 615
Fahrenheit scale, 224
Far point, 607
Farad, the, 360
Faraday, M., 340, 429, 430
Faraday, the, 381
Faraday's dark space, 494
Faraday's law of electrolysis, 381
Faraday's law of induction, 430
Ferromagnetism, 448
Field, within conductor, 340
  near conductor surface, 339
  of current element, 408
  of current loop, 412
  electric, motion in, 348
  gravitational, 165
  magnetic, 390
    motion in, 392
  of solenoid, 415
  of straight wire, 409
Filament, emission from, 471
Films, interference from thin, 644
  nonreflecting, 649
Filter, electric, 474
Fine structure of spectral lines, 702
Fizeau, A. H. L., 560
Fletcher, Harvey, 531
Flicker photometer, 624
Flight, curved, 101
Flotation, 201
Flow, continuity of, 205
  streamline, 203
  tube of, 204
Fluids, 190
  incompressible, 206
  sound in, 508
  velocity of waves in, 509
Fluorescence, 701
Fluorescent lamp, 495
Flux, luminous, 619, 626
  normal, 337
  radiant, 619, 626, 678, 684
  total, 624
Flywheel, 143
Focal distance, conjugate, 582
Focal length, 582
Focal line, primary, 585
  secondary, 585
Focal point, 587
Focus, 571
  conjugate, 582
  principal, 582
  virtual, 583
Foley, A. L., 524
Foot candle, 621
Foot pound, 111
Foot poundal, 111
Force, coercive, 450
  lines of, 334
  magnetic, 390, 400
  between two conductors, 410
  units of, 81, 82
Forces, 12
  addition of many, 14, 18
  addition of two, 13
  concurrent, 27
  coplanar, 30
  of friction, 41
  moment of, 30
  noncurrent, 32, 33
  resolution of, 16, 18
  resultant of, 13
  subtraction of, 16, 18
Foucault, J. B. L., 552, 560, 562
Fourier analysis, 503
Fourier theorem, 505
Fps system, 5
Fraunhofer, J., 575
Fraunhofer diffraction, 651
Fraunhofer lines, 576, 628
Free electrons, 330
Free fall, 67
Freezing, 246
Frequency, 75
  fundamental, 531
  of light, 563
  natural, 183
  of resonance, 457
Frequency modulation, 481
Fresnel, A. J., 552, 642
Fresnel diffraction, 651
  by circular aperture, 653
Fresnel double mirror, 642
Fresnel's half-period zones, 651
Friction, 41, 86, 106
  coefficient of, 86
  kinetic, 42
  in oscillation, 181

Friction, rolling, 42
  static, 42
Front (in meteorology), 321
Full-tone interval, 541
Fundamental frequency, 531
Fundamental tone, 533
Fusion, latent heat of, 238

G

Gain of amplifier, 477
Galilean telescope, 613
Galilei, Galileo, 10, 67, 77
Galvanometers, 404
  ballistic, 405
Gamma rays, 564, 711
  ionization by, 716
Gamow, G., 329
Gas, amplification, 490
  constant per mole, 263
  constant per molecule, 278
  critical values for, 269
  expansion of, 259
  kinetic theory of, 275
  liquefaction of, 271
  molar specific heat of, 264, 281
  pressure coefficient of, 261
  specific heats of, 264
  velocity of sound in, 510
  volume coefficient of, 261
Gauss, K. F., 336
Gauss (magnetic unit), 391
Gaussian image point, 598
Gaussian surface, 337
Gauss's law, 336
  for dielectric materials, 358
Gay-Lussac, 275
Gay-Lussac's law, 261
Geiger counter, 493
Generator, electric, 434
Geometrical optics, 566
Germer, L. H., 666, 668, 703
Glancing angle, 667
Glow discharge, 494
Governor, 104
Gram centimeter, 111
Grating, concave, 666
  diffraction, 663
  transmission, 666
Gravitation, universal, 159
Gravitational attraction, 213
Gravitational field, 165
Gravitational potential, 169
Gravitational units of force, 81
Gravity, 84, 85, 114
  center of, 35, 136
Grease-spot photometer, 623
Greek alphabet, 733
Gregorian telescope, 614
Grid, 475
  suppressor, 478
Grimaldi, F. M., 552
Gyration, radius of, 142
Gyroscopic motion, 154
Gyroscopic stability, 154

H

Hair hygrometer, 316
Half-life, 713
Half-period zones, 651
Half-tone interval, 541
Half-tone printing, 637
Hammock, 46
Harmonics, 503
Harrow, 106
Hartley oscillator, 480
Hazen, W. E., 716, 725
Heat, 233
  atomic, law of, 237
  capacity, 235
  of combustion, 242
  of formation, 243
  of fusion, 238
  insulators, 305
  mechanical equivalent of, 235
  molar specific, of gas, 264
  of solution, 243
  specific, 235
  transfer of, 303
  of vaporization, 238, 252
Heat death of the universe, 301
Heat pump, 301
Heaviside layer, 484
Heisenberg, W., 703
Heisenberg uncertainty principle, 704
Helmholtz, H. von, 531
Henry, Joseph, 430
Henry, the, 438
Hertz, Heinrich, 555

Hiero's crown, 197
Hooke, Robert, 50, 552
Hooke's law, 50, 55, 180
Horsepower, 121
Hot body, radiation from, 677
Hot spark, 699
Hue of color, 630
Humidity, 314
  absolute, 315
  atmospheric, 315
  relative, 315
Huygens, Christian, 78, 516, 551
Huygens eyepiece, 609
Huygens' principle, 516, 556
Hydraulic press, 196
Hydrodynamics, 203
Hydrogen, Bohr orbits for, 692
  continuous emission spectrum of, 627
  line spectrum of, 689
  spectral series of, 692
Hydrometer, 198
Hydrostatic paradox, 200
Hydrostatics, 190
Hygrometer, 315
Hyperfine structure of spectral lines, 702
Hyperopia, 607
Hysteresis, 56, 449
Hysteresis loop, 450

I

Iceberg, 202
Ice-pail experiment, 340
Iconoscope, 491
Ideal gas, 258
  bulk modulus of, 259
  equation of state of, 276
Illuminance, 619, 826
Image, 558, 570
  distortion of, 598
  formation, 587
  graphical construction of, 584, 588, 593
  multiple, 572
  perverted, 572
  virtual, 571, 583
Impedance, 185
  electrical, 457
  intrinsic, 555
  mechanical, 186
  phase of, 459
Impulse, 122
  angular, 152
Incidence, plane of, 673
Inclination, 427
Inclined plane, 44
Incompressible fluid, 206
Index of refraction, 562, 577
Indicator diagram, 300
Induced charges, 330
Induced emf, 429
Inductance, iron-cored, 452
Induction, electrostatic, 331
  Faraday's law of, 430
  magnetic, 390, 444
  motor, 466
  mutual, 437
Inductive capacity, specific, 355
Inductive charging, 331
Inductor, 439
  current growth in, 439
Inertia, 80
Infrared radiation, 309, 564
Initial conditions, 176
Insulators, 324, 305
Integration, 64
Intensity, in diffraction, 657
  level (sound), 543
  luminous, 620, 626
  magnetic 424, 444
  of magnetization, 421
  of sound, 528
Interference, constructive, 639
  destructive, 639
  fringes from film, 646
  in thin films, 644
  of waves, 513
Interferometer, Michelson, 643
Interior ballistics, 96
Intermolecular forces, 212
Intermolecular potential, 213
Internal-combustion engines, 300
Internal energy, 237, 264, 290
Internal force, 132
International Commission on Illumination, 619, 633
Inverse-square attraction, 169
Inversion temperature, 387
Ion, 378
Ionization, by alpha and beta rays, 715
  by collision, 492

Ionization, by gamma rays, 716
energy, 694, 698
gauge, 288
potential, 693, 694
Ionized atoms, 697
Ionosphere, 313, 427, 484
Iron-cored inductance, 452
Irradiance, 678, 684
Irrational spectrum, 665
Isothermal expansion, 259
work done in, 260
Isotopes, 394

J

Joule, J. P., 111
Joule, the, 111
Joule-Thomson experiment, 267
Jupiter's satellites, 551

K

K-series X-ray lines, 700
K shell, 695
Kelvin, Lord, 298
Kelvin scale, 225, 262
Kepler, Johann, 163, 551
Kerr effect, 677
Kilocalorie, 235
Kilogram, 82
Kilogram meter, 111
Kilowatt, 121
Kilowatt hour, 121
Kinematics, 61
Kinetic energy, 109, 115, 710
of rotation, 149
Kinetic friction, 42
Kinetic pressure, 206
Kinetic theory of gases, 275
Kirchhoff's laws, 372, 678
Kundt's method, 540

L

L shell, 695
Ladder, 38
Lambert, the, 625
Lambert's cosine law, 625
Lamina, 139
Laplace, P. S. de, 552
Lapse rate, 313
Latent heats, 238
Lattice, 132
Laue, M. von, 666
Laue spot (X-ray diffraction), 668
Lawrence, E. O., 399
Lead storage cell, 384
Lens, 593
aberration in, 597
achromatic, 604
combination of, 596
converging, 593
diverging, 593
electron, 497
negative, 593
positive, 593
rectilinear doublet, 615
telephoto, 597
thick, 596
thin, 593
Zeiss "Tessar," 615
Lenz's law, 431
Leslie, Sir John, 310
Leucippus, 10
Lever, 43, 119
Lift, on airplane wing, 209
cross-wind, 96
Light, 551
electromagnetic theory of, 552
plane-parallel, 568
polarization of, 671
refraction of, 562, 567, 574
through transparent slab, 568
velocity of, 412, 558
Linde, K., 271
Line, of force, 334
Fraunhofer, 576, 628
primary focal, 585
secondary focal, 585
Line spectrum, 628, 575
Liquefaction of gases, 271
Liquids, boiling point of, 252
expansion of, 229
superheated, 273
undercooled, 248
vapor pressure of, 251
Limiting angle of repose, 87
Lodestone, 419
Logarithms, 736
Looming, 589

Loop of wave, 514
Loschmidt number, 284
Loudness, 525, 542
Lumen, 620
Luminance, 624
Luminous efficiency, 621
Luminous flux, 619, 626
Luminous intensity, 620, 622, 626
Lummer-Brodhun photometer, 623
Lux, 621
Lyman series, 692

M

M shell, 696
Machines, 40
  efficiency of, 119
Mach number, 211
McLeod gauge, 288
Magnet, 419
  torque on, 423
Magnetic dipole, 420, 427
Magnetic dipole moment, 415, 421
Magnetic energy, 441
  density, 452
Magnetic equator, 427
Magnetic field, 390, 408
  of current loop, 413
  of solenoid, 415
Magnetic force, 390, 400
Magnetic induction, 390, 444
Magnetic intensity, 424, 444
Magnetic permeability, 445
Magnetic poles of earth, 427
Magnetic storms, 427
Magnetic susceptibility, 445
Magnetism, terrestial, 427
Magnetite, 419
Magnetization, of short specimens, 453
Magnetization current, 464
Magnetization curve, 448
Magnetization intensity, 421, 444
Magnetometer, 425
Magnetostriction, 452, 548
Magnification, 584, 596
Magnifier, simple, 609
Malus' law, 673
Manometer, 194
Mars, 705
Mass, 79
  of atom, 284
Mass, center of, 132
  reduced, 692
  relativistic increase of, 702
  rest, 710
Mass defect, 721
Mass-energy relation, 709, 718
Mass number, 394
Mass spectograph, 395
Maxwell, J. C., 391
Maxwell's distribution of molecular velocities, 275, 442, 553
Maxwell's equations, 442
Mayer, J. R., 265
Mayevski, 182
Mean free path, 283
Mean square, 277
Mean temperature, 314
Mean velocity, 66
Measurements, 2
Mechanical advantage, 41
Mechanical energy, conservation of, 114, 117
Mechanical impedance, 186
Megohm sensitivity, 405
Melde's experiment, 515
Melting, 246
Melting point, 247, 687
Membrane, basilar, 546
  vibration of, 538
Mendelyeev, 686
Meson, 726
Mesotron, 725
Metacenter, 201
Metacentric height, 202
Meteorology, 312
Metric pound, 6
Metric ton, 6
Metric units, 6
Mho, the, 369
  per meter, 369
Michelson, A. A., 560, 562
Michelson interferometer, 643
Microfarad, 360
Micrograph, electron, 663
Micron, 6
Microscope, compound optical, 610, 611
  electron, 496, 661
  limit of resolution of, 660
  resolving power of, 660
  simple optical, 608

Microwaves, 561, 564
Milky Way, 706
Miller, D. C., 532
Millibar, 192
Millikan, R. A., 350
Minimum deviation by prism, 375
Minimum potential energy principle, 213
Mirage, 589
Mirror, concave, 581
  convex, 583
  equation, 582
  plane, 571
  spherical, 581–591
Mixtures, eutectic, 250
  method of, 239
Mks system, 5, 89
Mobility, of charge, 367
  of ions, 378
Modulation, 480
Moduli of elasticity, 53
Molecular diameter, 213
Molecular spectra, 700
Molecular velocities, 278
Momentum, 122
  angular, 152
  conservation of, 123
  of photon, 716
Moment of dipole, 334
  of force, 30
  of inertia, 141
Monkey of pile driver, 130
Monochromatic light, 574
Monochromator, 616
Moon, 84, 165, 561
Moseley, H. G. J., 699
Moseley's formula, 699
Motion, 10
  angular simple harmonic, 180
  on curves, 98
  gyroscopic, 154
  history of, 10
  on incline, 87
  in magnetic field, 392
  natural, 10
  Newton's laws of, 11
  in plane, 70
  simple harmonic, 173–182
  transient, 184
  uniform circular, 73
  violent, 10
Motors, electric, 405
  induction, 466
  squirrel-cage, 467
Mt. Palomar telescope, 614
Multiple images, 572
Multiplier resistor, 403
Multiplier tube, 494
Murphy, G. M., 692
Musical scales, 541
Musical sounds, 525
Mutual conductance, 476
Mutual induction, 437
Myopia, 607

N

N shell, 696
Natural frequency, 183
Natural motion, 10
Natural oscillations, 183
Near point, 607
Nebulae, 706
Negative ions, 698
Negative lens, 593
Neutral position, 167
Neutral temperature, 387
Neutrino, 722
Neutron, 719
  potential energy of, 720
Newton, Sir Isaac, 11, 575, 647
Newton, the, 81
Newton meter, 111
Newtonian telescope, 614
Newton's law of cooling, 241
Newton's laws of motion, 11, 79, 105
Newton's rings, 647–649
Nicol prism, 675
Nodal lines, 538
Node, 514
Noise, electrical, 496
Nonconcurrent forces, 32
  equilibrant of, 33
  resultant of, 33
Nonreflecting films, 649
Normal flux, 337
Normal spectrum, 665
Nuclear fission, 723
Nuclear radius, 709
Nuclear reactor, 724
Nuclear spin, 702

Nucleus, 169, 329
binding energy of, 720
compound, 718
size of, 708
Number, atomic, 686, 694
Nutation, 154

O

Object, virtual, 571, 587
Ocean currents, 303
Oersted, H. C., 408
Oersted, the, 425
Ohm, G. S., 369
Ohm meter, the, 369
Ohmmeter, 404
Ohm's law, 367
Oil-drop experiment, 350
Opera glass, 613
Optic nerve, 606
Optical center, 595
Optical instruments, 506
Optical path, 645
Optical shutter, 677
Optics, geometrical, 566
physical, 566
Ordinary ray, 674
Oscillation, 173
damped, 181
electric, 468
forced, 183
Oscillator, Hartley, 480
tuned-grid, 479
Otto cycle, 300
Overtones, 531

P

Paint, colors of, 636
Parachute, 97
Parallax, 557
Parallel axes theorem, 144
Parallel capacitors, 362
Parallel resistances, 371
Parallelogram law, of forces, 16
of vectors, 22
Paramagnetism, 447
Paraxial rays, 581
Parmenides, 10
Particle in equilibrium, 27
Pascal, B., 194
Paschen, F., 692
Peltier, J. C. A., 386
Peltier emf, 386
Pencil of rays, 558
Pendulum, ballistic, 127
conical, 102
physical, 179
simple, 178
Pentode, 478
Penumbra, 557
Period of rotation, 75
in simple harmonic motion, 173, 502
Periodic chart of elements, 688
Periodic motion, 173
Periodic system of elements, 686
Permeability, magnetic, 445
Permittivity, 356
Perverted image, 571
Pfund series for hydrogen, 692
Phase, 175
angle, 187
change of, in reflection, 518
constant, 175, 502
of impedance, 459
Phosphor, 701
Phosphorescence, 701
Photoelasticity, 677
Photoelectric effect, 490
Photometer, bar, 623
Bunsen grease-spot, 623
flicker, 624
Lummer-Brodhun, 623
shadow, 623
Photometry, 619
Photon, 490, 555, 690
momentum of, 716
Phototube, 490
Photovoltaic cell, 495
Physical constants, 731
Physical dimensions, 7, 88
Physical optics, 566
Physical pendulum, 179
Piezoelectricity, 496
Pile, atomic, 724
Pincushion distortion, 602
Pinhole image, 558
Pipe, closed, 536
open, 536
Pirani gauge, 288

Microwaves, 561, 564
Milky Way, 706
Miller, D. C., 532
Millibar, 192
Millikan, R. A., 350
Minimum deviation by prism, 375
Minimum potential energy principle, 213
Mirage, 589
Mirror, concave, 581
  convex, 583
  equation, 582
  plane, 571
  spherical, 581–591
Mixtures, eutectic, 250
  method of, 239
Mks system, 5, 89
Mobility, of charge, 367
  of ions, 378
Modulation, 480
Moduli of elasticity, 53
Molecular diameter, 213
Molecular spectra, 700
Molecular velocities, 278
Momentum, 122
  angular, 152
  conservation of, 123
  of photon, 716
Moment of dipole, 334
  of force, 30
  of inertia, 141
Monkey of pile driver, 130
Monochromatic light, 574
Monochromator, 616
Moon, 84, 165, 561
Moseley, H. G. J., 699
Moseley's formula, 699
Motion, 10
  angular simple harmonic, 180
  on curves, 98
  gyroscopic, 154
  history of, 10
  on incline, 87
  in magnetic field, 392
  natural, 10
  Newton's laws of, 11
  in plane, 70
  simple harmonic, 173–182
  transient, 184
  uniform circular, 73
  violent, 10
Motors, electric, 405
  induction, 466
  squirrel-cage, 467
Mt. Palomar telescope, 614
Multiple images, 572
Multiplier resistor, 403
Multiplier tube, 494
Murphy, G. M., 692
Musical scales, 541
Musical sounds, 525
Mutual conductance, 476
Mutual induction, 437
Myopia, 607

N

N shell, 696
Natural frequency, 183
Natural motion, 10
Natural oscillations, 183
Near point, 607
Nebulae, 706
Negative ions, 698
Negative lens, 593
Neutral position, 167
Neutral temperature, 387
Neutrino, 722
Neutron, 719
  potential energy of, 720
Newton, Sir Isaac, 11, 575, 647
Newton, the, 81
Newton meter, 111
Newtonian telescope, 614
Newton's law of cooling, 241
Newton's laws of motion, 11, 79, 105
Newton's rings, 647–649
Nicol prism, 675
Nodal lines, 538
Node, 514
Noise, electrical, 496
Nonconcurrent forces, 32
  equilibrant of, 33
  resultant of, 33
Nonreflecting films, 649
Normal flux, 337
Normal spectrum, 665
Nuclear fission, 723
Nuclear radius, 709
Nuclear reactor, 724
Nuclear spin, 702

Nucleus, 169, 329
  binding energy of, 720
  compound, 718
  size of, 708
Number, atomic, 686, 694
Nutation, 154

O

Object, virtual, 571, 587
Ocean currents, 303
Oersted, H. C., 408
Oersted, the, 425
Ohm, G. S., 369
Ohm meter, the, 369
Ohmmeter, 404
Ohm's law, 367
Oil-drop experiment, 350
Opera glass, 613
Optic nerve, 606
Optical center, 595
Optical instruments, 506
Optical path, 645
Optical shutter, 677
Optics, geometrical, 566
  physical, 566
Ordinary ray, 674
Oscillation, 173
  damped, 181
  electric, 468
  forced, 183
Oscillator, Hartley, 480
  tuned-grid, 479
Otto cycle, 300
Overtones, 531

P

Paint, colors of, 636
Parachute, 97
Parallax, 557
Parallel axes theorem, 144
Parallel capacitors, 362
Parallel resistances, 371
Parallelogram law, of forces, 16
  of vectors, 22
Paramagnetism, 447
Paraxial rays, 581
Parmenides, 10
Particle in equilibrium, 27
Pascal, B., 194
Paschen, F., 692
Peltier, J. C. A., 386
Peltier emf, 386
Pencil of rays, 558
Pendulum, ballistic, 127
  conical, 102
  physical, 179
  simple, 178
Pentode, 478
Penumbra, 557
Period of rotation, 75
  in simple harmonic motion, 173, 502
Periodic chart of elements, 688
Periodic motion, 173
Periodic system of elements, 686
Permeability, magnetic, 445
Permittivity, 356
Perverted image, 571
Pfund series for hydrogen, 692
Phase, 175
  angle, 187
  change of, in reflection, 518
  constant, 175, 502
  of impedance, 459
Phosphor, 701
Phosphorescence, 701
Photoelasticity, 677
Photoelectric effect, 490
Photometer, bar, 623
  Bunsen grease-spot, 623
  flicker, 624
  Lummer-Brodhun, 623
  shadow, 623
Photometry, 619
Photon, 490, 555, 690
  momentum of, 716
Phototube, 490
Photovoltaic cell, 495
Physical constants, 731
Physical dimensions, 7, 88
Physical optics, 566
Physical pendulum, 179
Piezoelectricity, 496
Pile, atomic, 724
Pincushion distortion, 602
Pinhole image, 558
Pipe, closed, 536
  open, 536
Pirani gauge, 288

Pitch, 525
Pitot indicator, 208
Plan-position indicator, 487
Planck, Max, 679
Planck's constant of action, 170, 237, 489, 555, 690
Planck's radiation law, 679
Planetary motion, 163
Plane, of incidence, 673
  principal, 596
Plane diffraction grating, 664
Plane mirror, 571
Plane-polarized light, 671
Plate resistance, 476
Plates, pile of, 673
  vibration of, 538
Plumb bob, 92
Plutonium, 724
Point, aplanatic, 600
  far, 607
  near, 607
  principal, 596
Poiseuille, J. L. M., 219
Poisson, S. D., 58
Poisson's ratio, 58
Polar compounds, 696
Polar front, 320
Polariscope, 674
Polarization, electric, 353, 383
  of light, 671
Polarized light, stress analysis by, 676
Polarizer, 674
Polarizing angle, 672
Polaroid film, 675
Pole, consequent, 421
  magnetic, 419
  strength of, 419
Polygon rule, 15, 22
Porro prism, 614
Positive electron, 722
Positive lens, 593
Potential, electric, 341
    of charged spheres, 345
    of earth, 344
Potential difference, 344
Potential energy, 115, 167, 212
  of charged particle, 709
  of neutron near nucleus, 720
  principle of minimum, 213
Potentiometer, 384
Pound, 5, 6
  metric, 6
Poundal, 81
Power, 120–149
  in a-c circuit, 462
  available, 383
  factor, 463
  thermoelectric, 387
Precession, 154
Pressure, 53, 191
  amplitude, 529
  critical, 270
  head, 206
  kinetic, 206
  tube, 208
  wave, 529
Primary of transformer, 464
Primary colors, 632
Principal axis, 581
Principal plane, 596
Principal points, 596
Principal section, 674
Principal series, 694
"Principia mathematica," 11
Printing, half-tone, 637
  three-color, 637
Prism, achromatic, 578
  binocular, 613
  crown-glass, 579
  direct-vision, 578, 617
  erecting, 570
  flint-glass, 579
  Nicol, 675
  Porro, 614
  refraction of light through, 574
  spectrometer, 616
  total-reflecting, 569
Prismatic spectrum, 575
Projectile, 93
  range of, 94
  time of flight of, 94
Projection, 19
Projection lantern, 615
Pulley, 43
  differential, 44
Pump, diffusion, 287
  rotary vacuum, 287
Pupil, 606
  entrance, 612
  exit, 612

Purity of color, 630, 635
Pyrometer, optical, 224
  thermoelectric, 223
Pythagoras, 3

Q

Q factor, 469
Quality of sound, 525, 530
Quantum, of energy, 282
  of radiation, 490, 680
Quantum mechanics, 697, 703
Quantum number, 170
  total, 695
Quantum theory, 237, 282, 555, 680

R

Radar, 486, 561, 564
Radial velocity, 705
Radian, 4
Radiant emittance, 678, 684
Radiant energy, 677, 684
Radiant flux, 619, 626, 678, 684
Radiation, 303, 490
  black-body, 678
  electromagnetic, 483
  exchange, 683
  from hot body, 677
  laws of, 681
  thermal, 677
Radio reception, 483
Radio transmission, 483
Radio waves, 564
Radioactive series, 713
Radioactive tracer, 722
Radioactivity, 711
  artificial, 721
Radiophosphorus, 721
Radiosodium, 721
Radius, of curvature, 573, 581, 586
  of gyration, 142
Rain gauge, 314
Rainbow, 590
Raindrops, 78, 97
Ramsden eyepiece, 609
Range, of alpha particle, 715
  of projectile, 94
  of proton, 715
Rankine, William J., 299
Rankine cycle, 299
Rational spectrum, 665
Ray, 516, 556, 566
  conjugate, 596
  extraordinary, 675
  method, 581
  ordinary, 674
  paraxial, 581
Rayleigh, Lord, 660
Rayleigh-Jeans radiation law, 682
Reactance, 457
Reaction, 105
Recalescence, 449
Rectangular components, 17
Rectification, 473
Rectifier, crystal, 495
Rectilinear doublet lens, 615
Rectilinear propagation, 557
Reduced mass, 692
Reference circle, 178
Reflectance, 630, 678, 684
Reflecting telescope, 614
Reflection, angle of, 566
  change of phase on, 518
  by concave mirror, 581
  by convex mirror, 583
  diffuse, 566
  law of, 518
  polarization by, 672
  through prism, 574
  regular, 566
  of spherical waves, 570
  total, 568
  of waves, 517
Reflector, corner, 572
Refracting angle, 574
Refracting telescope, 611
Refraction, atmospheric, 589
  double, 674
  index of, 562, 577
  of light, 562, 567
  at plane surface, 572
  at spherical surfaces, 585
  of spherical waves, 572
Refrigerators, 298
Regelation, 249
Regular reflection, 566
Relative aperture, 614
Relative humidity, 315
Relative luminosity, 619
Relativistic increase of mass, 702

Relativity, theory of, 85
Remanence, 449
Resistance, 367
  in circuits, 371
  plate, 476
Resistivity, 368
  temperature coefficient of, 369
Resistor, 369
  multiplier, 403
Resolution, of forces, 16, 18
  limit of, 660
  minimum angle of, 660
Resolving power, of microscope, 660
  of telescope, 659
Resonance, 185, 457, 539
  frequency of, 457
Rest energy, 711
Rest mass, 710
Resultant force, 13, 27, 33
Retentivity, 449
Retina, 606
Reverberation time, 547
Reynolds' number, 204
Rigid body, 132, 137
Rock-salt crystal lattice, 667
Rockets, 124
Rod, vibrations of, 537
Römer, Olaf, 559
Roentgen W. K. von, 409
Roentgen ray, 489
Rolling cylinder, 151
Rolling friction, 42
Rolling wheel, 150
Root-mean-square current, 456
Root-mean-square velocity, 278
Rotating vector, 461
Rotation, 140
  analogies to translation, 153
Rowland, H. A., 666
Rowland grating, 666
Rowland ring, 444
Rutherford, E., 329, 717
Rutherford model of atom, 690
Rydberg constant, 689, 692

S

SAE, 220
Sagitta, 573
Satellite, artificial, 165, 171
Satellites of Jupiter, 559
Saturated color, 635
Saturated vapor pressure, 251
Saturation of color, 630
Saturation curve, 270
Scalar quantities, 22
Scale, equally tempered, 542
  just, 542
  musical, 541
Scattering of alpha particles, 708
Sclerotic membrane, 606
Schrödinger, E., 703
Screw, 45
Search coil, 436
Secondary of transformer, 464
Secondary electron emission, 494
Section, principal, 674
Seebeck, T. J., 386
Seebeck emf, 386
Selective absorption, 675
Self-inductance, 439
Semiconductor, 495
Sensitivity, of balance, 39
  curve of eye, 620
  megohm, 405
Series capacitors, 362
Series circuit, 456
Series motor, 406
Series resistances, 371
Shadow, 557
Shadow photometer, 623
Shear modulus, 53
Shells of electrons, 694
Shock wave, 527
Shunt, 402
Shunt motor, 405
Shutter, optical, 677
Side bands, 481
Sigma (sum), 21
Significant figures, 8
Simple harmonic motion, 173
Simple harmonic wave, 501
Slug, 89
Smythe, H. D., 397
Snell's law, 521, 562, 567
Sodium, light from, 563
Solar constant, 684
Solar spectrum, 706
Solar system, 163
Solenoid, magnetic field of, 415

Solid angle, 4
Sound, 523
  intensity, 528
  musical, 525
  quality, 530
  velocity in fluid, 509
  wall, 210
  wave in fluid, 508
Specific gravity, 191
Specific heat, by continuous flow method, 242
  by method of cooling, 241
Specific inductive capacity, 355
Spectra, of alkali atoms, 693
  astronomical, 704
  molecular (band), 700
  X-ray, 699
    (*See also* Spectrum)
Spectral emittance curve, 680, 685
Spectrograph, crystal, 667
  mass, 395, 616
Spectrometer, constant-deviation, 617
  prism, 616
Spectrophotometer, 631
Spectrophotometry, 624
Spectroscope, 616
Spectroscopic binaries, 706
Spectrum, absorption, 628
  of Arcturus, 705
  band, 706
  characteristic X-ray, 667
  continuous emission, 627
  electromagnetic, 564
  irrational, 665
  line, 575, 628
  of Mars, 705
  normal, 665
  prismatic, 575
  rational, 665
  solar, 706
  types of, 627
  of Venus, 705
    (*See also* Spectra)
Speed, curve, 66
  of efflux, 209
  of escape, 169
Sphere, of action, 212
  capacitance of, 358
  gravitation on, 161
  moment of inertia of, 143
Spherical aberration, 573, 584, 598
Spherical drop, 216
Spherical mirror, 581
Spherical shell, 162
Spherical surface, refraction at, 585
Spherical wave, 530
  reflection of, 570
  refraction of, 572
Spin, 155
  electron, 448, 702
  nuclear, 702
Squirrel-cage motor, 467
Standard meter, 6
  in terms of light waves, 644
Standing wave, 501, 514, 533, 536
Stars, energy production in, 724
  twinkling of, 589
Statampere, 366
Statcoulomb, 326
Static friction, 42
Static tube, 208
Statics, 12
Stationary wave, 514
  in pipe, 536
Statvolt, 342
Stefan, Joseph, 310
Stefan-Boltzmann law, 683, 310
Steinmetz relation, 451
Steradian, 4
Stiffness constant, 51
Stokes' law, 220, 701
Stoner, E. C., 693
Strain, 51
  permanent, 55
Stratosphere, 313
Streamline flow, 203
Strength, of couple, 35
  of electric field, 333
Stress, thermal, 51, 228
Stress analysis, 676
String, standing waves in, 533
  vibrations in, 532
Sublimation, 246, 255
Subtraction, of colors, 636
  of forces, 16, 18
Sunshine recorder, 315
Supersonics, 548
Suppressor grid, 478
Surface, equipotential, 344
Surface charge density, 339

Surface current, equivalent magnetic, 421
Surface current density, 417
Surface tension, 215
Susceptibility, electric, 355
  magnetic, 445
Systems of units, 5

T

Teardrop, 210
Telephoto lens, 597
Telescope, astronomical, 612
  Cassegrain, 614
  Galilean, 613
  Gregorian, 614
  McDonald, 614
  Mt. Palomar, 614
  Newtonian, 614
  reflecting, 614
  refracting, 611
  resolving power of, 659
  terrestial, 613
  Yerkes, 614
Television, 485, 491
Temperature, absolute, 224, 260, 298
  coefficient of resistivity, 369
  Curie, 449
  definition of, 222
  gradient, 305
  inversion, 387
  neutral, 387
  scales of, 224
Tensile strength, 55
Terminal velocity, 97, 220
Terrestial magnetism, 427
Terrestial telescope, 613
Thermal capacity, 312
Thermal conductivity, 306
Thermal expansion, 687
Thermal radiation, 677
Thermal stress, 228
Thermionic emission, 471
Thermocouple, 386
Thermodynamics, cycles in, 294
  first law of, 264, 290
  reversible processes in, 293
  second law of, 291
Thermoelectric power, 387
Thermometer, clinical, 231
  constant-volume gas, 223
Thermometer, liquid-in-glass, 223
  maximum, 231
  mercury, 231
  metallic, 223
  minimum, 231
  resistance, 223
  standard gas, 225, 226
Thermopile, 387
Thermoscope, 222
Thomson, Sir J. J., 393, 702
Thomson emf, 385
Three-color printing, 637
Three-phase current, 466
Threshold of audibility, 544
Thrust, 126
Thyratron, 393
Time of flight, 94
Time constant, 376, 440
Ton, metric, 6
Tone, 525
  combination, 539, 542
  difference, 539, 542
  fundamental, 533
  pure, 530
  summation, 542
Toroid, magnetic field of, 416
Torque, 30
  condition for equilibrium, 32
  on current-carrying coil, 401
  on magnet, 423
  as vector, 155
Torricelli, E., 209
Torsion, 57
Total-reflecting prism, 569
Tourmaline, 675
Townsend coefficient, 493
Tractor, 106
Trade winds, 318
Train of waves, 500
Trajectory, 96
Transfer of heat, 303
Transformer, 464
Transients, 184
Translation analogies to rotation, 153
Transmission grating, 666
Transmutation of elements, 717
Transverse wave, 501
Triad, major, 541
  minor, 541
Trichromatic coefficients, 633

Tricolor stimulus, 632
Trigonometric functions, 735
Triode, amplifier, 476
  characteristic curves of, 475
Triple point, 256
Tristimulus coefficients, 632
Tropical calm, 318
Troposphere, 313
Truss, 46
Tube of flow, 204
Tuned-grid oscillator, 479
Tuning fork, 538
Turbulence, 203
Turn ratio, transformer, 464
Turntable, 152
Twist, 57
Tyndall, John, 310

U

Ultraviolet, 564
Ultrasonics, 548
Umbra, 557
Uncertainty principle, 703, 704
Undercooled liquid, 248
Undercooled vapor, 273
Unit line of force, 334
Units, 2, 729
  electrical, table of, 732
  of force, 81
  metric, 6
  systems of, 5
  of work, 111
Universe, expanding, 707
Urey, H., 692

V

V-2 rocket, 126, 523
Vacuum, diffusion pump for, 287
  production of high, 286
  rotary pump for, 287
Valence electron, 694
Van de Graaff generator, 347
Van der Waals' equation, 272
Vapor pressure, 251
Vaporization, latent heat of, 238, 246
Vector, rotating, 461
Vector calculus, 22
Vector difference, 71
Vector differential, 72
Vector quantities, 22
Velocity, 62
  amplitude, 186
  angular, 75
  of approach, 129
  average, 62
  of gas molecules, 278
  instantaneous, 63, 71
  of light, 412, 558
  Maxwellian distribution of, 279
  radial, 705
  relative, 72
  root-mean-square, 279
  of separation, 129
  of sound in gas, 510, 540
  terminal, 97, 220
  of transverse wave in string, 505
  of waves in fluid, 509
Vena contracta, 209
Venturi meter, 207
Venus, 705
Vibrations, of air column, 534
  of bell, 539
  complex, 503
  forced, 533
  free, 533
  of plates and membranes, 538
  of rods, 537
  of strings, 532
Violent motion, 10
Virtual focus, 571, 583
Virtual image, 571, 583
Virtual object, 571, 588
Viscosity, 214
  coefficient of, 219
Visibility, 317
Vision, defects of, 607
  distance of distinct, 607
Vitreous humor, 606
Volt, the, 342
Voltameter, 381
Voltmeter, 402
Volume, atomic, 686
Von Helmholtz, H., 531
Von Laue, M., 666

W

Warm front, 321
Water equivalent, 240

Water, expansion of, 230
Water waves, 512
Watt, James, 121
Watt, the, 121
Watt-hour meter, 418
Wattmeter, 417, 464
Wave, bow, 527
  carrier, 481
  compressional, 509
  diffraction, 517
  electron, 703
  in fluid, 508
  front, 516, 556
  interference of, 512
  longitudinal, 501
  motion, 500
  number, 689
  reflection, 517
  refraction, 520
  shock, 527
  simple harmonic, 501
  spherical harmonic, 530
  standing, 501, 514
  stationary, 514
  transverse, 501
  water, 512
Wave equation, general, 506
Wave train, 500, 556
Wavelength, 563
  of color, dominant, 630, 635
Weather Bureau, 195
Weber, the, 391
Weber-Fechner law, 543
Weighing, double, 40
Weight, 12
Weighted mean, 36
Weston standard cell, 384
Wet-and-dry-bulb hygrometer, 315
Wheatstone bridge, 374, 377
Wheel and axle, 44
Wien's displacement law, 682
Wilson, C. T. R., 715
Wind instruments, 534
Winds, 304
Work, 108, 118
  definition of, 109
  units of, 111
Work function, 471, 490

X

X rays, 489, 564
  characteristic spectrum, 667
  diffraction of, 666
  K-series lines, 700
  Laue spot, 668
  reflection from crystal, 667
  spectra, 699

Y

Y connection, 466
Yaw, 96
Young, Thomas, 552, 556, 639
Young's experiment, 639
Young's modulus 53

Z

Zeiss "Tessar" lens, 615
Zero-point energy, 704

INDEX OF TABLES

A

Acceleration due to gravity, 85
Angles of contact, 218
Atmospheric pressure, 199
Atmospheric temperature, 199
Atom, electron shells of, 695
  shell structure of, 696

B

Beaufort scale of wind velocities, 314
Boiling points, 254
Bulk modulus, 54

C

Capacity, specific inductive, 357
Combustion, heat of, 243, 244
Conductivity, electric, 369
  thermal, 306
Critical densities of gases, 270
Critical pressures of gases, 270
Critical temperatures of gases, 270

D

Densities, 191
Dispersive power, 578

E

Electric moments of molecules, 355
Electrical oscillations, forced, compared to mechanical, 457
Electron shells in atom, 695
Expansion, linear, coefficients of, 227
  of liquids, coefficients of, 231

F

Force, intermolecular and gravitational, 214
  units of, 82
Formation, heats of, 244
Fraunhofer lines, 576
Free fall, kinematics of, 61
Friction, coefficients of, 86
Fusion, heats of, 247

G

Gases, critical densities of, 270
  critical pressures of, 270
  critical temperatures of, 270
  ratio of specific heats, 282
Greek alphabet, 733

H

Heats, of combustion, 243, 244
  of formation, 244
  of fusion, 247
  of vaporization, 254
Humidity, relative, 316

I

Inertia, moments of, 144
Intensity levels of sound, 543
Ionization energy, 698
Isotopes, 395
  mass of, 396

L

Liquids, coefficient of expansion, 231
Logarithms, 736

M

Major scale, 541
Mathematical formulas, 3, 734
Mechanical oscillations, forced, compared to electrical, 457
Melting points, 247
Metric units, 6
Moments of inertia, 144
Moments of molecules, electric, 355
Motion, rotation and translation, 154

P

Periodic table of atoms, 688
Permeability, 447
Permittivity, 357
Physical constants, 731
Power, units of, 121
Pressure, atmospheric, 199

R

Radioactive series, 714
Ratio of specific heats, gases, 282
Refractive index, 578
Resistivity, electric, 369
temperature coefficient of, 369
Rotation, comparison with translation, 153, 154

S

Shear modulus, 54
Solar system, 164
Sound, absorption coefficients, 547
intensity level, 543
Specific heat, 236
of gases, 282
at atmospheric pressure, 266
Specific inductive capacity, 357
Spectral lines, Fraunhofer, 576
Surface tension, 216
Susceptibility, electric, 357
magnetic, 447

T

Temperature, atmospheric, 199
coefficient of electrical resistivity, 369
scales, 226
Thermal conductivities, 306
Thermoelectric constants, 388
Translation, comparison with rotation, 153, 154
Trigonometric functions, 735

U

Units, 729
electrical, 732
of fps, cgs, mks systems, 5

V

V-2 rocket, 127
Vapor pressure of water, saturated, 251
Vaporization, heats of, 254
Viscosity, coefficient of, 220

W

Wind velocities in Beaufort scale, 314
Work, units of, 112

Y

Young's modulus, 54