MARTIN C. MARTIN
Clarkson College of Technology
Potsdam, New York

and

CHARLES A. HEWETT
Rochester Institute of Technology
Rochester, New York

ELEMENTS OF CLASSICAL PHYSICS

PERGAMON PRESS INC.

New York · Toronto
Oxford · Sydney · Braunschweig

PERGAMON PRESS INC.
Maxwell House, Fairview Park, Elmsford, N.Y. 10523

PERGAMON OF CANADA LTD.
207 Queen's Quay West, Toronto 117, Ontario

PERGAMON PRESS LTD.
Headington Hill Hall, Oxford

PERGAMON PRESS (AUST.) PTY. LTD.
Rushcutters Bay, Sydney, N.S.W.

PERGAMON GmbH
Burgplatz 1, Braunschweig

Library of Congress Cataloging in Publication Data

Martin, Martin C
Elements of classical physics.

1. Physics. I. Hewett, Charles A., joint author.
II. Title.
QC21.2.M38 1973 530.1'32 73-3450
ISBN 0-08-017098-6

Printed in the United States of America

The Authors

Martin C. Martin (Ph.D. University of Alberta, Edmonton, Alberta, Canada) is Associate Professor of Physics at Clarkson College of Technology. His teaching efforts have been in the area of classical physics at the undergraduate and graduate levels and modern physics at the undergraduate level. He has carried on research in the field of solid state, specifically on the effect of elastic and plastic deformation on the physical properties of metals, and the effect of heat treatment on the physical properties of metal alloys.

Charles A. Hewett (Ph.D. University of Missouri-Columbia) is Professor of Physics at the Rochester Institute of Technology. His primary teaching efforts have been in the areas of classical and modern physics for engineering and science students. Currently he is working with Dr. Hrishikesh Banerjee to develop studies in direct energy conversion and in the physics of semiconductor electronic devices in response to requests of students and faculty members of the College of Engineering. Collaboration with Dr. James Karpick, a local industrial physicist, is expected to result in a program of independent study and undergraduate research in fluorescence and photoluminescence of solids that will meet the needs of gifted high school students as well as the more usual group of physics majors.

Preface

This book is being published at a time when many of the nation's leading physicists already have prepared textbooks which implement in an outstanding fashion new approaches developed for the teaching of elementary physics to today's students. The appearance of still another textbook should therefore be accompanied by a presentation of the ideas behind the structure and content of the book.

It has been our experience that students retain an uneven coverage of the various areas of physics from their high school courses. Areas such as elementary heat and light are, in general, more easily understood and remembered than mechanics, thermodynamics, sound, and electricity and magnetism, which necessarily involve a greater degree of abstraction in their presentation as well as greater mathematical sophistication if quantitative discussions are desired.

With very few exceptions, the courses in general physics that have been developed for presentation to entering freshmen begin with mechanics, which represents a relatively serious intellectual challenge for the average student, with or without the use of calculus. This is then followed by heat, light, and sound, which are in turn followed by electricity and magnetism. As a result, the students are confronted by a course that is rather frustrating in the unevenness of its demands upon their understanding. If, in addition, calculus is used from the outset, the average student is initially discouraged, not only by the relatively foreign quantitative concepts of mechanics, but also by the presentation of these concepts in a mathematical "language" that is equally foreign. As a result, many students succumb to these demands and terminate their study of physics or engineering with the mistaken impression that an understanding of the physical world is beyond their grasp.

With these considerations in mind, we have written a text which: (a) begins with the material most readily understood with a minimal mathematical framework, thus providing a smooth transition from high school; (b) progresses as uniformly as possible to areas of increasing conceptual difficulty; (c) introduces at appropriate places the necessary mathematical concepts at a time when they would already have been presented in a typical concurrent mathematical course, and (d) stresses throughout the course the physical concepts and the manner in which these concepts can be used to provide quantitative understanding in a wide variety of specific situations. Many of these are carefully discussed as examples, and the remainder are presented in the problem sections as suitable tests of understanding.

In writing the text, we presume the student has at least a qualitative understanding of the meaning of the terms force, pressure, work, and energy.

After a section devoted to the discussion of dimensions and units, the areas presented are heat, light, mechanics, thermodynamics, sound, and electricity and magnetism, respectively. It is our opinion that this sequence of topics is best suited for achieving the goals outlined above. However, those desiring a more conventional course sequence could with very little difficulty begin with mechanics (Chapter 14), followed by the remaining topics in their usual order. In addition, Chapter 15 on special relativity may be omitted without prejudice to subsequent chapters. No effort has been made to include an extensive presentation of modern physics. There are two reasons for this: first, the material covered represents a thorough coverage of classical physics which is a necessary prerequisite to the proper discussion of modern physics, and second, the mathematical preparation required for modern physics in our opinion necessarily makes it a second year subject in contrast to the contents of this text. It might also be noted that excellent texts covering the subject are available.

We have emphasized the rationalized MKS units throughout the text. We have also made use of some of the various other systems of units in common use today. We introduce the necessary conversions for these systems in an appendix. Every system of units has its proponents and opponents (who are equally convinced of the correctness of their point of view); our choice simply represents a personal preference.

The authors of this book owe much to many people: our teachers; our colleagues; but, most important, our students, whose desire to understand physics has prompted us to try to make their pathway to knowledge as natural as our abilities permit.

Introduction:
Dimensions and Units

In the study of physics, the dimensions and units to be encountered must be understood. A student must remember that one can only equate the same kinds of quantities. A test on the correctness of an equation can be obtained by checking whether or not the dimensions on one side of the equation are the same as those on the other side.

DIMENSIONS

A dimension may be defined as a name describing certain physical quantities. Therefore, a large number of dimensions are possible. This number can be reduced by the fact that certain descriptions can be expressed in terms of other more basic descriptions (dimensions). For example, length, area, and volume are dimensions, but area can be measured as a length squared and volume as a length cubed. Therefore, the dimensions of area and volume can be stated in terms of the more fundamental dimension of length.

In physics, there are five fundamental dimensions: length, mass, time, temperature, and electric charge, which we shall denote by $[l]$, $[m]$, $[t]$, $[T]$, and $[q]$, respectively. We have already shown how area and volume are expressed in terms of length. Now let us consider a few more physical quantities which should be familiar from high school physics.

$$\begin{aligned}
\text{Velocity} &= \text{length} \div \text{time} = [lt^{-1}] \\
\text{Acceleration} &= \text{velocity} \div \text{time} = [lt^{-2}] \\
\text{Force} &= \text{mass} \times \text{acceleration} = [mlt^{-2}] \\
\text{Pressure} &= \text{force} \div \text{area} = [ml^{-1}t^{-2}] \\
\text{Work} &= \text{force} \times \text{length} = [ml^{2}t^{-2}] \\
\text{Power} &= \text{work} \div \text{time} = [ml^{2}t^{-3}] \\
\text{Density} &= \text{mass} \div \text{volume} = [ml^{-3}] \\
\text{Current} &= \text{charge} \div \text{time} = [qt^{-1}]
\end{aligned}$$

UNITS

A unit may be defined as a particular amount of the dimension or quantity to be measured. Three different systems are used today in science and engineering. They are the meter-kilogram-second (MKS) system, the centimeter-gram-second (cgs) system, and the foot-slug-second (British) system. The rationalized MKS and cgs systems are used universally in scientific work, with the MKS actually superseding the two other systems.

We shall use the rationalized MKS system throughout most of the book. Much of the literature is still written in the other two systems of units, so we feel that students should eventually become familiar with the three systems of units. In Table 1 are the MKS units for the physical quantities given in Section 1. For comparison purposes, the equivalent cgs unit is also indicated.

Table 1 MKS Units for Some Physical Quantities.

Physical Quantity	MKS Unit	cgs Unit
length	meter	centimeter
mass	kilogram	gram
velocity	meter/second	centimeter/second
acceleration	meter/second^2	centimeter/second^2
force	newton	dyne
pressure	newton/meter^2	dyne/centimeter^2
work or energy	joule	erg
power	watt	erg/second
density	kilogram/meter^3	gram/centimeter^3
charge	coulomb	statcoulomb
current	ampere	statampere

The student should be familiar with the above physical quantities and the MKS units associated with them, even though a detailed discussion of them will not appear until later in the book.

Table 2 then gives some prefixes used to represent divisions and multiples of metric quantities.

Table 2 Prefixes Used for Divisions and Multiples of Metric Quantities.

deci-	10^{-1}	deca-	10^{1}
centi-	10^{-2}	hecto-	10^{2}
milli-	10^{-3}	kilo-	10^{3}
micro-	10^{-6}	mega-	10^{6}
nano-	10^{-9}	giga-	10^{9}
pico-	10^{-12}	tera-	10^{12}

1 Temperature and Thermometry

Consistent with our intent to progress by degrees from the easiest to the most difficult topics, we will begin by discussing various aspects of thermal energy, heat, and temperature.

1-1 THERMAL ENERGY, HEAT, AND TEMPERATURE

The concept of temperature is fundamental to the study of heat. The idea of temperature developed from man's sense of hotness and coldness. Therefore, a body's temperature represents its degree of hotness or coldness. In order to define temperature in a more concrete way than the above, we will first define thermal energy and heat.

We shall define thermal energy as follows:

> Thermal energy is the energy of the atoms in a substance because of their motion on a microscopic scale.

We shall later see in our discussion of kinetic theory that the average motional or kinetic energy of atoms is directly proportional to the temperature.

We shall now define heat as follows:

> Heat is thermal energy transferred between two or more material substances or from one portion of the substance to another on a macroscopic scale.

This means that a hot body can give up some of its thermal energy and therefore affect a neighboring body. The quantity of thermal energy it gives up depends upon the nature and condition of the neighboring body and upon the medium separating the two bodies.

Now we are able to give a somewhat better definition of temperature; we will define it as follows:

> The temperature of a body is a measure of its ability to transfer heat to other bodies.

A body is at the same temperature as another body if there is no net flow of heat from one to the other when they are placed in contact or separated only by a conducting wall. We can also say that a body X is at a higher temperature than a body Y if heat flows from X to Y when they are placed in contact or separated only by a conducting wall.

The above definition of temperature gives us a way of defining absolute zero, which is the lower limit of temperature. That is, no body can be colder or have less thermal energy than a body at absolute zero. Therefore, we can define absolute zero as follows:

> Absolute zero is the temperature of a body that is not capable of transferring any thermal energy to another body.

There is a criticism of the above definitions on the basis that we do not observe the flow of heat. All we can observe is that the cold body gets hotter and the hot body gets cooler. However, many physical processes occur especially in atomic and nuclear physics that we cannot observe directly, so this is not a critical problem. Other definitions of temperature will be given later on in the book which remove this criticism. However, we feel that at the beginning it is good to have a definition of temperature even though it may be subject to later slight modifications.

1-2 TEMPERATURE MEASUREMENT

We shall now discuss thermometry, which is the science of measuring temperatures. A number of physical properties of substances change with temperature. Temperature measuring devices or thermometers make use of these properties. Some of these properties which change with temperature and can be used for thermometric purposes are:

1. Volume of liquids.
2. Length of solids.
3. Volume of gases.
4. Pressure of gases.
5. Pressure of saturated vapors.
6. Electrical resistance of metals.
7. Thermoelectric currents.
8. Color of radiated light.
9. Total radiation intensity.
10. Magnetic susceptibility of paramagnetic salts.

The construction of a useful thermometer requires a choice of some substance whose thermometric properties indicate changes in temperature and a choice of the design of the thermometer. We then require a thermometer scale, which necessitates a choice of one or more fixed temperature points and a selection of numbers to be associated with each division.

1-3 CENTIGRADE AND FAHRENHEIT TEMPERATURE SCALES

The definitions of the Centigrade and Fahrenheit scales involve in each case the same two distinct fixed thermal points. The lower fixed point is the temperature of the melting point of pure ice (ice point), which is the temperature at which pure ice coexists with air saturated water at one atmosphere pressure. The upper fixed point is the temperature of the boiling point of pure water (steam point), which is the temperature of equilibrium between pure water and pure steam at one atmosphere pressure. On the Centigrade scale, the ice point is taken as 0 degrees and the steam point as 100 degrees. On the Fahrenheit scale, the ice point is taken as 32 degrees, the steam point as 212 degrees. There are 100 degrees between the ice and steam points on the Centigrade scale, while there are 180 degrees between the ice and steam points on the Fahrenheit scale. We leave it to the student as an exercise to show that the conversion from the Centigrade scale to the Fahrenheit scale or vice versa is given by the equations

$$t°\mathrm{C} = \frac{5}{9}(t°\mathrm{F} - 32),$$

$$t°\mathrm{F} = \frac{9}{5}t°\mathrm{C} + 32. \qquad (1\text{-}1)$$

Note that we are using (t) to indicate Centigrade and Fahrenheit temperatures. Later in the text this same symbol will be used to denote time. The double usage of this symbol should not lead to ambiguity if dimensional analysis is used.

Example 1. At what temperatures do the Centigrade and Fahrenheit scales coincide?

SOLUTION

This means the $t°\mathrm{C} = t°\mathrm{F}$. Substitute $t°\mathrm{C} = t°\mathrm{F}$ in Eq. (1-1).

$$t°\mathrm{F} = \frac{9}{5}t°\mathrm{F} + 32$$

or

$$\frac{4}{5}t°\mathrm{F} = -32.$$

Therefore,

$$t°\mathrm{F} = \frac{5(-32)}{4} = -40°,$$

since

$$t°C = t°F$$

$$t°C = -40°.$$

1-4 GENERAL DEFINITION OF TEMPERATURE

Let us consider a general definition of temperature based on the thermometric properties of a substance. Let X_0 be some thermometric property as, for example, length of a solid or height of a liquid in a capillary tube measured at 0°C, and X_{100} and X_t the thermometric property at 100°C and some unknown temperature t°C. The size of a unit interval or degree is $X_{100} - X_0/100$, and the size of the interval from 0°C to t°C is $X_t - X_0$. By definition, t°C is the number of degrees in the interval $X_t - X_0$, and is given by the equation

$$t°C = \frac{X_t - X_0}{\frac{X_{100} - X_0}{100}}$$

or

$$t°C = \left(\frac{X_t - X_0}{X_{100} - X_0}\right)100. \qquad (1\text{-}2)$$

Various thermometric properties can be used. For example, let us consider a mercury in glass thermometer, one with which we are all familiar, and a constant volume gas thermometer, which we will discuss later. If the length of the mercury column is substituted in Eq. (1-2) for the thermometric property X, we have

$$t°C = \left(\frac{l_t - l_0}{l_{100} - l_0}\right)100. \qquad (1\text{-}3)$$

If the gas pressure of the gas in the gas thermometer is substituted in Eq. (1-2) for X, we have

$$t°C = \left(\frac{p_t - p_0}{p_{100} - p_0}\right)100. \qquad (1\text{-}4)$$

A word of caution is necessary here. Even though both the above thermometers read the same at 0°C and 100°C, they do not necessarily read the same at in-between temperatures. There is no reason to expect them to be the same since the temperature as defined by Eq. (1-2) depends upon the thermometric property of the substance. Even though two substances like mercury and a particular gas might have the same thermometric property, they might be contained in glass which has different thermometric properties, so the properties relative to the glass would be different. Experiments have shown that if we compare scales of temperatures based on different liquids or gases, or even on the same liquid or gas in different kinds of glass, the scale reads slightly differently. However, in the range between 0°C and 100°C they generally agree to within a few tenths of a degree.

1-5 KELVIN AND RANKINE TEMPERATURE SCALES

We shall see later that there is a limit to the lowest temperature that can ever be attained. This lowest temperature is -273.15°C; it corresponds to absolute zero which we previously defined. The absolute temperature scale has absolute zero as its zero point. The absolute scale with degree intervals equal to those on the Centigrade scale is called the Kelvin or absolute scale, and the absolute scale with degree intervals equal to those on the Fahrenheit scale is called the Rankine scale. The temperature of absolute zero on the Fahrenheit scale is -459.67°F. For most temperature measurements, it is only necessary to have three figure accuracy so as to reduce memory work. We shall in the calculation in this text relate the Kelvin and Centigrade, and the Rankine and Fahrenheit scales by the following equations:

$$T°K = t°C + 273, \qquad (1\text{-}5)$$

and

$$T°R = t°F + 460. \qquad (1\text{-}6)$$

Both the ice point and the steam point are difficult to measure with good accuracy. The main difficulty is that when ice melts it becomes surrounded with pure water and causes poor contact between the ice and the air saturated water. This prevents the ice point from being accurately measured. Since the boiling point of water is sensitive to small changes of pressure, the value of the steam point can be in error unless the pressure is kept very constant.

In 1854, Kelvin proposed a temperature scale known as the thermodynamic scale (discussed in Chapter 26), which is independent of the properties of any substance. He pointed out that the scale could be determined by the use of a single fixed point. At the Tenth General Conference of Weights and Measures in 1954, it was decided to choose this fixed point as the temperature and pressure at which ice, water, and water vapor coexist in equilibrium (the triple point of water), and to assign it the value of 273.16°K. This point can be measured with greater accuracy than the ice point and the steam point.

1-6 MERCURY THERMOMETER

The mercury thermometer is the most frequently used thermometer in laboratories because of its simplicity, its fairly large temperature range (−39°C to 357°C), and its quick response to slight temperature changes. It is made by first blowing a bulb at one end of a fine bore glass tube. Mercury is poured into the tube, heated to drive out any air, and the other end of the tube is sealed. When the bulb is brought into contact with a body, the mercury either expands or contracts relative to the glass tube depending on whether the body is hotter or colder than the original surroundings of the bulb.

The chief errors in a mercury thermometer are caused by a non-uniform bore in the tube and imperfections of the glass bulb. However, if these and other smaller errors are corrected, the mercury thermometer is very accurate. A good mercury thermometer is only surpassed in accuracy by a platinum resistance thermometer in the range 0°C to 200°C.

1-7 GAS THERMOMETERS

Gas thermometers are usually not used in laboratories because they are somewhat elaborate and cumbersome. They are used generally in standardizing other more simple thermometers. For a given temperature rise, gases expand about 20 times as much as mercury and about 120 times as much as the glass in the thermometer. This reduces error due to uneven changes in the volume of the glass. Gas thermometers have a temperature range of −269°C to 1600°C.

There are two types of gas thermometers —namely, constant volume and constant pressure thermometers. With the first type, the temperature is measured by the change in pressure of the gas with temperature. With the second type, the temperature is measured by the change in volume of the gas with temperature. Constant volume thermometers are more satisfactory because it is easier to measure accurately a change of pressure than a change of volume. Figure 1-1 is a simple form of a constant volume gas thermometer.

The volume of the gas in the bulb is kept constant by raising or lowering tube C so that the top of the mercury is always kept at point B. The pressure is measured by measuring the difference between the heights h of the mercury surfaces in the tube. The pressure of the gas in the bulb is equal to the atmospheric pressure plus the pressure due to the height h of the mercury in the tube. If we let p_0 be the pressure when the bulb is surrounded by melting ice, p_{100}—the pressure when the bulb is surrounded by steam, and p_t—the pressure when the bulb is surrounded by the substance whose temperature is to be measured, the temperature of the substance is given by Eq. (1-4), namely,

$$t°\text{C} = \left(\frac{p_t - p_0}{p_{100} - p_0}\right) 100. \qquad (1\text{-}7)$$

Hydrogen is the gas usually used, except for extreme low temperatures where helium is used and extreme high temperatures where nitrogen is used.

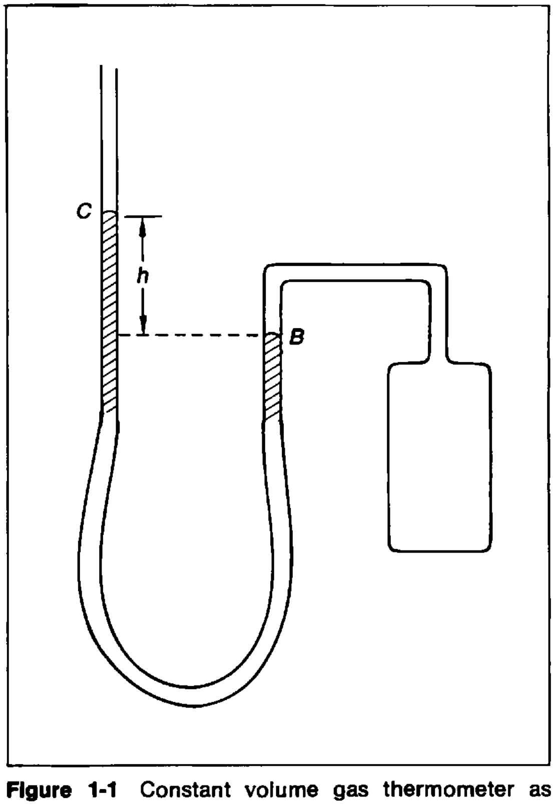


Figure 1-1 Constant volume gas thermometer as described in text.

Helium is used at very low temperatures because it does not liquefy until a temperature of about −269°C is reached. Nitrogen is used at high temperatures because both hydrogen and helium diffuse through the bulb at very high temperatures.

1-8 THE PLATINUM RESISTANCE THERMOMETER

It has been previously stated that the electrical resistance of metals changes with temperature. It is best to use a noble metal for a thermometer so that there is no oxidation. In 1887, Callendar investigated how the resistance of platinum varied with temperature. He showed that the change in electrical resistance of platinum could be used for a very accurate scale of temperatures over a wide temperature range. Today, it is regarded as the most accurate thermometer in the range −190°C to 660°C.

If we define the resistance temperature scale by applying Eq. (1-2), then

$$t\,^{\circ}\mathrm{C} = \left(\frac{R_t - R_0}{R_{100} - R_0}\right) 100, \qquad (1\text{-}8)$$

where R_0, R_{100}, and R_t are the resistances of the platinum wire at the ice point, steam point, and the temperature t, respectively. Callendar also showed experimentally that the resistance R_t of a platinum wire at temperature t is given by the equation

$$R_t = R_0(1 + at + bt^2), \qquad (1\text{-}9)$$

where R_0 is the resistance at the ice point, and a and b are constants. In order to determine the constants a and b, three fixed points are necessary. These fixed points have been chosen as the ice point, the steam point, and the sulphur point which is the boiling point of sulphur (444.6°C) at atmospheric pressure.

Other resistance thermometers of cheaper metals (for example, copper) are sometimes used where extreme accuracy is not necessary. Also, some semi-conductors whose resistance decreases with temperature increase are used. They are known as thermistors; they have a very large decrease in resistance with increase in temperature and have a greater sensitivity than a platinum thermometer, but they are less accurate.

1-9 THERMOCOUPLES

Thermocouples are based on the principle that electric currents are set up in a closed circuit consisting of two dissimilar metals in contact with one another when the junctions are at different temperatures (Chapter 34). The reference junction is usually put in melting ice and the other junction in contact with the material whose temperature is to be measured. The current (or voltage across a circuit element) is then related to the temperature by appropriate calibrations.

The most common thermocouples are constructed of iron and constantan (an alloy of copper and nickel), copper and constantan, and platinum and an alloy of platinum-rhodium. Thermocouples have the advantage that they can be used for measuring temperatures at a particular point in an experimental system.

PROBLEMS

1. In a certain city the greatest temperature variation recorded during a 24 hour period was 45°F. This corresponds to how many degrees C?
2. (a) A Centigrade degree is what fraction of a Fahrenheit degree?
   (b) The conversion from the Centigrade scale to the Fahrenheit scale is given by the equation
   $$t\,^{\circ}\mathrm{F} = \frac{9}{5}t\,^{\circ}\mathrm{C} + 32.$$
   Justify this relation.
3. The reading of a thermometer for the temperature of a room is 77°F. What is the reading on the Centigrade scale?
4. (a) What Centigrade temperature corresponds to 233° Kelvin?
   (b) What Fahrenheit temperature corresponds to 77° Kelvin?
5. At what temperature will the reading of a Fahrenheit thermometer be three times that of a Centigrade thermometer?

6. (a) The temperature of liquid nitrogen is about −196°C. What is its temperature on the Kelvin scale? on the Rankine scale?
   (b) What is the difference between 0°C and 180°F in degrees C? in degrees F?
7. A thermometer was graduated to read 0°C at the boiling point of water, and 155° at the freezing point. Find the reading on the Fahrenheit and Centigrade scale corresponding to 80° on this thermometer.
8. The resistance of a platinum resistance thermometer is found to be 10.000 ohms at the ice point, 13.855 ohms at the steam point, and 25.261 ohms at the sulphur point. Find the constants a and b in Eq. (1-9), and plot R against t in the range 0° to 600°C.

2 Quantity of Heat and Calorimetry

2-1 QUANTITY OF HEAT

When a hot and a cold body are brought into contact, there is a transfer of heat from the hot body to the cold body until they come to the same temperature or reach a state that we term thermal equilibrium. The transference of heat from a hot body to a cold body is analogous to the flow of water from a high level to a low level.

The increase in temperature of a body when heated depends on its mass and the material of which it is composed. For example, to bring a large container of water to the boiling point with a laboratory burner requires a long time. The final temperature may be relatively low, even though it receives a large quantity of heat. If a wire is held in a burner flame, it comes to a very high temperature in a very short time even though it receives only a small quantity of heat. Therefore, the final temperature reached by two bodies in contact initially at different temperatures depends on their masses, the materials of which they are composed, and their initial temperatures.

The net result of heat transfer to or from bodies until thermal equilibrium is reached is given by the relation:

heat gained by cold bodies = heat lost by hot bodies.

In order to make use of this relation, we must define a unit for measuring quantities of heat or thermal energy. The unit we use depends on the system of units we adopt. In the metric system of units, the unit of heat or thermal energy is either the calorie or kilocalorie. If we use centimeter-gram-second (cgs) units, the unit of heat is the calorie; while if we use meter-kilogram-second (MKS) units, the unit of heat is the kilocalorie. The calorie is the quantity of heat required to raise the temperature of one gram of water through one centigrade degree, while the kilocalorie is the quantity of heat required to raise the temperature of one kilogram of water through one centigrade degree.

The above definition of the heat unit depends slightly on the location of the one degree interval. It has been agreed to choose this interval between 14.5°C and 15.5°C.

2-2 SPECIFIC HEAT AND HEAT CAPACITY

The specific heat of a body is the heat required to raise the temperature of a unit mass of the body one degree.

In the MKS system of units, it is the heat required to raise the temperature of a one kilogram mass of the body one degree centigrade. From this definition and the definition of the kilocalorie, we see that the specific heat of water is 1 kilocalorie per kilogram per degree centigrade. Although we are concerned here with the MKS system of units, it seems noteworthy to mention that the numerical value of the specific heats is the same in the cgs system of units.

The quantity of heat necessary to raise the temperature of a body of mass m and specific heat c by Δt degrees is given by

$$Q = mc\Delta t. \quad (2\text{-}1)$$

The specific heat of a substance is roughly constant at ordinary temperatures provided the temperature interval is not too great. As the temperature is lowered, the specific heats of all substances show a decrease. The specific heat of a substance is changed by (1) change of state such as from a solid to a liquid, liquid to a vapor, and vice versa, (2) presence of impurities, and (3) change in temperature.

It is convenient, as we shall see in calorimetry, to define a quantity called heat capacity.

> The heat capacity of a body is the quantity of heat required to raise the temperature of the body one degree. It is the product of the mass of the body and its specific heat.

In the MKS system, the unit of heat capacity is kilocalories per degree centigrade.

2-3 CALORIMETRY

Calorimetry refers to the laboratory science of making measurements of quantities of heat. In making measurements of quantities of heat, a container called a calorimeter is used. The ordinary calorimeter is a vessel placed within another larger vessel, with the two vessels insulated from one another. In this way, exchange of heat to the surroundings is minimized. The inner vessel of the calorimeter must be provided with a stirrer so as to keep its contents at a uniform temperature.

The fact that when two bodies at different temperatures are placed in contact, heat is transferred from the hotter one to the cooler one until their temperatures are equal, gives us one of the most convenient methods of determining specific heat, known as the *method of mixtures.*

2-4 METHOD OF MIXTURES

(a) Determination of the Specific Heat of a Solid

Let a solid of mass m, temperature t, of unknown specific heat c_x be immersed in a mass m_1 of water at temperature t_1 less than t contained in a calorimeter of mass m_2, known specific heat c, and also at temperature t_1. Let t_2 be the final temperature after thermal equilibrium is reached.

Heat lost by the solid = $mc_x(t - t_2)$.
Heat gained by the water = $m_1(t_2 - t_1)$.
Heat gained by the calorimeter = $m_2c(t_2 - t_1)$.

Heat lost by the solid = heat gained by the water + heat gained by the calorimeter. Therefore,

$$mc_x(t - t_2) = m_1(t_2 - t_1) + m_2c(t_2 - t_1), \quad (2\text{-}2)$$

from which c_x can be determined.

(b) Determination of the Specific Heat of a Liquid

The specific heat of a liquid may be determined in the same way by the method of mixtures. If a solid of known mass, specific heat, and temperature is immersed in a known mass of the liquid contained in a calorimeter at a known temperature, the specific heat of the liquid can be calculated.

Example 1. A 0.10 kg calorimeter of specific heat 0.090 kcal/kg°C contains 0.15 kg of a liquid at 20°C. Into this container is placed a 0.20 kg block of copper of specific heat 0.095 kcal/kg°C and temperature 100°C. The final temperature is 40°C. What is the specific heat of the liquid?

SOLUTION

Heat gained by liquid + heat gained by calorimeter = heat lost by copper block.

$$(0.15\ \text{kg})(c_x)(40-20)^\circ\text{C} + (0.10\ \text{kg})(0.090\ \text{kcal/kg}^\circ\text{C})(40-20)^\circ\text{C} = (0.20\ \text{kg})(0.095\ \text{kcal/kg}^\circ\text{C})(100-40)^\circ\text{C}$$

$$(c_x)(3.0\ \text{kg}^\circ\text{C}) + (0.18\ \text{kcal}) = 1.14\ \text{kcal}$$

$$c_x = \frac{0.96\ \text{kcal}}{3.0\ \text{kg}^\circ\text{C}} = 0.32\frac{\text{kcal}}{\text{kg}^\circ\text{C}}.$$

The method of mixtures also can be used to determine the specific heat of gases, but it is very difficult to perform experimentally.

2-5 METHOD OF COOLING

Another way of determining the specific heat of a liquid is by the method of cooling. The apparatus consists of a calorimeter with a test tube for the liquid suspended in the inner calorimeter (Figure 2-1). A and B are thermometers and C is a stirrer which enables the liquid to be stirred prior to its temperature being observed. The space D is filled with ice.

The test tube is first filled with hot water, and a curve similar to that in Figure 2-2 is plotted. It is then filled with an equal volume of the liquid at the same temperature as the hot water, and a cooling curve for the liquid is plotted. From the graph, the intervals of time are obtained for which the liquid and water cool through the same temperature interval Δt. Let t_1 be the time for water, and t_2—the time for the other liquid to cool through the same temperature interval. The mass of the water m_1 and that of the liquid m_2 is also determined.

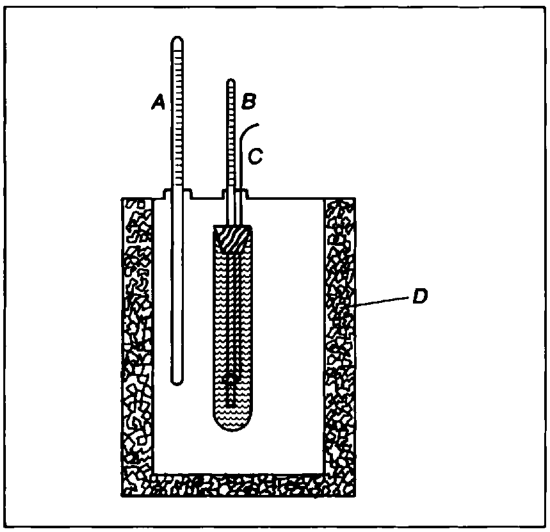


Figure 2-1 Cooling apparatus for determining the specific heat of a liquid.

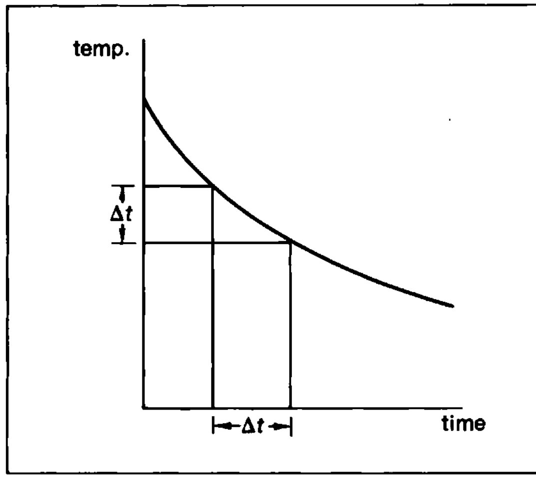


Figure 2-2 Cooling curve for a liquid.

The specific heat of the liquid is calculated as follows:

$$\frac{\text{Heat lost by the liquid}}{\text{Heat lost by the water}} = \frac{m_2 c(\Delta t)}{m_1(1)(\Delta t)} = \frac{t_2}{t_1}.$$

Therefore,

$$c = \frac{m_1 t_2}{m_2 t_1}. \tag{2-3}$$

Table 2-1 Specific Heats of Solids, Liquids, and Gases (in kcal/kg°C).

Metallic solids		Non-metallic solids		Liquids		Gases	
Aluminum	0.212	Clay	0.22	Alcohol	0.55–0.60	Air	0.24
Copper	0.094	Coal	0.3	Mercury	0.033	Hydrogen	3.40
Brass	0.090	Concrete	0.16	Oil	0.5	Nitrogen	0.25
Gold	0.030	Glass	0.12–0.20	Water	1.00	Oxygen	0.22
Iron	0.11	Granite	0.19	Turpentine	0.41		
Lead	0.030	Ice	0.48				
Nickel	0.106	Limestone	0.22				
Platinum	0.032	Marble	0.21				
Silver	0.056	Paraffin	0.69				
Steel	0.11	Rubber	0.48				
Tin	0.055	Wood	0.3–0.7				
Zinc	0.092						

Other methods of determining the specific heat of substances are methods based on change-of-state and electrical methods. Table 2-1 is a representative listing of the specific heats of various solids, liquids, and gases.

It can be seen from the above table that water has a greater specific heat than most substances making up the solid surface of the earth. This fact is important in explaining why the climate is milder near large bodies of water. In the winter, the temperature of the water does not fall very low and thus tends to prevent the temperature of the land from falling very low, while in the summer the water does not rise much in temperature and tends to moderate the temperature of the land.

2-6 LATENT HEAT OF FUSION

We can add heat to a substance without raising its temperature if the substance is undergoing a change of state. If, for example, we add heat slowly to a mixture of ice and water which is initially at a uniform temperature, we will observe that the temperature does not rise until all the ice is melted. The heat supplied is utilized in changing the ice to water. This heat is stored as thermal energy in the water, and is given out when the water changes back to ice. As this heat cannot be detected by a thermometer, it is usually referred to as latent.

> The latent heat of fusion is the quantity of heat that must be added to a unit mass of a solid so as to melt it, or the quantity of heat that must be removed from a unit mass of a liquid so as to solidify it, without a change in temperature and pressure.

The latent heat of fusion can be determined by the method of mixtures.

2-7 LATENT HEAT OF VAPORIZATION

In the last section, we stated that while a solid is changing to a liquid or vice versa there is no change in temperature. Similarly, when a liquid is changing to a vapor or vice versa at constant pressure there is no change in temperature. To change a given mass of water at the boiling point into vapor or steam, a definite amount of heat must be supplied; when an equal mass of steam condenses to water, an equal amount of heat is given out. The heat supplied is stored as thermal energy in the steam, and is given out when the steam changes back to water; this is called the latent heat of vaporization.

> The latent heat of vaporization is the quantity of heat that must be added to a unit mass of a liquid so as to vaporize it, or the quantity of heat that must be removed from a unit mass of a vapor so as to condense it, without a change in temperature and pressure.

The latent heat of vaporization can also be determined by the method of mixtures.

Example 2. A 40 gm copper calorimeter contains 50 gm of ice; both are at 0°C. Ten grams of steam at atmospheric pressure are passed into the calorimeter. What is the final temperature of the calorimeter and its contents?

SOLUTION

Heat gained by ice + heat gained by calorimeter = heat lost by steam.

$$[(50\text{ gm})(80\text{ cal/gm})] + [(50\text{ gm})(t-0)(1\text{ cal/gm}^\circ\text{C})] + [(40\text{ gm})(t-0)(0.094\text{ cal/gm}^\circ\text{C})] = (10\text{ gm})(540\text{ cal/gm}) + 10\text{ gm}(100^\circ\text{C}-t)(1\text{ cal/gm}^\circ\text{C})$$

$$4000\text{ cal} + 50t\text{ cal/}^\circ\text{C} + 3.76t\text{ cal/}^\circ\text{C} = 5400\text{ cal} + 1000\text{ cal} - 10t\text{ cal/}^\circ\text{C}$$

$$63.76t = 2400^\circ\text{C}$$

$$t = 37.6^\circ\text{C}.$$

Table 2-2 gives the melting points, boiling points, latent heat of fusion, and vaporization of various materials.

2-8 HEAT OF COMBUSTION

The heat given out when substances burn is of importance in engineering, chemistry, and physics.

> The heat of combustion of a substance is the quantity of heat given out per unit mass or per unit volume when the substance is completely burned.

The heats of combustion of solids and liquids are usually expressed as the heat per unit mass, while the heats of combustion of gases are usually expressed as the heat per unit volume. The heats of combustion of solids and liquids are usually mea-

Table 2-2 An Illustrative List of Melting and Boiling Points, Latent Heats of Fusion, and Vaporization at One Atmosphere Pressure.

Substance	Melting point °C	Boiling point °C	Heat of fusion kcal/kg	Heat of vaporization kcal/kg
Alcohol (ethyl)	−115	78	25	205
Alcohol (methyl)	−98	65	22	267
Aluminum	660	2060	93	2000
Ammonia	−75	−34	108	327
Copper	1080	2600	51	1760
Gold	1060	2970	16	446
Hydrogen	−259	−253	15	107
Iron	1540	2740	65	1620
Lead	327	1744	6.3	222
Mercury	−39	357	2.7	71
Methane	−182	−161	14.5	138
Nitrogen	−210	−196	6.2	48
Oxygen	−219	−183	3.3	51
Platinum	1774	4407	27	640
Silver	960	2212	24	552
Tin	232	2270	14	650
Tungsten	3400	5930	44	1180
Water	0	100	80	540
Zinc	419	907	24	362

sured with a bomb calorimeter (Figure 2-3). A certain mass of fuel is put in the bomb (a steel cylinder fitted with a gas tight cover) with pure oxygen under pressure so as to insure complete combustion. The bomb is surrounded by water in a calorimeter. The fuel is ignited by sending an electric current through a heater wire within the bomb. The heat of combustion is calculated from the rise in temperature of the water, bomb, and calorimeter.

The heat of combustion of gases is usually measured by using a continuous flow calorimeter. Here, water is let flow continuously through a tube heated with the gas flame. From the change in temperature of the water, the mass of water that flows through the tube, and the volume of gas consumed by the flame, all in the same interval of time, the heat of combustion of the gas can be calculated.

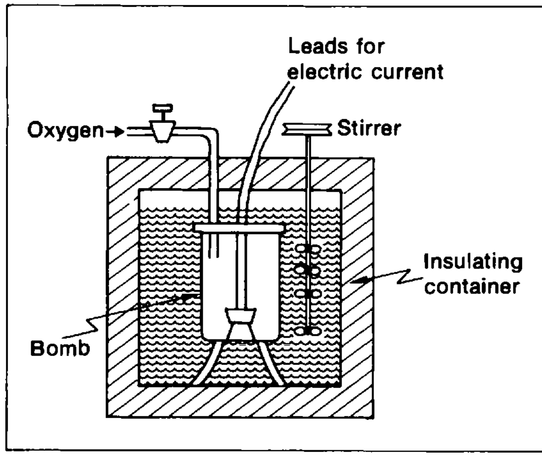


Figure 2-3 Bomb calorimeter for measuring the heats of combustion of solids and liquids.

Table 2-3 contains an illustrative list of the heats of combustion of a number of solids, liquids, and gases.

Table 2-3 Heats of Combustion.

Substance	kcal/kg	Substance	kcal/m^3
Solids:		Gases:	
Coal	6000–7800	Acetylene	12,900
Wood	4500–5000	Coal gas	4300
Liquids:		Natural gas	8000–18,000
Alcohol	6400	Propane	20,600
Gasoline	11,400		
Diesel oil	10,600		
Kerosene	11,000		

PROBLEMS

1. Find the equilibrium temperature when 80 gm of iron at 100°C are dropped into 200 gm of water contained in a 50 gm iron vessel which are both at 20°C.
2. A 50 gm sample of a solid material, at a temperature of 100°C, is dropped into a 100 gm copper calorimeter containing 200 gm of water initially at 20°C. The final temperature of the calorimeter and its contents is 22°C. What is the specific heat of the sample?

3. A piece of solid metal of mass 0.150 kg was heated at 659°C and then plunged into 0.100 kg of water at a temperature of 25°C. When the water and metal reached equilibrium in temperature, their combined mass was found to be 0.247 kg.
   (a) What became of the mass that was lost?
   (b) What was the final equilibrium temperature?
4. Fifty grams of aluminum at 98°C is dropped into 37.2 gm of alcohol at 10°C in a 50 gm copper calorimeter. The final temperature of the mixture is 35°C. What is the specific heat of the alcohol?
5. To find the specific heat of methyl alcohol, a 40 gm mass of aluminum at 200°C is added to 120 gm of the alcohol, which is initially at 20°C. The final temperature of the mixture is 40°C. What does this give for the specific heat of the alcohol?
6. A 0.300 kg aluminum vessel contains 0.600 kg of water at 20°C. How much steam at 100°C must be added to raise the temperature of the water to 100°C?
7. How much heat is required to melt 10 gm of ice (already at 0°C), heat the resulting water to 100°C, and boil it all away?
8. Ice at − 10°C and steam at 110°C come into thermal equilibrium as water at 50°C in a calorimeter. Take the specific heat of steam at 0.50 kcal/kg°C. What is the ratio of the mass of ice to the mass of steam?
9. Twenty grams of steam at 100°C were condensed in 362 gm of water at 10°C. The heat capacity of the container is 40 cal/°C, and the final temperature is 40°C. Calculate the heat of vaporization.
10. Two kilograms of molten tin at its melting point are dropped into a 1 kg copper vessel containing 3 kg of water at 10°C. What is the final equilibrium temperature?
11. If 200 gm of iron at 120°C, 20 gm of ice at 0°C, and 100 gm of water at 20°C are all mixed, and it is found that there is no change in the temperature of the water, what must be the specific heat of the iron?
12. In an insulated container are 50 gm of ice and 200 gm of water at 0°C. Ten grams of steam at 100°C and atmospheric pressure is allowed to flow into the mixture, condensing therein. Neglecting the heat absorbed by the container, compute the final steady state temperature of the mixture.
13. Calculate the equilibrium temperature
   (a) when we mix 100 gm of ice at − 20°C with 400 gm of water at 30°C,
   (b) when we mix 100 gm of ice at − 20°C with 200 gm of water at 30°C.
14. A mixing faucet is supplied with cold water at a rate of 0.6 gallons per minute and hot water at a rate of 1 gallon per minute. If the temperature of the cold water is 50°F and the hot water is 150°F, what is the temperature of the water from the faucet?
15. If 10 gm of gasoline are burned in a bomb calorimeter containing 3 kg of water, what is the heat of combustion of the gasoline if the heat capacity of the calorimeter is 1 kcal/°C and the rise in temperature is 25°C?
16. Twenty grams of water contained in a 30 gm copper calorimeter was found to cool from 70°F to 50°F in 10 minutes. An equal volume of another liquid with a mass of 15 gm cools from 70°F to 50°F in 6 minutes in the same calorimeter. Find the specific heat of the liquid.

3 Expansion of Solids, Liquids, and Gases

3-1 EXPANSION OF SOLIDS

Most substances expand when heated and contract when cooled. Solids have the property of retaining their shape when heated without the support of a container. All three dimensions change with temperature. If the body expands isotropically, that is, if the body expands equally in all directions, we may then take any dimension as the length of the body. Solid bodies of different materials, having different space arrangements of their atoms, expand different amounts for equal temperature changes. To express numerically the expansion of a particular substance, we need a quantity characteristic of that substance. This quantity is called the coefficient of expansion.

3-2 COEFFICIENT OF LINEAR EXPANSION OF SOLIDS

Suppose a rod of length l_0 is heated through a temperature interval Δt until its length is l_t. The difference $l_t - l_0 = \Delta l$ is the amount the rod has expanded on heating. It is found experimentally that for moderate temperature intervals the quantity $\Delta l / l_0 \Delta t$ equals a constant which is characteristic of the material in the rod. This constant α is called the coefficient of linear expansion. That is,

$$\alpha = \frac{\Delta l}{l_0 \Delta t}. \tag{3-1}$$

The coefficient of linear expansion may thus be defined as follows:

> The coefficient of linear expansion is the increase in length per unit length per degree rise in temperature.

Equation (3-1) can be written in a different form by replacing Δl by $l_t - l_0$, namely,

$$l_t = l_0(1 + \alpha \Delta t). \tag{3-2}$$

This is sometimes more convenient than Eq. (3-1).

The units of α are $°C^{-1}$ or $°F^{-1}$. Since the Fahrenheit degree is only $\frac{5}{9}$ as large as the Centigrade degree, α (per °F) is equal to $\frac{5}{9}\alpha$ (per °C). Typical values of the coefficient of linear expansion of various solids are given in Table 3-1.

Table 3-1 Coefficients of Linear Expansion of Solids.

Substance	α in units of $10^{-6}\,°C^{-1}$	Substance	α in units of $10^{-6}\,°C^{-1}$
Aluminum	24	Magnesium	26
Brass	19	Nickel	13
Copper	14	Platinum	9
Glass (ordinary)	9	Quartz (fused)	0.4
Glass (pyrex)	3	Silver	19
Gold	14	Steel	12
Iron	11	Tin	20
Invar	0.9	Tungsten	4.5
Lead	30	Zinc	30

Example 1. An aluminum rod is 200 cm long at 0°C. What is its length at 100°C?

SOLUTION

$$\begin{aligned}\text{Length at } 100°\text{C} &= l_{100} = l_0(1+\alpha\Delta t)\\ &= 200\text{ cm}\,[1+(24\times10^{-6}\,°\text{C}^{-1})\\ &\quad\times(100°\text{C})]\\ &= 200\text{ cm} + (200\text{ cm})(24\times10^{-4})\\ &= 200.48\text{ cm.}\end{aligned}$$

3-3 COEFFICIENT OF AREA EXPANSION

Consider a square sheet of length l_0 heated through a temperature interval Δt until its length is l_t. If α is the coefficient of linear expansion of the material, its new area is given by

$$\begin{aligned}A_t &= l_t^2 = [l_0(1+\alpha\Delta t)]^2 = l_0^2(1+2\alpha\Delta t+\alpha^2\Delta t^2)\\ &= A_0(1+2\alpha\Delta t+\alpha^2\Delta t^2).\end{aligned}$$

Since $\alpha\Delta t$ is a very small quantity for all solids, $\alpha^2\Delta t^2$ may be neglected in comparison with $2\alpha\Delta t$. Therefore,

$$A_t = A_0(1+2\alpha\Delta t), \tag{3-3}$$

and the coefficient of area expansion is twice the coefficient of linear expansion. Although the above result was derived for a square sheet, it holds for a sheet of any shape.

> The coefficient of area expansion of a solid is the increase in area per unit area per degree rise in temperature, and is twice the coefficient of linear expansion for an isotropic solid.

It is worth noting that Eq. (3-3) is applicable to the increase in area of a hole in a sheet of material as well as the increase in area of the material itself. Can you explain this?

3-4 COEFFICIENT OF VOLUME EXPANSION

Consider a cube of material of edge length l_0 heated through a temperature interval Δt until its length is l_t. If α is the coefficient of linear expansion of the material, its new volume is given by

$$\begin{aligned}V_t &= l_t^3 = [l_0(1+\alpha\Delta t)]^3\\ &= l_0^3(1+3\alpha\Delta t+3\alpha^2\Delta t^2+\alpha^3\Delta t^3)\\ &= V_0(1+3\alpha\Delta t+3\alpha^2\Delta t^2+\alpha^3\Delta t^3).\end{aligned}$$

Since we can neglect the terms containing $\alpha^2\Delta t^2$ and $\alpha^3\Delta t^3$,

$$V_t = V_0(1+3\alpha\Delta t) \tag{3-4}$$

is an equation which holds for an isotropic substance of any shape.

> The coefficient of volume expansion of a solid is the increase in volume per unit volume per degree rise in temperature, and is three times the coefficient of linear expansion for an isotropic solid.

Equation (3-4) is applicable to the increase in volume of a cavity in a solid as well as the increase in volume of the solid itself. Can you explain this?

3-5 IMPORTANCE AND APPLICATIONS OF THE EXPANSION OF SOLIDS

In pipelines, bridges, and non-welded rails, allowance must be made for temperature changes, otherwise they will bend with a rise in temperature. This is taken care of in pipelines by expansion joints and in bridges and rails by leaving a small space between the bridge sections or rails. The joint is usually made by means of two plates, one on each side of the bridge sections or rails. Bolts are put through the plates and through slotted holes in the bridge section or rail so that they are free to slide. Clock pendulums must be compensated for expansion in order for them to keep time correctly.

The fact that different metals have different coefficients of expansion has been applied to the construction of thermometers and thermostats. If two thin strips of different metals welded or riveted

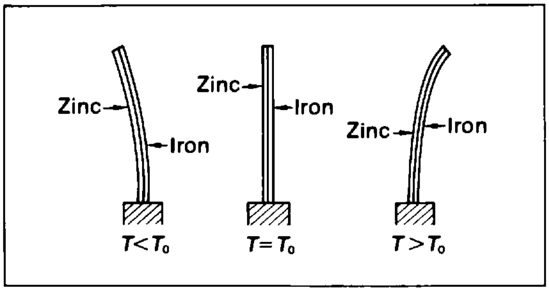


Figure 3-1 Bimetallic thermometer or thermostat element.

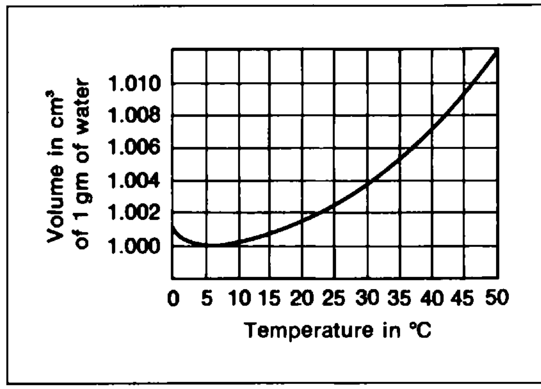


Figure 3-2 The anomalous expansion of water.

together are heated, the resultant strip will bend into a curve (Figure 3-1). The metal with the larger coefficient of expansion will be on the convex side of the curve. When used as a thermostat, one end is fixed and the other end opens or closes an electrical circuit. That is, it operates as a switch which is governed by the temperature. The temperature control of a cooking oven operates on this principle.

3-6 EXPANSION OF LIQUIDS

Since a liquid must be held in a solid container, the true expansion of the liquid can only be obtained if we also know the expansion of the container.

The apparent volume expansion of the liquid is equal to the difference between the true volume expansion of the liquid and the volume expansion of the container. The new volume of the container is given by

$$ {}_cV_t = V_0(1 + 3\alpha_c \Delta t), \quad (3\text{-}5) $$

where α_c is the linear expansion coefficient of the container. The new true volume of the liquid is given by

$$ {}_lV_t = V_0(1 + \beta_l \Delta t), \quad (3\text{-}6) $$

where β_l is the volume expansion coefficient of the liquid.

The apparent volume expansion of the liquid is

$$ {}_lV_t - {}_cV_t = V_0(\beta_l - 3\alpha_c)\Delta t. \quad (3\text{-}7) $$

Hence, the apparent volume expansion coefficient of the liquid is

$$ \beta_l - 3\alpha_c = \frac{\Delta V}{V_0 \Delta t}, \quad (3\text{-}8) $$

where we have replaced ${}_lV_t - {}_cV_t$ by ΔV.

In the above experiment, we could measure the value of β_l if we knew α_c. If we had filled the container with water both enclosed in a bath at 0°C and raised the temperature of the bath to 4°C, we would have found that the water decreased in volume. The reason is that water has a peculiar expansion property. It has a negative expansion coefficient between 0°C and 4°C or , in other words, when water is heated from 0°C it contracts with temperature rise until it reaches 4°C and then begins to expand. The volume per gram of water at temperatures between 0°C and 100°C is shown graphically in Figure 3-2.

Water has its greatest density at 4°C. This is important in nature. It is the reason why the surface of lakes and rivers freeze first. The water at the bottom stays at 4°C, enabling the fish to enjoy a warmer winter than some land dwellers. There are a number of accurate methods for determining the expansion coefficient of liquids. The interested student may read about these methods in books on

Table 3-2 Coefficients of Volume Expansion of Liquids.

Substance	$\beta\ 10^{-5}\ °C^{-1}$
Alcohol (ethyl)	112
Alcohol (methyl)	120
Benzene	124
Carbon tetrachloride	124
Ether (ethyl)	166
Glycerin	50
Mercury	18
Petroleum	95
Turpentine	97
Water	21

heat that are available in the library. Typical values of the coefficient of volume expansion of various liquids are given in Table 3-2.

Example 2. A pyrex glass bottle is filled with mercury at 20°C. The volume of the bottle is 100 cm^3. How many cm^3 spill over when the bottle is placed in steam at 100°C?

SOLUTION

Volume that spills over $= \Delta V = (\beta_m - 3\alpha_c)V_0\Delta t$

$= [(180 - 3 \times 3) \times 10^{-6}\,°C^{-1}](100\,cm^3)(80°C)$

$= (171 \times 10^{-6}\,°C^{-1})(100\,cm^3)(80°C)$

$= 1.37\,cm^3.$

3-7 EXPANSION OF GASES

Gases do not have a definite volume. If a gas is put into a container, it fills the complete volume of the container. It also exerts a definite pressure upon the walls of the container.

Suppose we have a container of definite volume, containing a definite mass of gas, which exerts a definite pressure on its walls. If the container is a cylinder fitted with a piston and we change the volume by pushing the piston in or out of the cylinder, the pressure changes. If we keep the piston fixed and change the temperature of the gas, the pressure changes. If we allow the piston to move freely and change either the pressure or the temperature, the volume changes.

3-8 BOYLE'S LAW

In 1660, Robert Boyle found a relation between the pressure and volume of a gas if the temperature of the gas is held constant. The relation is known as Boyle's law, and is stated as follows:

> The product of the pressure and the volume of a given mass of gas at constant temperature is a constant.

Mathematically, it is written

$$pV = \text{constant} \quad (\text{for constant } t), \qquad (3\text{-}9)$$

or, alternately,

$$p_1V_1 = p_0V_0 \quad (\text{for constant } t), \qquad (3\text{-}10)$$

where p_1 is the gas pressure when its volume is V_1, and p_0 is the initial pressure when its initial volume is V_0, both states being at the same temperature.

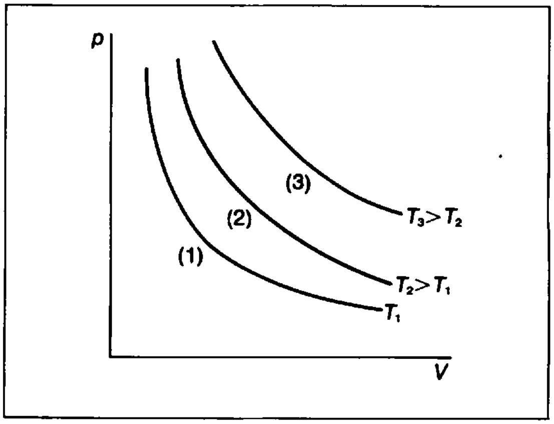


Figure 3-3 Graphic illustration of pV=constant for various temperatures for one amount of gas.

The constant in Eq. (3-9) depends upon the temperature of the gas. Figure 3-3 is a graphical illustration of Eq. (3-9). Curve (1) represents Boyle's law for a given mass of gas at absolute temperature T_1, curve (2) for an absolute temperature T_2, etc. They are called isothermal curves (curves of constant temperature).

3-9 CHARLES' LAW

The volume of gases, like most solids and liquids, increases with an increase in temperature. The volume expansion coefficient of a gas is defined in the same way as that of a liquid, namely

$$\beta = \frac{\Delta V}{V_0\Delta t}. \qquad (3\text{-}11)$$

Gases, however, differ from liquids as they all have very nearly the same value of β at constant pressure. At 0°C, the volume expansion coefficient of all permanent gases is approximately $\frac{1}{273}\,°C^{-1}$.

Consider a gas whose volume is V_0 at atmospheric pressure p_0 and temperature 0°C. We heat the gas, keeping its pressure constant, to a temperature t. Its new volume is given by

$$V_t = V_0(1 + \beta t)$$

or

$$V_t - V_0 = \frac{V_0 t}{273} = V_0\frac{(T - 273)}{273} = \frac{V_0 T}{273} - V_0$$

or

$$V_t = \frac{V_0 T}{273}, \qquad (3\text{-}12)$$

where T is the absolute temperature in °K.

We see that the volume of the gas is directly proportional to the absolute temperature at constant pressure. This is known as Charles' law, and is stated as follows:

> At constant pressure, the volume of a given mass of a gas varies directly as the absolute temperature.

Symbolically, it is stated as

$$\frac{V}{T} = \text{constant} \quad \text{(at constant pressure).} \tag{3-13}$$

Alternately, it is written

$$\frac{V_1}{T_1} = \frac{V_0}{T_0} \tag{3-14}$$

or

$$\frac{V_1}{V_0} = \frac{T_1}{T_0}. \tag{3-15}$$

Similarly, at constant volume, the pressure of a given mass of gas varies directly as the absolute temperature; an alternative form of Charles' law is

$$\frac{p_1}{p_0} = \frac{T_1}{T_0}. \tag{3-16}$$

We may well ask what happens to the pressure and volume of a gas when T equals zero. Actually, all gases liquefy at temperatures above absolute zero, so the question has no physical significance. Figure 3-4 is a graph of a volume of a particular mass of gas at constant pressure versus its temperature in both Centigrade and Kelvin degrees. We see that the volume of the gas is directly proportional to its absolute temperature.

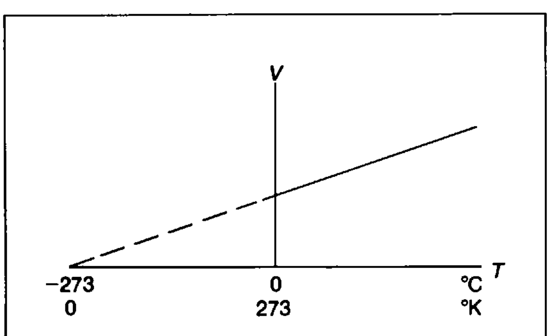


Figure 3-4 Variation of volume of a gas with temperature at constant pressure.

3-10 THE GENERAL GAS LAW

Boyle's law and Charles' law provide the basis for a more general relation between pressure, volume, and temperature known as the general gas law. Combining Eqs. (3-10) and (3-15), we get

$$\frac{p_1 V_1}{T_1} = \frac{p_0 V_0}{T_0}. \tag{3-17}$$

Since p_1, V_1, and T_1 can be any set of pressure, volume, and temperature values, we can remove the subscripts. The value of V_0, if taken as the gas volume at atmospheric pressure p_0 and temperature T_0 (0°C), depends on the mass and kind of gas used. However, one kilomole (kilogram molecular weight) of any gas has a volume of 22.4 m^3 at atmospheric pressure and 0°C. Hence, if the gas contains n kilomoles, we can write Eq. (3-17) as

$$pV = nRT,$$

where

$$R = \frac{p_0 V_0}{T_0}. \tag{3-18}$$

The value of R has the same value for all gases, and is known as the universal gas constant.

$$R = 8.31 \times 10^3 \text{ joules/(kilomole °C)}.$$

Equation (3-18) is called the general gas law.

If we substitute $T_0 = 273$ and $T_1 = 273 + t$ in Eq. (3-17), we obtain

$$p_1 V_1 = p_0 V_0 \left(\frac{273 + t}{273}\right) \tag{3-19}$$

or

$$p_1 V_1 = p_0 V_0 \left(1 + \frac{t}{273}\right) \tag{3-20}$$

or

$$p_1 V_1 = p_0 V_0 (1 + \beta t), \tag{3-21}$$

where $\beta = \frac{1}{273}$ °C^{-1}.

It should be pointed out that these equations do not hold exactly at all pressures and temperatures. However, it is quite remarkable that these simple equations specify the behavior of all gases without too much error except at high pressures and very low temperatures.

Example 3. The volume of a gas at 27°C and atmospheric pressure is 1 m^3. What is the volume when the pressure is equal to four times the atmospheric pressure and the temperature is 327°C.

SOLUTION

$$\frac{p_2V_2}{T_2} = \frac{p_1V_1}{T_1} \qquad \frac{4p_1V_2}{600°K} = \frac{(p_1)(1\ m^3)}{300°K} \qquad V_2 = \frac{600}{1200} = 0.50\ m^3.$$

PROBLEMS

1. At 25°C an iron rod is 2 cm in diameter. A brass ring has an interior diameter 1.995 cm at the same temperature. At what temperature, both the rod and ring being heated, will the ring just slide onto the rod?
2. A square brass plate with a circular hole through its center is heated from 5°C to 605°C. When cold, the plate is 32 cm on an edge and the diameter of the hole is 8 cm. What is the diameter of the hole when the plate is hot?
3. The density of a liquid is defined by the expression $\rho = m/V$, where m is the mass and V is the volume of the liquid. Show that
$$\Delta\rho = -\beta\rho\Delta T,$$
where ΔT is the change in temperature.
4. A copper container has a capacity of 500 cm^3 at 5°C. If 470 cm^3 of methyl alcohol also at 5°C are placed in the container, at what temperature will the container be just filled with alcohol?
5. A thin aluminum wire is bent into the form of a rectangle 50 cm long and 40 cm wide. When the temperature of the wire is changed from 10°C to 100°C, what is the change in area enclosed by the wire?
6. A glass flask holds exactly 1000 cm^3 of water at 5°C. When the flask and the water are heated to 95°C, 15 cm^3 of water overflow. What is the linear coefficient of expansion of the glass?
7. We have two square parallel plates in contact, one of steel and the other of aluminum. At 0°C, the aluminum plate is 1 m on a side, and the steel plate is 1.001 m on a side. It is arranged to have an etching process begin when the plates have precisely the same area. Find the temperature at which the etching process begins.
8. A pyrex glass flask in a dark cupboard contains 1000 cm^3 of ethyl alcohol at 20°C. The flask is removed from the cupboard and set in sunlight, so that in 20 minutes the temperature rises to 60°C. What volume of alcohol is present at the end of 20 minutes?
9. A narrow-necked ordinary glass bottle is just filled with 100 cm^3 of a liquid when the bottle and the liquid are at − 20°C. The coefficient of volume expansion of the liquid is 12.4×10^{-5} per °C. What volume of the liquid has spilled over by the time enough heat has been added to raise the temperature of the bottle and the liquid to 10°C?
10. A certain medical prescription requires 10 cm^3 of ethyl alcohol at 20°C. A freshly sterilized graduated pyrex flask is used for measuring out the alcohol. The flask is at a temperature of 96.3°C.
    (a) What volume of alcohol at 96.3°C must be used?
    (b) What volume on the graduated pyrex flask will give the required amount of alcohol if the flask was calibrated at 20°C?
11. Two rods of metal A and one rod of metal B are connected as shown in Figure 3-5. The overall length is L and the length of B is d. If the coefficients of linear expansion are α_A and α_B, determine d in terms of L, α_A, and α_B, if L does not change with temperature.

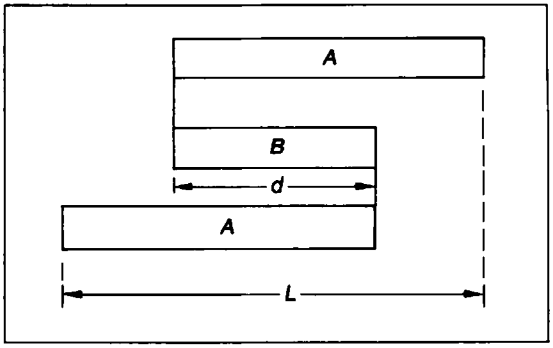


Figure 3-5

12. If 1 m^3 of a certain gas at atmospheric pressure and 0°C has a mass of 2 gm, what would be the mass of the same volume of the same gas at 0.50 atmospheric pressure and 100°C?

13. A steel rod and a brass rod are fastened to the same end plate as shown in Figure 3-6. If the distance between the ends of the rods d is to remain 8 cm over a range of temperatures, what must be the length of each rod?

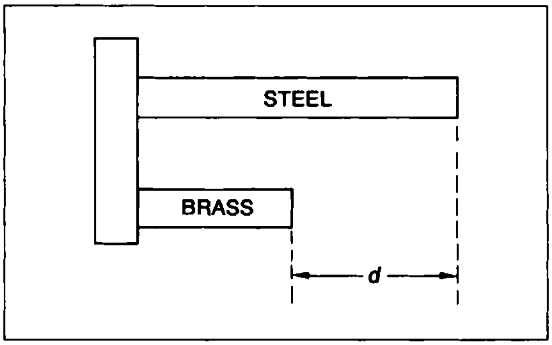


Figure 3-6

14. A hand pump with a cylinder 0.50 m long is used to pump air into a tire. Assuming the temperature of the air to remain constant, find how far the piston must be moved before air can enter the tire if the pressure of air in the tire is twice the atmospheric pressure at the beginning of the stroke.

15. A certain mass of gas occupies a volume of 380 cm^3 at 12°C and pressure 1.053 atmospheres. What will its volume be at −10°C and pressure 1 atmosphere?

16. An air bubble of 10 cm^3 volume is at the bottom of a lake, where the pressure is 4 times that at the surface and the temperature is 5°C. The bubble rises to the surface, which is at a temperature of 20°C. If the temperature of the air in the bubble is the same as the surrounding water, what is its volume just before it reaches the surface?

17. An aluminum container holds 400 cm^3 of a liquid when completely filled at 10°C. When the temperature of the container and liquid is raised to 80°C, 25 cm^3 of the liquid spill out of the container. Determine the coefficient of volume expansion of this liquid.

18. (a) If 1 m^3 of air at 1 atmosphere pressure and 27°C temperature is compressed to 0.50 m^3 at 3 atmospheres, what will be the resulting temperature?
    (b) If the gas is now allowed to cool down to its original temperature at constant volume, what will be the final pressure?

19. The cross-sectional area of the mercury column in a barometer is 1.20 cm^2, the length of the vacuum at the top is 8 cm when the barometer reads 76.4 cm. Calculate the volume of external air that must be inserted into the tube in order to lower the mercury column to 38.2 cm. Assume the atmospheric pressure and the temperature remain constant.

20. Owing to air above the mercury in a barometer, it reads 74.5 cm when the actual pressure is 75.2 cm, and 73.3 cm when the actual pressure is 73.7 cm of mercury. Calculate the reading of the barometer when the true pressure is 76 cm.

21. A metal container has an internal volume of 1000 cm^3 when its temperature is 0°C, but when its temperature is 100°C the container holds 1008 cm^3.
    (a) Determine the coefficient of linear expansion for this metal.
    (b) If a liquid just fills the container at 0°C, and 10 cm^3 spill out at 100°C, determine the coefficient of volume expansion of the liquid.

22. A thermometer is made of a capillary tube of ordinary glass of 0.020 mm^2 cross section at 0°C. At 0°C there are 2 cm^3 of mercury in the thermometer. How far does the mercury move up the tube at 30°C?

4 Heat Transfer

4-1 METHODS OF HEAT TRANSFER

There are three ways in which heat is transferred from one place to another. The three ways are by conduction, convection, and radiation.

> Conduction is the transfer of heat in which thermal energy is transferred from molecule to molecule in a material with no perceptible motion of the material.
>
> Convection is the transfer of heat by mass motion of the heated material.
>
> Radiation is the transfer of thermal energy by electromagnetic waves.

In the first two methods, a material medium is required. In radiation, no material is needed, and the heat is transferred with the speed of light.

4-2 COEFFICIENT OF THERMAL CONDUCTIVITY

Materials differ widely in their ability to conduct heat. Metals are good conductors while non-metallic materials are poor conductors. However, there is a fair difference in the conductivity of different metals as well as of different non-metals.

These differences may be illustrated in a simple way for metals. Suppose we take three equal diameter wires of aluminum, steel, and copper, and twist them together at one end as illustrated in Figure 4-1. At equal intervals along the wires we attach small metal balls by means of wax. We then heat the ends which were twisted together. As the heat is conducted along the wires, the balls drop off when the temperature of the wire is sufficient to melt the wax. After a particular time interval, the balls will stop dropping off, and the temperature is said to be in a steady state, which means that the heat passing along the wire is equal to the heat that escapes from it. When this state is reached, it is found that the copper has lost the largest number, the aluminum the next largest number, and the steel the least number of balls. This means that the copper wire is the best conductor, the aluminum is next best, and the steel is poorest.

Consider the conduction of heat through a slab of

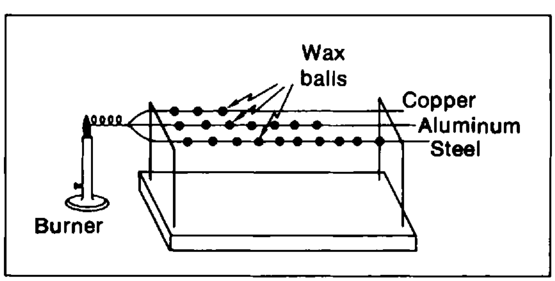


Figure 4-1 Comparison of conductivity of metals.

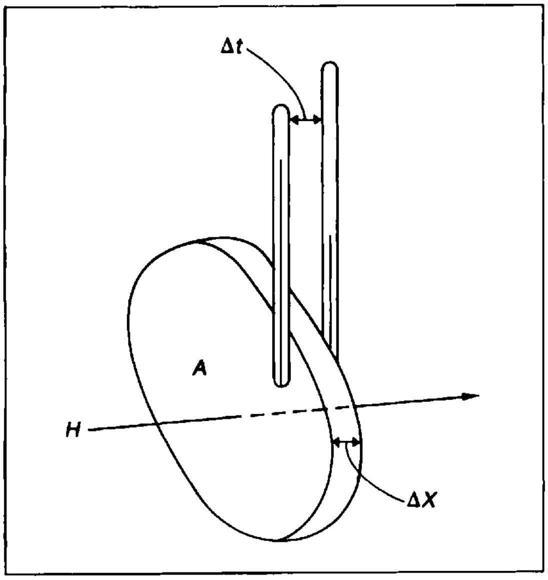


Figure 4-2 Conductivity of heat through a slab of material.

material of face area A and thickness ΔX, with a difference of temperature between the faces of Δt after a steady state is reached, as illustrated in Figure 4-2. H, the heat per unit time, passing through the slab is found experimentally to be proportional to the temperature gradient $\Delta t/\Delta X$ and the area A. Then

$$H \propto A\frac{\Delta t}{\Delta X}$$

or

$$H = KA\frac{\Delta t}{\Delta X}, \qquad (4\text{-}1)$$

where the proportionality constant K is called the coefficient of thermal conductivity of the material, and is defined as follows:

> The coefficient of thermal conductivity of a material is the time rate of heat flow by conduction per unit area per unit temperature gradient.

Various systems of units are used in specifying K. In MKS units, K is expressed in kcal/sec m°C. Table 4-1 gives typical values of K for a number of substances.

Example 1. An aluminum pan has a 0.10 m² heating surface, 0.20 cm thick. How much water evaporates per minute if the outer surface is kept at 110°C?

SOLUTION

From Eq. (4-1),

$$H = KA\frac{\Delta t}{\Delta x} = (480 \times 10^{-4}\ \text{kcal/sec m°C}) \times (0.10\ \text{m}^2)\frac{[(110-100)°\text{C}]}{(2\times 10^{-3}\ \text{m})} = 24\ \text{kcal/sec}.$$

Therefore, the heat flow in 1 minute

$$= (24\ \text{kcal/sec})(60\ \text{sec/min}) = 1.44 \times 10^3\ \text{kcal/min}.$$

L_v, the heat of vaporization of water, is 540 kcal/kg. Therefore, the mass of water vaporized in 1 minute is given by

$$(1.44 \times 10^3\ \text{kcal/min})\left(\frac{1\ \text{kg}}{540\ \text{kcal}}\right) = 2.67\ \text{kg/min}.$$

4-3 CONDUCTIVITY OF LIQUIDS AND GASES

We can see from Table 4-1 that all liquids have very low conductivities. As a demonstration, we can partially fill a test tube with water and boil the upper portion of it without appreciably heating the bottom, as illustrated in Figure 4-3.

Table 4-1 Typical Values of the Coefficient of Thermal Conductivity (K) for Various Substances.

Substance	K 10^{-4} kcal/sec m°C
Metals:	
Aluminum	480
Copper	920
Brass	260
Iron and Steel	110
Silver	1000
Non-Metals:	
Brick (fire)	2.5
Brick (insulating)	0.4
Concrete	2.0
Cork	0.1
Glass	2.0
Ice	4.0
Wood	0.2
Gases:	
Air	5.7×10^{-2}
Hydrogen	33.0×10^{-2}
Oxygen	5.6×10^{-2}

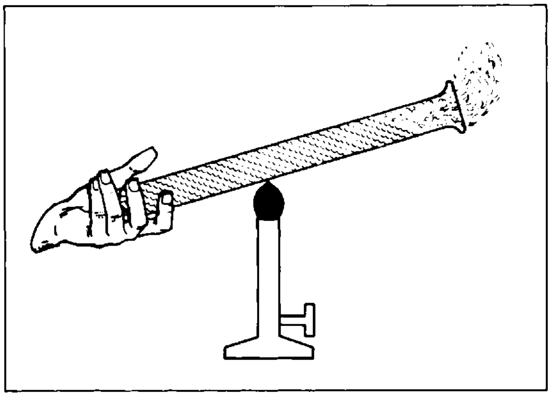


Figure 4-3 Demonstration that water is a poor heat conductor.

The conductivity of gases is extremely small as shown in Table 4-1. Many solid materials are good insulators because they are porous and contain much air. Storm windows on a house are effective chiefly because of the enclosed air between them and the regular windows. The most effective insulator of all is a vacuum, because heat can only be transferred through a vacuum by radiation.

4-4 CONVECTION

Convection occurs only in liquids and gases. As previously stated, it is the transfer of heat by the mass motion of the heated material. The material motion is due to the difference in density of a hot and cold portion of the material.

When a container of water is heated on a stove, the heat passes by conduction through the bottom of the container to the water. The lowest layer of water is heated and expands, thus becoming less dense than the colder water above. It is forced upward by the colder water, which sinks. This circulation continues until all the water is heated to the boiling point. The process is an example of convection. Hot water and hot air heating of a building are also due to convection.

Unequal heating of land and water gives rise to winds. Local land and sea breezes, as illustrated in Figure 4-4, are accounted for by convection currents owing to the higher specific heat of water. Also, the unequal heating of large bodies of water gives rise to ocean currents.

In determining the quantity of heat conducted through a wall or window, it is not correct to use the difference between the outdoor and indoor temperature as the difference in temperature between the outer and inner surfaces of the wall or window. Because of convection effects, the outer surface temperature may be well above the outdoor temperature, and the inner surface temperature well below the indoor temperature. Only part of the temperature drop occurs in the wall or window; the rest occurs in the layers of air in contact with them. In the case of an ordinary window pane in a home with an indoor temperature of 70°F and outdoor temper-

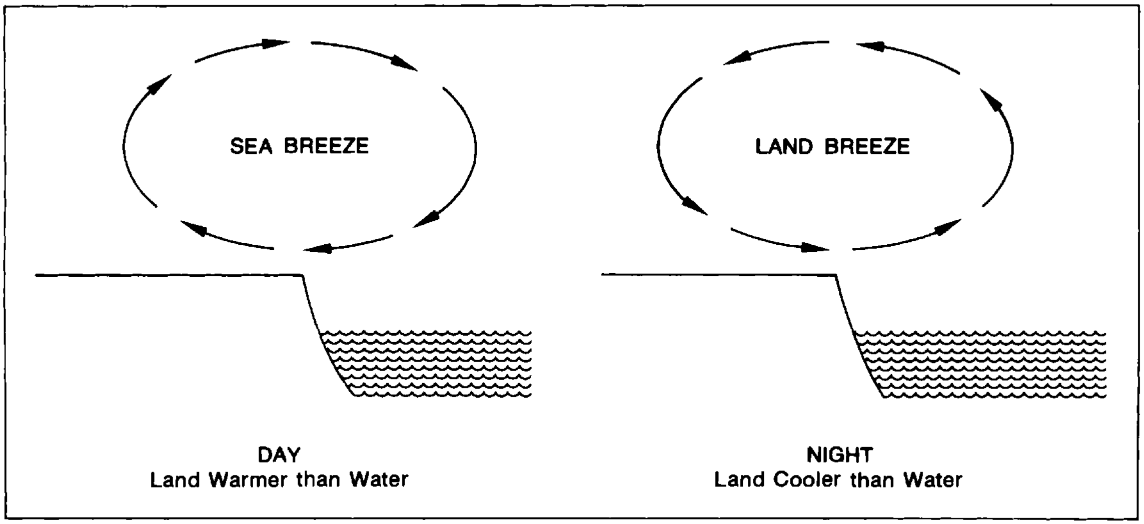


Figure 4-4 Illustration of sea and land breezes.

ature of 0°F, the temperature of the glass will be about 35°F. The temperature difference between the outside and inside of the glass will be 0.2°F.

The mathematical theory of convection is quite complex. The heat per unit time (H) transferred by convection to or from a surface can be calculated from the following equation:

$$H = hA\,\Delta t, \tag{4-2}$$

where h is the convection coefficient, A is the area of the surface, Δt is the temperature difference between the surface and the main body of the fluid.

The value of h depends on many circumstances, such as the shape and orientation of the surfaces; the density, viscosity, specific heat, thermal conductivity, and velocity of the fluid; and whether or not evaporation or condensation takes place. A large amount of research has been done in the determination of h. As a result, there are in existence many formulae, graphs, and tables from which h can be obtained for various conditions.

A law discovered by Newton relates the rate of cooling or heating of a given substance to the difference in temperature between it and its surroundings. The law is stated as follows:

> The time rate of heat flow from a body to the surroundings or vice versa is directly proportional to the difference in temperature between the body and the surroundings provided the difference in temperature is small.

Mathematically, it is stated as

$$H = C\Delta t, \tag{4-3}$$

where C is a constant. The law agrees well with experiment when the heat transfer is chiefly by convection.

Example 2. The temperature of a room is 20°C, and the inside temperature of a window pane 0.2 cm thick is 0°C. The value of the convection coefficient for the pane is 9×10^{-4} kcal/sec m^2°C. What is the heat transferred per unit area? What is the outside temperature of the window and the outdoor temperature?

SOLUTION

$$H = hA\,\Delta t = (9 \times 10^{-4}\ \text{kcal/sec m}^2\,°\text{C})(1\ \text{m}^2)(20°\text{C}) = 18 \times 10^{-3}\ \text{kcal/sec m}^2.$$

Hence, 18×10^{-3} kcal/sec m^2 must be conducted through the window pane itself. Using Eq. (4-1),

$$18 \times 10^{-3}\ \text{kcal/sec m}^2 = (2 \times 10^{-4}\ \text{kcal/sec m}\,°\text{C})\left(\frac{\Delta t}{2 \times 10^{-3}\ \text{m}}\right)$$

$$\Delta t = 0.18°\text{C}.$$

Therefore, the outside temperature of the pane = − 0.18°C.

Since the same quantity of heat must be transferred to the outdoors from the outside of the window as from the room to the inside of the window, the difference in temperature in both cases must be equal. Therefore, the outdoor temperature will be − 20.18°C.

4-5 RADIATION

All bodies not at absolute zero temperature emit electromagnetic radiation. Electromagnetic radiation includes thermal radiation, light, and X-rays. The principal difference in these types of radiation is wavelength (see Section 10-5). The quantity of radiation emitted by a body depends on its temperature and the nature of its surface. When electromagnetic radiation of any kind falls on a material surface, some may be absorbed, some reflected, and some transmitted. The part absorbed is transformed into thermal energy or other forms of energy within the absorbing material.

Heat transfer by radiation does not require a material medium and the heat in transit travels with the speed of light. An example of this is the transfer of heat from the Sun to the Earth. It travels through ninety million miles of space in which there is no material substance in about 8 minutes.

4-6 PREVOST'S THEORY OF EXCHANGES

If a number of bodies at different temperatures are placed in an evacuated container with insulated walls, they transfer thermal radiation between themselves and the container walls until they finally all reach the same temperature.

After they all reach the same temperature, the radiation does not stop. A thermal equilibrium is reached and each body absorbs as much radiation as it emits per unit time. This observation was first made by Prevost in 1792. From his observations, he proposed a theory which can be stated as follows:

> In a state of equilibrium, the amount of energy radiated per unit time from an object is equal to the energy absorbed by it in the form of radiations from surrounding objects.

The amount of radiation absorbed or emitted by a body is dependent on the nature of its surface. A polished surface is a poor absorber and a poor emitter, whereas a black surface is a good absorber and a good emitter. This can be demonstrated with the apparatus shown in Figure 4-5. The two bulbs A and B are connected to an air thermometer C. If a can half painted black and the other half left shiny is filled with hot water and placed between the bulbs, it can be demonstrated that the black surface is the better radiator. Next, if we blacken one of the bulbs and use a totally black can filled with hot water, we can demonstrate that a black surface is a better absorber than a shiny one.

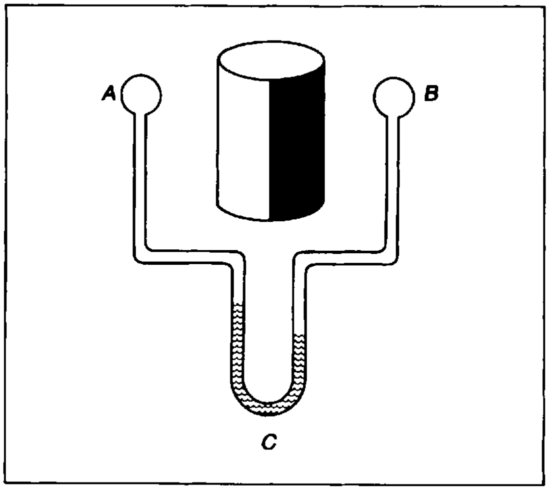


Figure 4-5 Demonstration that a black surface is a better radiator than a polished surface.

If radiation falls on an opaque body, some is absorbed and some is reflected. If α denotes the fraction that is absorbed and ρ the fraction that is reflected, then

$$\alpha + \rho = 1, \tag{4-4}$$

where α is called the absorption factor and ρ the reflection factor.

We previously gave a demonstration that a good radiator is a good absorber. This means that the radiation rate of a surface must be proportional to the absorption factor of the surface. The following law, called *Kirchhoff's Law of Radiation*, states this fact:

> The ratio of the rates of radiation of any two surfaces is equal to the ratio of the absorption factors of the two surfaces.

Symbolically, the law is written as follows:

$$\frac{W_1}{W_2} = \frac{\alpha_1}{\alpha_2}, \tag{4-5}$$

where W_1 and W_2 are the radiation rates or radiation powers of the two surfaces and are expressed in joules/sec m^2 or watts/m^2. (Note: 1 watt = 1 joule/sec.)

> A perfect radiator is a body that is a perfect absorber, and is called a black body.

No surface is a perfect absorber with $\alpha = 1$. An approximation to a perfect absorber may be obtained by a hollow sphere having a small opening with the inside walls having a rough, dull surface as shown in Figure 4-6. The radiation enters or leaves the cavity through a small hole. Part of the radiation entering the cavity will be absorbed by its walls and part reflected. Only a very small part of the reflected radiations escape through the hole, so that after many internal reflections nearly all the radiation is absorbed and the body approximates a black body.

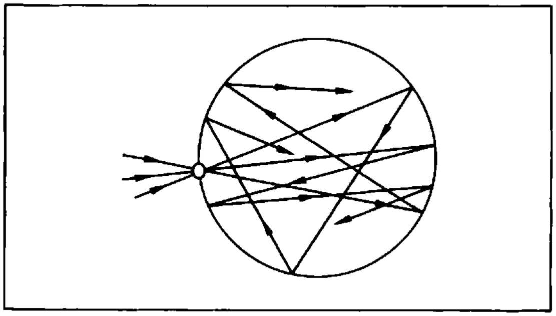


Figure 4-6 An approximation of a perfect absorber.

It was mentioned at the beginning that the quantity of radiation emitted by a body depends on its temperature. In fact, the total radiation emitted by a body increases very rapidly as the temperature is raised. The quantitative relation between the rate of radiation from a body and its surface temperature is given by the following law, known as the *Stefan–Boltzmann Law*:

> The rate at which a radiator emits radiant energy is directly proportional to the fourth power of its absolute temperature.

Symbolically, the law is written as follows:

$$W = \alpha\sigma T^4, \qquad (4\text{-}6)$$

where W is the radiation power per unit area in joules/sec m^2 or watts/m^2, α is the absorption factor which varies between zero and unity, and σ is the Stefan–Boltzmann constant with a numerical value of 5.669×10^{-8} watts/m^2°K^4.

Example 3. If the radiation from a small opening in a coal stove approximates black body radiation, and it radiates 2.84×10^5 watts/m^2, calculate its temperature in degrees Centigrade.

SOLUTION

$$W = \alpha\sigma T^4$$

$$2.84 \times 10^5 \text{ watts/m}^2 = (1)(5.67 \times 10^{-8} \text{ watts/m}^2 \text{°K}^4)(T^4)$$

$$T = \left[\frac{2.84 \times 10^5 \text{ watts/m}^2}{5.67 \times 10^{-8} \text{ watts/m}^2 \text{°K}^4}\right]^{1/4}$$

$$= 1496\text{°K}$$

$$t = 1223\text{°C}.$$

PROBLEMS

1. Water in a glass beaker is boiling away at a rate of 30 gm/min. The bottom of the beaker has an area of 300 cm^2 and it is 0.2 cm thick. Calculate the temperature at the underside of the bottom of the beaker.
2. A boiler made of steel plate 1.5 cm thick has a surface area of 8 m^2. The boiler contains water at 100°C, and its outer surface is 80°C. How much heat does it lose by conduction?
3. A certain object radiates energy at the rate of 10 kcal/sec m^2 when it is at 27°C. What is its rate of radiation at 127°C?
4. If 2.80 kcal/sec is the rate of conductive heat transfer through a flat 4 m^2 slab of glass, find the temperature gradient in the glass.
5. We have determined that a very hot blue-white star has a surface temperature of 23,000°K. Measurements and simple calculation show that this star radiates at the rate of 1.4×10^{10} watts per square meter of its surface. Find the reflection factor for the star.
6. The filament in a light bulb has a diameter of 0.20 mm and an absorption factor of 0.30. What will its temperature be when it radiates 5 watts per cm length?
7. The rear wall of a fireplace has an effective area of 0.50 m^2. It can be considered a black body surface, and has a temperature of 327°C. How many watts does it radiate to the room?
8. Why is a thermos bottle
   (a) evacuated?
   (b) double-walled?
   (c) silvered on the inside?
9. By measurement and simple calculations, it has been found that the Sun radiates at the rate of 6.25×10^7 watts per square meter of its surface. A certain physical model of a hot gas predicts that the Sun should have a reflection factor of 0.10. Find the Sun's surface temperature if this model holds.
10. Consider two spherical bodies A and B of the same size in a black enclosure. Each is free to radiate thermal energy to the other. Each body is in thermal equilibrium with its immediate surroundings and with the other. The temperature of the bodies and their immediate surroundings has a fixed value T. A has a reflection factor of 0.70 and B has an absorption factor of 0.60.
    (a) Which body, A or B, radiates energy at the greater rate?
    (b) Explain your answer.

11. A beer cooler made of oak contains 8 kg of ice at 0°C. Its dimensions are $0.60 \times 0.40 \times 0.40$ m. The wood is 5 cm thick, and has a thermal conductivity of 6×10^{-5} kcal/m sec °C. If the outside temperature is 20°C,
(a) what is the rate of heat flow into the ice box?
(b) how long will it take to melt the ice?

12. A long rod, insulated to prevent heat losses, has one end immersed in boiling water (at atmospheric pressure) and the other end in a water-ice mixture. The rod consists of 1 m of copper (one end in steam) and a length L_2 of steel (one end in ice). Both rods are of cross-sectional area 5 cm^2. The temperature of the copper iron junction is 60°C, after a steady state has been set up.
(a) How much heat flows per second from the steam bath to the ice water mixture?
(b) How long is L_2?

13. Heat, sufficient to vaporize water at 100°C at a rate of 1.80 kg per hour, passes through the bottom of an aluminum pan, 2 mm thick and 250 cm^2 in area, with a certain flame under the pan. Calculate the temperature of the underside of the pan next to the flame.

14. The operating temperature of a tungsten filament in an incandescent lamp is 2227°C, and its absorption factor is 0.30. Find the surface area of the filament of a 100 watt lamp.

15. In a brick house, the air in a room has a temperature of 20°C when the outside temperature is -18°C. The convection coefficient is 10^{-3} kcal/sec m^2 °C inside and 2×10^{-3} kcal/sec m^2 °C outside. The rate of heat transfer per unit area through the walls is 2.25×10^{-2} kcal/sec m^2. Find:
(a) the temperature of the inside surface of the wall and
(b) the temperature of the outside wall.

16. An uninsulated steam pipe 10 cm in diameter with an outside temperature of 95°C passes through a room. The absorption factor is 0.70, and the convection coefficient when the room is 25°C is 2×10^{-3} kcal/sec m^2 °C.
(a) What is the heat loss per meter of pipe by radiation?
(b) What is the heat loss per meter of pipe by natural convection?

5 Thermal Properties of Substances

5-1 EXPLANATION OF CHANGE IN STATE

We will now consider the various states of matter and the interconversions among them. If we apply heat to a crystalline solid in which the atoms or molecules are arranged in a regular pattern, the temperature, energy, and amplitude of atomic vibrations increase until the atoms can no longer hold together in their regular position. The atoms become free to move around and to slide over each other, and the solid changes to a liquid.

In the case of pure crystalline substances such as ice, the fusion or melting point is sharply marked. For a specific pressure, there is a definite temperature above which the substance is wholly liquid and below which it is solid. In the case of amorphous substances, such as glass, there is no definite melting point. Such substances are pliable over a range of temperatures. Figure 5-1 is a temperature-time graph for both types of substances when heat is being added to them. The flat portion of curve (b) in Figure 5-1 represents the time interval during which the change in state occurs. The constant temperature indicated by this flat portion is the temperature corresponding to the change in state, for example, melting point. Curve (a) of Figure 5-1 does not have a flat portion, consistent with the fact that there is no definite melting point. If a substance expands on solidifying, an increase of pressure will lower its melting point; if it contracts on solidifying, an increase in pressure will raise its melting point. This is an example of Le Chatelier's principle which says that a system will react to applied forces by assuming a configuration that seeks to minimize the forces applied. Thus, applying a pressure to a solid causes a decrease in volume. Since the

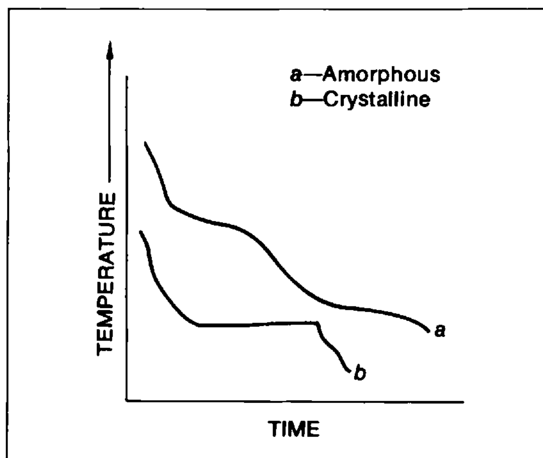


Figure 5-1 Temperature-time graph for a crystalline and amorphous substance.

volume decreases upon melting for the first case, an increase in pressure lowers the melting point and vice versa for the second case. The pressure required is quite large for any significant change in the melting point. Consider a wire with a weight on it hung around a block of ice, as illustrated in Figure 5-2. The ice directly below the wire melts because its melting point is lowered below the temperature of the surrounding ice. The water resulting from the melted ice flows to the top of the wire and freezes again into ice because of the reduced pressure.

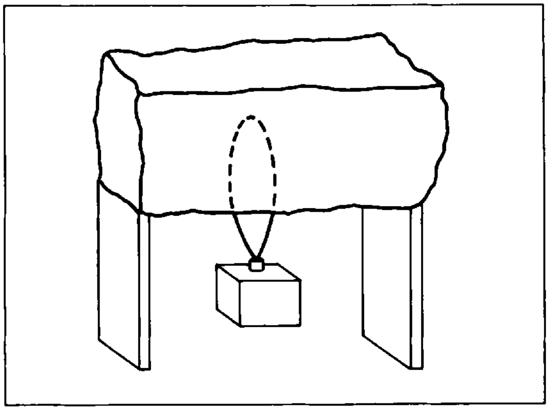


Figure 5-2 Wire passing through block of ice.

Eventually, the wire will pass through the block of ice leaving it intact. The process of melting under pressure and freezing again, as in the above example, is called *regelation.* Another example of regelation is ice skating. The ice beneath the skates melts because of the pressure due to the weight of the skater. The skater therefore glides on a thin film of water which freezes again behind him.

In the change in state from liquid to vapor, the process is different than from solid to liquid. In a liquid, not all the molecules have the same energy. There are always a few which have energies and velocities greater than the average, just as there are a few with energies and velocities that are lower. When a liquid is exposed to the open air, some of these molecules with high energy escape from the liquid surface and may be carried away by air currents. Any liquid exposed to the air evaporates. Evaporation occurs at all temperatures. If, however, a liquid is put in a closed container, molecules leaving the liquid accumulate in the space above the liquid. Some of these vapor molecules will settle on the liquid surface and recondense. Soon, a steady or equilibrium state is reached, where the number of molecules leaving the liquid surface becomes equal to the number reentering it. When this equilibrium is reached, the space above the liquid is said to be saturated with vapor; the pressure of such a vapor is called the *saturated vapor pressure.* The rate of evaporation depends on the temperature, and the rate of recondensation depends on the vapor pressure. For this reason, there is a direct connection between the liquid temperature and the vapor pressure. The saturated vapor pressure of any substance is found to depend only on the temperature. It does not depend on the amount of vapor present. The saturated vapor pressures for water at various temperatures is given in Table 5-1. Figure 5-3 is a typical saturated vapor pressure versus temperature graph for a substance. The end of the vaporization curve is called the *critical point.* Above the critical point, no distinct liquid-to-vapor phase transition can be detected. The temperature and pressure at which this happens are called the critical temperature (t_c) and critical pressure (p_c), respectively. Table 5-2 gives the critical temperatures and pressures for some common gases.

Since the molecules of higher energy escape from the liquid, the average molecular energy of the remaining liquid is decreased. Therefore, the liquid is cooled by evaporation unless heat is supplied to compensate for this energy loss.

A liquid boils in the open atmosphere at a temperature for which the vapor pressure just above the liquid surface is equal to the atmospheric pres-

Table 5-1 Pressure of Saturated Water Vapor.

Temperature (°C)	Pressure (mm Hg)	Temperature (°C)	Pressure (atm)
0	4.58	100	1.0
5	6.51	110	1.41
10	8.94	120	1.96
15	12.67	140	3.57
20	17.5	160	6.10
30	31.8	180	9.90
40	55.1	200	15.4
50	92.5	220	22.8
60	149.0	250	39.3
70	234.0	300	84.8
80	355.0	350	163.2
90	526.0	374	218.4

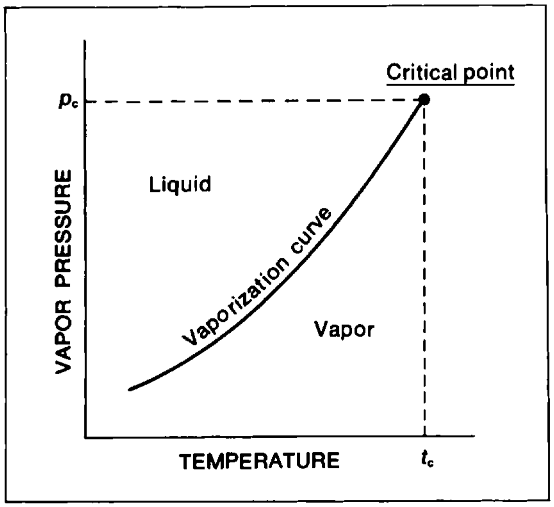


Figure 5-3 Vapor pressure versus temperature.

Table 5-2 Critical Temperatures and Pressures for Various Gases.

Substance	Critical temperature (in °C)	Critical pressure (in atmospheres)
Air	− 140.7	37.2
Ammonia	132.4	111.5
Argon	− 122.0	48.0
Carbon dioxide	31.0	73.0
Helium	− 267.9	2.26
Hydrogen	− 240.0	12.8
Nitrogen	− 147.1	33.5
Oxygen	− 118.8	49.7
Water	374.0	218.4

sure. Figure 5-3 is, therefore, also a graph of the variation of the boiling point with external pressure. The higher the external pressure, the higher the boiling point must be. This is why water boils at a higher temperature in a pressure cooker. The boiling points of various substances at standard atmospheric pressure were given in Table 2-2 of Chapter 2.

5-2 PRESSURE-TEMPERATURE DIAGRAM FOR A PURE SUBSTANCE

We have seen that the vapor pressure curve represents the pressure and temperature at which the liquid and vapor phases exist in equilibrium with

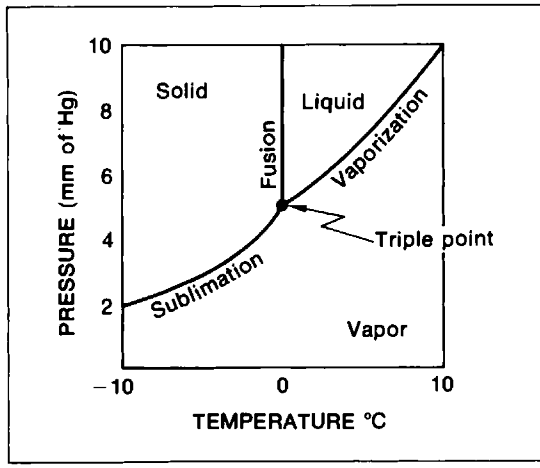


Figure 5-4 Triple point diagram for water.

one another. Similar curves representing the pressure and temperature at which the solid and liquid and the solid and vapor are in equilibrium can be drawn. For a suitable range of temperature and pressure, a single pressure-temperature diagram will show all three curves. These three curves intersect at one point—called the *triple point.* As a result, the diagram is usually given the name *triple point diagram.* Figure 5-4 is such a diagram for water. The curves representing the pressure and temperature at which the solid and liquid phase are in equilibrium is called the *fusion* or *melting point curve,* the solid and vapor in equilibrium—the *sublimation curve,* and the liquid and vapor in equilibrium—the *vaporization* or *boiling point curve.* The point at which all three phases can coexist in equilibrium is, of course, the *triple point.*

While the pressure corresponding to the triple point of water lies well below atmospheric pressure, this pressure for some substances, such as CO_2, lies above atmospheric pressure. At atmospheric pressure, CO_2 sublimes or, in other words, passes directly from the solid to the vapor at − 78°C. Solid CO_2 is commonly known as dry ice. It is used to keep things cool; it is more desirable than ordinary ice because it is less messy since it does not go through the liquid state.

5-3 REAL GASES

Real gases do not obey Boyle's law or the general gas law perfectly. This is best illustrated by a $p - V$ diagram (Figure 5-5) for a real gas. Suppose the gas

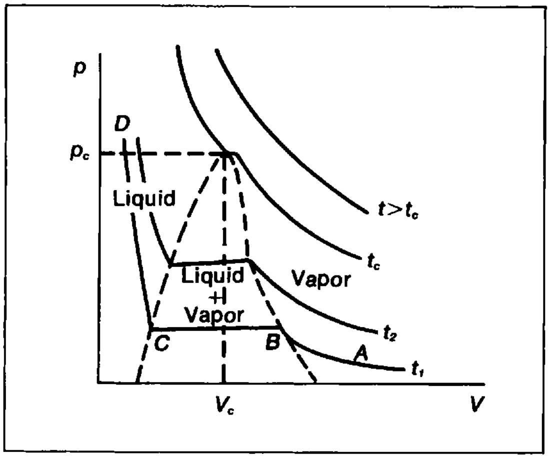


Figure 5-5 Pressure-volume diagram for a real gas.

is compressed at a temperature t_1 along the curve $ABCD$ in the diagram. If the pressure is low, AB in the diagram, the curve follows closely that of an ideal gas, and can be fitted approximately by Boyle's law. When point B is reached, however, a sharp change occurs in the nature of the curve and the volume continues to decrease without further increase of pressure (BC in the diagram). This is because, at point B, the gas begins to condense and gradually changes to a liquid along BC. At point C, all the gas has completely changed to liquid and a large increase in pressure is required for a small decrease in volume (CD in the diagram).

At a higher temperature, t_2, condensation begins at a higher pressure and at a smaller volume. This is because the effect of the attractive forces between the molecules is smaller since the molecules are moving faster at a higher temperature. Finally, if the temperature is increased sufficiently, a temperature will be reached where the thermal agitation will be so rapid that condensation into the liquid state is prevented. This is, as we stated in the previous section, the critical temperature (t_c on the diagram).

The dashed line and the critical temperature curve divide the $p - V$ diagram of Figure 5-5 into four regions. Below the dashed curve, the substance is a mixture of liquid and vapor in equilibrium. Below the critical temperature to the right of the dashed curve, the substance is a vapor or gas; to the left of the dashed curve, it is a liquid. Above the critical temperature curve, the substance is a gas.

The general gas law, which applies to an ideal gas, does not take into account the forces between the molecules of the gas. These forces vary with pressure, volume, and temperature of the confined gas. There are two types of intermolecular forces, repulsive and attractive. The repulsive forces are short-range forces which become strong when the molecules come very close together. The attractive forces are of longer range and are weaker than the repulsive forces. They are usually called van der Waal's forces. Van der Waal studied the character of intermolecular forces and developed an equation of state for a real gas. This equation takes into account the volume of, and the forces between, the molecules. The equation is

$$\left(p + \frac{a}{V^2}\right)(V - b) = RT. \qquad (5\text{-}1)$$

The quantities a and b are constants for a particular gas. The term a/V^2 takes account of the intermolecular forces, and the term b is proportional to the volume occupied by the molecules themselves.

5-4 EXPANSION OF GASES

If a gas is compressed in a cylinder fitted with a piston, work is done on the gas, and the gas becomes heated. If, on the other hand, a gas is allowed to expand, the gas does work in pushing the piston out against atmospheric pressure, and the gas is cooled.

Consider the following question: If the gas were to expand without doing work against atmospheric pressure, that is, expand freely into a vacuum, would the gas cool? In other words, does the thermal energy of a gas change on free expansion? The answer is no for an ideal gas. The thermal energy of an ideal gas is independent of volume and pressure changes, provided the temperature does not change. However, there is no ideal gas. The thermal energy of a real gas depends on pressure as well as on temperature. At temperatures not too far below the critical temperature, a free expansion causes a decrease in temperature of the gas.

5-5 LIQUEFACTION OF GASES

All gases can be liquefied. Air has a critical temperature of $-140°C$, and cannot be liquefied by merely compressing it at room temperature. The following is a brief description of one commercial method used for the liquefaction of air. It is known as the Linde process, and a simple drawing of the apparatus used in this process is shown in Figure 5-6. Air is compressed in the compressor, circulated through the coils in the cooler, and allowed to expand through a small opening at *O*. As a result of the expansion, the air cools further. This cooled air is again compressed and circulated through the coils in the cooler so that the approaching air becomes progressively cooler before expansion. By the continuous operation of this cycle, the expanding air is finally cooled to the point where it liquefies and is collected at the bottom of the container below the expansion tube.

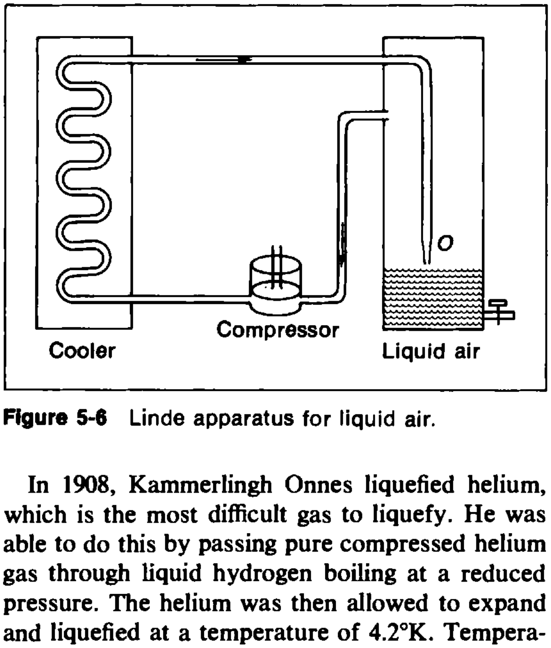


Figure 5-6 Linde apparatus for liquid air.

In 1908, Kammerlingh Onnes liquefied helium, which is the most difficult gas to liquefy. He was able to do this by passing pure compressed helium gas through liquid hydrogen boiling at a reduced pressure. The helium was then allowed to expand and liquefied at a temperature of 4.2°K. Temperatures as low as 0.7°K can be obtained by the rapid evaporation of liquid helium. Temperatures approaching absolute zero have been reached by demagnetizing magnetic materials which were initially at liquid helium temperatures.

All gases, except helium, which have first been liquefied can subsequently be solidified by pumping away the vapor above the liquid. The solid state of helium only exists under high pressure.

5-6 ATMOSPHERIC HUMIDITY

Water vapor is present in the air. Water vapor is lighter than air. When water is evaporated into the air, it displaces a volume of air equal to its own volume. The term humidity is used to describe the water vapor content of the atmosphere. The actual vapor pressure at any time or place cannot exceed the saturation pressure for the existing temperature (Figure 5-3), for condensation begins to take place as soon as this value is reached. The higher the temperature, the more water vapor the air is capable of holding. When warm air is cooled, some of the water vapor condenses and reappears as water. The temperature at which the water vapor in the air is sufficient to saturate it is called the *dewpoint.*

The measurement of humidity is called *hygrometry.* In hygrometry, either the absolute humidity or relative humidity is measured.

The ***absolute humidity*** is the mass of water vapor per unit volume in the atmosphere at a given temperature. ***The relative humidity*** is the ratio of the amount of water vapor present in the atmosphere to the amount required to saturate it at the same temperature.

Since the vapor pressure is proportional to the mass of vapor present, the relative humidity is equal to the ratio of the actual pressure of the vapor present to the saturation vapor pressure at the same temperature. It is usually expressed in percent:

$$\text{relative humidity (\%)} = \left(\frac{p}{p_m}\right)(100), \qquad (5\text{-}2)$$

where p equals the vapor pressure of the water vapor present in the air, and p_m equals the vapor pressure of the water vapor if the air were saturated at the same temperature.

Example 1. Find the relative humidity if the vapor pressure of the water vapor in the air is 5 mm of mercury at a temperature of 10°C.

SOLUTION

From Table 5-1, the saturated water vapor pressure at 10°C is 8.94 mm of mercury. Therefore,

$$\text{relative humidity} = \left(\frac{5 \text{ mm of Hg}}{8.94 \text{ mm of Hg}}\right)(100) = 56\%.$$

There are various ways of measuring the relative humidity. One of the most accurate ways is by determining the dewpoint. This temperature may be found by partly filling a brightly polished container with water and dropping pieces of ice into it, while stirring it with a thermometer. The dewpoint is the temperature indicated by the thermometer when moisture first appears on the polished surface. The saturated vapor pressure of water at the dewpoint is a measure of the water vapor present in the air. Therefore, if we know the dewpoint and the temperature of the air, we can calculate the relative humidity.

Example 2. If the room temperature is 20°C and the dewpoint is 15°C, find the relative humidity.

SOLUTION

From Table 5-1, the saturated vapor pressures at 15°C and 20°C are 12.67 and 17.5 mm of mercury, respectively. Therefore,

$$\text{relative humidity} = \left(\frac{12.67\ \text{mm of Hg}}{17.5\ \text{mm of Hg}}\right)(100) = 72\%.$$

Another way of measuring the relative humidity is with the *wet and dry bulb hygrometer,* which makes use of the principle of cooling by evaporation. It consists of two mercury thermometers placed side by side. The bulb of one thermometer is kept dry, while that of the other is kept continually wet by a piece of porous cloth attached to a wick which dips into a container of water as illustrated in Figure 5-7. The temperature of the wet bulb thermometer will read lower than that of the dry bulb thermometer because of the evaporation taking place at its surface. If the air is already saturated, no evaporation takes place and both thermometers give the same reading. Thus, the drier the air, the greater the difference in the readings of the two thermometers. Tables have been constructed permitting the relative humidity and dewpoint to be determined from the readings of the two thermometers.

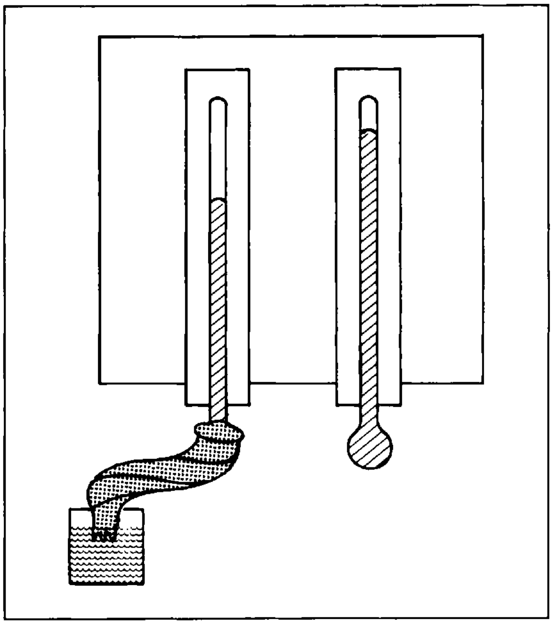


Figure 5-7 Wet and dry bulb hygrometer.

PROBLEMS

1. For a given substance, it is observed that the melting point decreases with an increase in pressure.
   (a) Can the substance be liquefied?
   (b) Can the substance be solidified?
   (c) Is the heat of fusion positive or negative?
   Give reasons for your answers.
2. An air conditioning unit works in the summer time by cooling the air below the desired temperature to condense out excess water vapor, and then reheating the air. If the outside air is at 86°F and 70% relative humidity, and the delivered air is at 68°F, to what temperature must it be cooled to lower the water content in the delivered air to 50% relative humidity?
3. The measured volume of a quantity of hydrogen collected over water is 780 cm^3, the temperature being 16°C and the barometer reading 740 mm. The volume is measured with the water level the same inside and outside the bottle. If the quantity of hydrogen collected has a volume of 705 cm^3 at 0°C and 1 atmosphere pressure when dried, calculate the vapor pressure of saturated water vapor at 16°C.
4. The temperature in a room is 20°C. A can containing water is gradually cooled by adding ice to it. At 10°C, the surface of the can clouds over. What is the relative humidity in the room?

5. In the morning, a hygrometer in a room at a temperature of 20°C shows the relative humidity to be 37%. What is the dewpoint? Later on in the day, the hygrometer is broken, making it necessary to determine the relative humidity by an alternative method. The temperature in the room remains at 20°C. The dewpoint is measured by the method given in Problem 4 and found to be 10°C. What is the relative humidity?
6. A small amount of liquid is put in a pyrex glass tube. The tube is evacuated and sealed off. Describe the behavior of the liquid surface when the temperature is raised
   (a) if the volume of the tube is much less than the critical volume,
   (b) if the volume of the tube is much greater than the critical volume.
7. The pressure of water vapor in equilibrium with liquid water is given approximately by

$$p \text{ (atmospheres)} = 10^{-4}t^2$$

   for t between 10°C and 100°C. The volume of a cylinder completely filled with water at 20°C is doubled by partly withdrawing a tight fitting piston. After equilibrium is restored, the temperature is 10°C. What is the pressure of the water vapor which now fills one-half of the cylinder?
8. What pressure must be maintained in a pressure cooker to boil water at 110°C, when the ordinary boiling point is 100°C?
9. What is the maximum value that could be obtained for the melting point of ice?

6 Reflection and Refraction at a Plane Surface

6-1 INTRODUCTION

In the next eight chapters we will discuss optical phenomena. A truly complete understanding of this subject becomes possible only after three basic questions have been answered:

What is the nature of light?
How does it propagate from point to point?
How does the presence of matter affect its propagation?

Full answers for these questions would necessarily involve a discussion of a major part of physics as we understand it today. We are not prepared for such detail at present, so what follows in the next several chapters must be regarded as incomplete.

On the other hand, much of our present knowledge of light and its behavior was already known to early civilizations (although our present understanding is due mainly to the work of scientists from the seventeenth to the nineteenth century). It is therefore meaningful to consider first the more readily observed characteristics of light and some practical applications.

6-2 RECTILINEAR PROPAGATION

In this chapter and the two that follow, we shall be concerned with phenomena that can be explained by noting that a narrow beam of light traveling in an isotropic medium (one whose properties are the same in all directions) will travel in a straight line—a property called rectilinear propagation. This is consistent with our experience that objects illuminated by point sources of light apparently cast sharply defined shadows, and that we cannot see around corners.

Most common sources of light are extended sources, not point sources. By using appropriate apertures, such a source can be used to define a narrow pencil of light.† If the cross section of the pencil is reduced to the ideal limit of a point, the beam which results is called a ray. The concept of a light ray is very useful, for it allows us to use straight lines and geometrical constructions to discuss on paper the behavior of light as it reaches the boundary of an isotropic medium or passes from one isotropic medium to another of different optical character. This method of discussing optical phenomena is called geometrical optics; we will use it in the following pages to discuss mirrors, lenses, and optical instruments.

†This statement does not apply to laser light, which is discussed in chapter 11.

6-3 REFLECTION

When a ray of light traveling in an isotropic medium arrives at a perfectly smooth boundary surface of a second medium, two different situations can occur. The ray may be reflected from the boundary surface, thus remaining in the first medium, or it may enter the second medium, in which case it is said to be refracted. The extent to which either of the two situations predominates depends upon several factors to be discussed later.

Reflection from a smooth, uniform surface is termed regular or specular reflection. The laws governing the situation are easily stated, and can be readily verified with a small hand mirror or similar reflector.

The laws of specular reflection are as follows:

1. The ray incident upon the smooth boundary, the ray reflected from the boundary, and a perpendicular line (the normal) drawn to the plane containing the boundary (reflecting surface) at the point of incidence all lie in a single plane.
2. The angle between the incident ray and the normal (the angle of incidence, i) is equal in magnitude to the angle between the reflected ray and the normal (the angle of reflection, r).

Figure 6-1 illustrates these laws and the quantities introduced in them. Thus, Law 1 makes it possible to draw a two-dimensional diagram, while Law 2 is a statement of the symmetry of the reflection process.

If the pencil of light is large in cross section, and it strikes a surface which is very irregular (for example, a crumpled sheet of aluminum foil), the laws of reflection will not seem to be successful in predicting the paths of the reflected light rays, which will be visible in all directions. This type of reflection is called diffuse reflection. To understand the apparently random directions of the reflected rays we note the following: If a normal to the irregular reflecting surface is constructed at the point of incidence of each (parallel) ray of a small pencil of light, it will be observed that the direction of each reflected ray will satisfy the laws of reflection. Therefore, we can see that it is sufficient to study specular reflection in detail, since diffuse reflection can be reduced to the specular case by suitably limiting either the cross section of the light pencil or the extent of the reflecting surface.

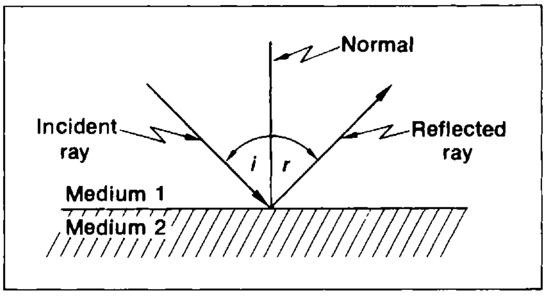


Figure 6-1 Reflection at a plane boundary.

6-4 THE FORMATION OF IMAGES BY A PLANE MIRROR

The image formed by a plane mirror of an extended illuminated object can be constructed by applying the laws of reflection to individual points on the object. A point source of light, in the absence of pin holes, will send out light rays in all directions. As seen in Figure 6-2, of the many rays striking a reflecting surface, only those intercepted by the observer's eye need be drawn. The reflected rays, extended back to their point of intersection, form an image of the source at I. To the observer, the light rays seem to come by rectilinear propagation directly from this image. Because the image is behind the reflecting surface where no real light rays exist, it is only an apparent or virtual image. Thus, it would not appear on a screen placed at I, although it certainly seems real to an observer. Since $i = r$

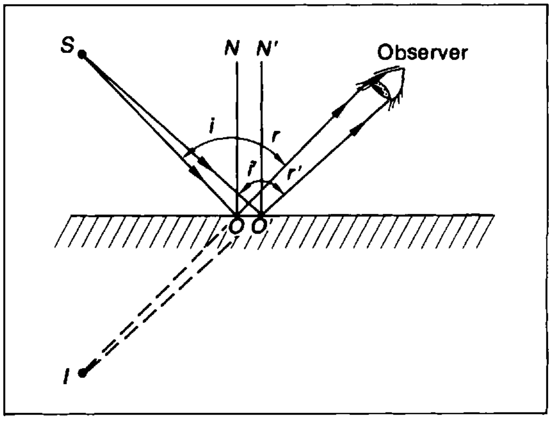


Figure 6-2 Image of a point source.

and $i' = r'$, triangles SOO' and IOO' are congruent (why?), which establishes the fact that S and I are symmetrically located with respect to the mirror. It is left as a problem to show that an extended object in front of a plane mirror will have a virtual image equal in size to the object and symmetrically located with respect to the mirror.

6-5 REFRACTION

A monochromatic (single color) light ray that is refracted at a plane boundary between two different isotropic media obeys the laws of refraction. These are as follows:

1. The incident ray, the refracted ray, and the normal to the plane boundary at the point of incidence lie in a plane.
2. The ratio of the sine of the angle of incidence to the sine of the angle of refraction is a constant, which is independent of the angle of incidence and depends only upon the two media and the color of the light used.

These laws are illustrated in Figure 6-3.

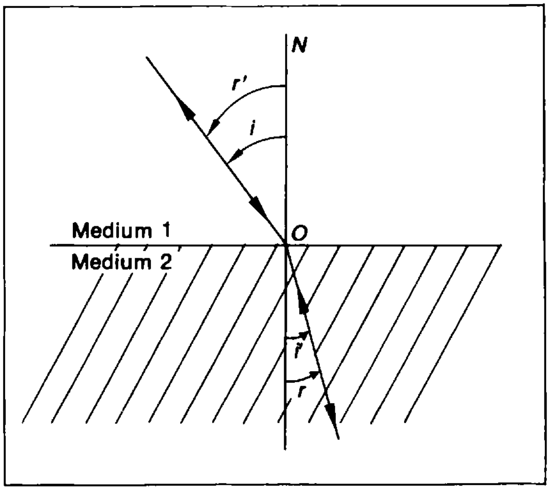


Figure 6-3 Refraction at a plane boundary.

The second law is called Snell's law after Willebrord Snell, who discovered it in 1621. Expressed in equation form, Snell's law reads

$$\frac{\sin i}{\sin r} = n_{21}, \qquad (6\text{-}1)$$

where n_{21} is a color dependent constant called the relative index of refraction of medium 2 with respect to medium 1. If the roles of the incident and refracted rays are reversed, it is found experimentally that the diagram remains the same, and only the directions of the rays are reversed.† In this case, Snell's law states that

$$\frac{\sin i'}{\sin r'} = \frac{\sin r}{\sin i} = n_{12}, \qquad (6\text{-}2)$$

and we conclude that

$$n_{12} = \frac{1}{n_{21}}. \qquad (6\text{-}3)$$

We can show that the relative index of refraction of one medium relative to another can be determined at once, providing the indices of refraction of the two media relative to a third are both known. If a ray of light enters a slab of material with parallel plane sides, the ray emergent from the slab is directed parallel to the incident ray but displaced laterally by an amount depending upon the thickness of the slab, the angle of incidence, and the relative index of refraction. By successively adding other plane parallel slabs, the ray emergent from the final slab will still be parallel to the incident ray, and the total lateral displacement is simply the sum of the individual displacements due to each slab. Figure 6-4 shows this situation (where $\phi_1 = \phi_4$).

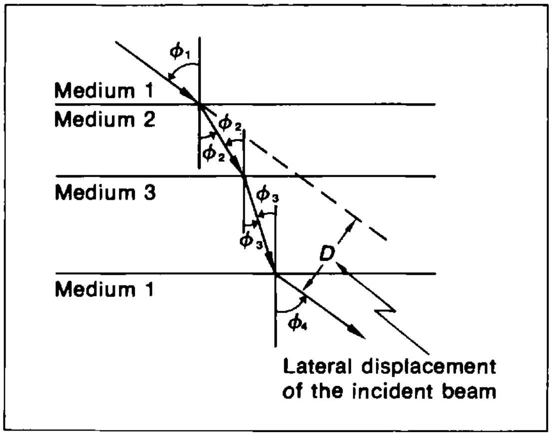


Figure 6-4 Multiple lateral displacement of a light ray.

†The reflected rays are *not* the same when the incident and refracted rays are reversed. See Section 6-6.

Repeated application of Snell's law yields

$$n_{21} = \frac{\sin \phi_1}{\sin \phi_2}, \quad n_{32} = \frac{\sin \phi_2}{\sin \phi_3}, \quad n_{13} = \frac{\sin \phi_3}{\sin \phi_4} = \frac{\sin \phi_3}{\sin \phi_1}.$$

Multiplication of these equations yields $n_{21}n_{32}n_{13} = 1$, from which it follows that

$$n_{21} = \frac{1}{n_{32}n_{13}} = \frac{n_{23}}{n_{13}} \qquad (6\text{-}4)$$

because of Eq. (6-3).

It is customary, when quoting indices of refraction, to compare a given material with a vacuum, which by convention has been adopted as a standard or reference medium. The index of refraction of a medium with respect to a vacuum is then identified as the absolute index of refraction, or simply the index of refraction of the medium. When the term index of refraction is used in this book, it is implicit that the comparison medium is the vacuum and that, hence, a second subscript is not necessary. Table 6-1 lists some approximate values for indices of refraction of various materials. It can be seen that the index of refraction is qualitatively related to the density of the material. Thus, for example, there is little difference between the various gases and a vacuum. One might, therefore, speculate that the index of refraction indicates the extent to which a given material interferes with or impedes the passage of light. More dense materials would be expected to give more effective interference than those of lesser density since there are more individual atoms or molecules to act. We shall return to this point in a later chapter.

Table 6-1 Typical Indices of Refraction.

Substance	Index of refraction
Dry air (STP)	1.0003
CO_2 (STP)	1.0004
Water	1.33
Alcohol	1.36
Glass	1.50–1.70
Quartz (fused)	1.46
Diamond	2.42

A more symmetric form of Snell's law is obtained if we use absolute indices of refraction as exhibited in Eq. (6-4). Thus, since

$$\frac{\sin \phi_1}{\sin \phi_2} = n_{21} = \frac{n_2 \text{ vacuum}}{n_1 \text{ vacuum}} = \frac{n_2}{n_1},$$

we can write

$$n_1 \sin \phi_1 = n_2 \sin \phi_2. \qquad (6\text{-}5)$$

It should be noted that the index of refraction is a function of the color of the light ray involved.

6-6 THE CRITICAL ANGLE AND TOTAL INTERNAL REFLECTION

A medium is said to have an optical density greater or less than that of a second medium if its index of refraction is greater or less than that of the second medium. If a ray of light goes from an optically less dense medium to an optically more dense medium ($n_2 > n_1$), we see from Snell's law that the refracted ray will be bent toward the normal. Since the angle of incidence in medium 1 cannot be greater than 90°, it follows that the angle of refraction can take on values between 0° (for $i = 0°$) and a maximum value θ_c (for $i = 90°$).

This latter angle, called the critical angle, has the following significance. Consider a series of light rays in medium 1 with incidence angles between 0° and 90°. The refraction angles in medium 2 will be as shown in Figure 6-5(a). On the other hand, if we now consider, as in Figure 6-5(b), a series of light rays originating in medium 2 with angles between 0° and 90°, it is clear that Snell's law cannot be satisfied when the angle exceeds the critical angle since it would imply $|\sin \theta| \geqslant 1$. The physical explanation of this situation is that the "refracted" ray for $i > \theta_c$ is not refracted at all. Instead, it is reflected back into medium 2 from the interface and, therefore, will obey the laws of reflection. It can be shown experimentally that rays originating in the optically denser medium will experience both reflection and refraction at an interface with a less dense medium. The fraction of the incident light that is refracted is greatest at $i = 0°$, and will be equal to 0 for angles equal to or greater than θ_c. This latter situation is called total internal reflection.

6-7 DISPERSION

In Section 6-5, it was stated that the index of refraction of a medium is a function of the color of light in the medium. A schematic illustration of this relationship is shown in the graph of Figure 6-6. Table 6-2 gives a set of typical values of the index of refraction for various colors. In Chapter 11, a more quantitative method of distinguishing colors

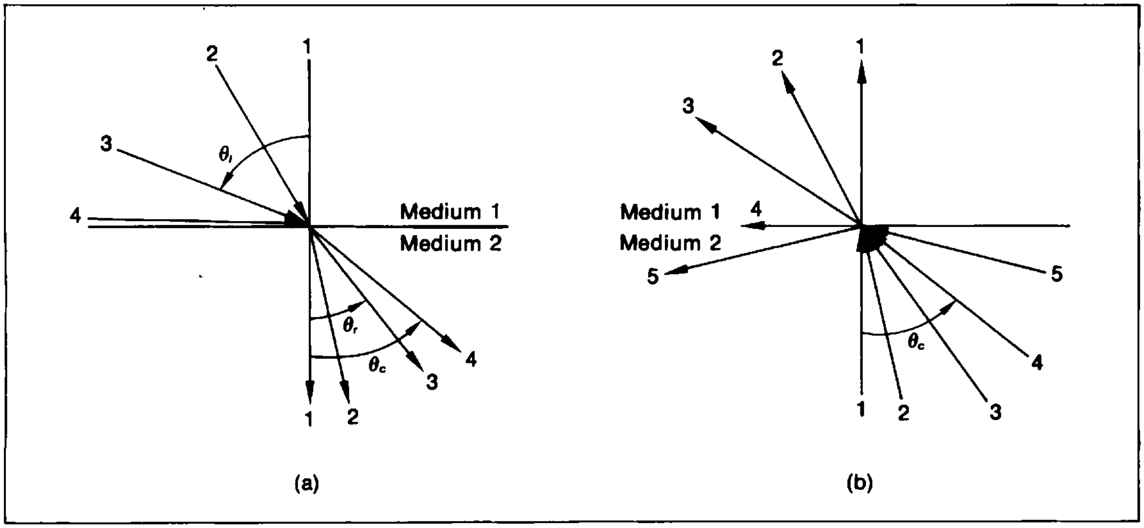


Figure 6-5 The critical angle and total internal reflection.

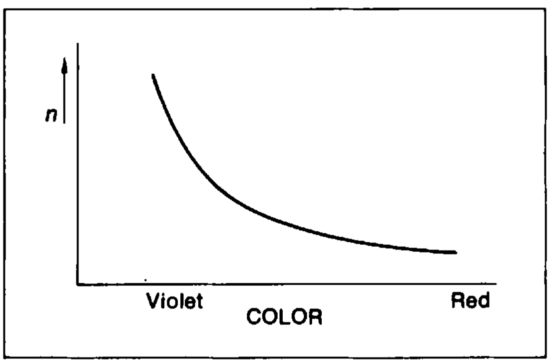


Figure 6-6 Variation of index of refraction with color.

Table 6-2 Color Dependence of the Index of Refraction for a Typical Flint Glass.

Red	1.622	Blue	1.639
Yellow	1.627	Violet	1.662

will be presented. For the present, we merely note that red light is refracted less than blue or violet light. We note also that the values for the index of refraction in Table 6-1 must be regarded as approximate or average values since a white light source is assumed.

White light is a combination of all the various colors of light. The percentages of the various colors in a beam of white light depends markedly upon the nature of the source. Now let us consider what effect the variation of index of refraction with color will have upon the appearance of a ray of white light passing through a transparent substance with non-parallel sides—for example, a prism. The situation is illustrated by Figure 6-7.

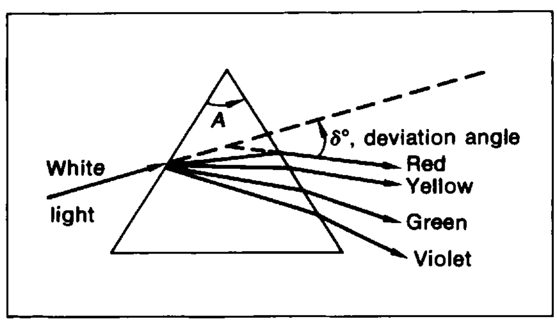


Figure 6-7 Dispersion of white light by a prism.

At the first surface, the rays of the various colors will be bent by an amount that increases from red to violet. In addition, the non parallel sides of the substance accentuate the separation of the various colors by providing different path lengths to be traversed by the separate rays. Finally, when these separated rays leave the transparent medium, the separation is further enhanced by a bending away from the normal that is greatest for violet and least for red light. Why? (It is implied here that the medium surrounding the prism of transparent material is a vacuum or, at any rate, a medium with an index of refraction lower than that of the prism.) It is clear from the preceding discussion that the deviation angles for the various colors depend not only upon the index of refraction but also upon the apex angle A of the prism. This "spreading-out" of the several colors in the ray of light is known as dispersion. If the constituent colors of a light ray are known, they can be used to deduce the refractive properties of a material. Conversely, if the refractive properties of the material are known, they can be used to analyze a beam of light whose chromatic makeup or color composition is unknown.

PROBLEMS

1. Show that a light ray reflected from a plane mirror will be rotated through an angle 2θ when the mirror is rotated through an angle θ. Hint: Consider the light ray originally at normal incidence and rotate the mirror.
2. Use the laws of specular reflection to show that an extended object in front of a plane mirror will have a virtual image equal in size to the object and symmetrically located with respect to the mirror.
3. A ray of light is incident at grazing incidence onto one face of a glass cube. The ray emerges from an adjacent face of the cube at an angle θ relative to the normal to that face. Show that $\sin\theta = \cot\theta_c$, where θ_c is the critical angle.
4. A circular disc floats on the surface of a liquid for which $n = 1.4$. What must its diameter be if a point object located 10 cm below the surface is not to be visible from above?
5. A ray of light strikes a plane sheet of glass ($n = 1.60$) that is 3 cm thick. The angle of incidence is 45°. Determine the lateral displacement D of the ray as it emerges from the glass. (See Figure 6-4.)
6. If a ray of light makes an angle of incidence of 53° at an air-diamond interface, what are the angles of reflection and refraction for the light?
7. When an object is under water, it appears to be closer to the surface than it actually is. Show that the apparent depth is equal to the actual depth divided by the index of refraction for the water, $n = 1.33$.
8. At what angle should a fish look up toward the surface of the water in order to see objects on the shore?
9. Two plane mirrors meet at an angle of 60°. A ray of light enters parallel to one mirror and strikes the other. By making a ray diagram, trace the ray through the system and tell how it leaves.
10. By means of ray diagrams, show the number of images that a person will see by standing in front of a pair of vertical plane mirrors joined along one edge and making an angle of 90° with each other. Consider both single and multiple reflections.
11. A ray of white light is incident on the face of a 60° prism so that the violet light traverses the prism parallel to the base of the prism as shown in Figure 6-8. The index of refraction for the violet light is 1.662. Calculate the deviation angle for violet light.

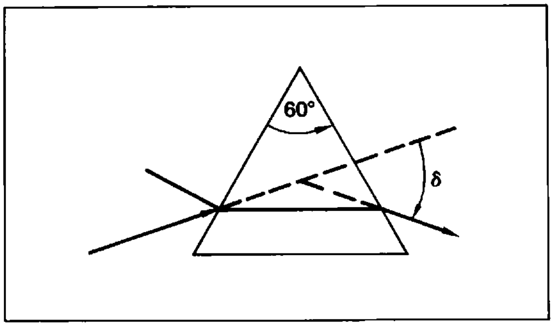


Figure 6-8

7 Reflection and Refraction at a Curved Surface

7-1 TERMINOLOGY AND SIGN CONVENTION

In the last chapter, the laws of reflection and refraction were applied to situations involving plane surfaces. The plane surface is a special case of a curved surface (a spherical surface with an infinite radius of curvature), so it is appropriate to consider next reflection and refraction for general curved surfaces. Experimental studies show that the laws of reflection and refraction as stated are sufficient, when applied to individual light rays, to provide a satisfactory analysis in all situations for which geometrical optics is appropriate.

It is found that the specific behavior of a given light ray depends upon the radius of curvature of the optical surface (the interface between two media of differing indices of refraction). Thus, a concave mirror with its center of curvature in front of the reflecting surface as in Figure 7-1(a) will tend to focus or bring into convergence a beam of light rays parallel to the axis of symmetry of the mirror. On the other hand, a convex mirror, with its center of curvature behind the reflecting surface as in Figure 7-1(b), will tend to defocus or cause a divergence of a similar light beam. Even though the magnitude of the two radii of curvature may be the same, the effects on a beam of light are markedly different.

The location of the image of an object providing the oncoming light beam is also affected by the radius of curvature. For a concave mirror the light rays converge to a real image at a point in front of the mirror. The image formed by a convex mirror is virtual since the reflected rays appear to be diverging from an image point behind the mirror where no light rays actually exist. To distinguish between these two cases, it is reasonable to identify object and image distances and radii of curvature as either positive or negative. (We shall see that the same designations are appropriate for all other distances relevant to reflection and refraction situations.)

Therefore, it is desirable to introduce a choice of signs for each of the several distances or dimensions to be encountered in any situation before beginning an analysis of any specific case. Such a set of choices is called a sign convention. The reader can, by a rapid survey of physics textbooks, discover that there are a number of different sign conventions in use. The effect of a given sign convention is to label some distances positive, others negative. Another sign convention may produce different signs for some or all of the distances involved, but the magnitudes of all distances will

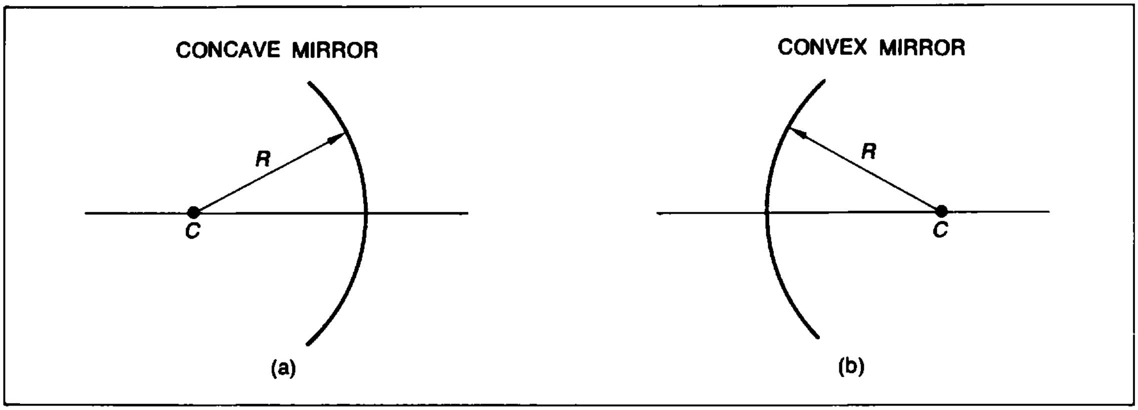


Figure 7-1

remain unchanged. The interpretation of algebraic results, however, can be extremely confusing for the reader who turns to one or more additional texts while studying geometrical optics. It is, therefore, mandatory that in consulting any text on geometrical optics one must first determine the sign convention adopted if confusion is to be minimized.

In this text we provide a simple sign convention and apply it to the case of reflection by way of illustration. We shall see, however, that it can be applied consistently to cases of refraction as well, with similar analytical results and interpretation.

The sign convention for geometrical optics is as follows:

1. All distances are to be measured from the optical surface to the point in question along the axis of the optical surface.
2. The distance from the surface to the object (the object distance p) is positive if the direction traveled is opposite to that of the approaching light.
3. The distance from the surface to the image (the image distance q) is positive if the direction traveled is the same as that of the departing light.
4. Any radius of curvature (R) is positive if the direction from the surface to the center of curvature is the same as that of the departing light.
5. Any object or image dimension above the axis of the optical surface is positive; below the axis—negative.

As used in geometrical optics, the term axis refers to the axis of symmetry of the optical surface. As an example, consider the heavy horizontal line in either Figure 7-1(a) or (b). An axis of symmetry implies a conic surface, and one commonly encounters discussions of spherical and paraboloidal optical surfaces. The reflection or refraction of light is physically less complicated for the latter case than for the former case. To see this, refer to Figure 7-2, where the laws of reflection are applied to the individual light rays striking a spherical [Fig. 7-2(a)] and a paraboloidal [Fig. 7-2(b)] reflecting surface. In the latter case, all the reflected rays cross the axis at a common point, called the focal point. For the spherical surface, the reflected rays cross the axis at a point which is dependent upon the vertical distance from the axis of the incoming horizontal ray. This is known as spherical aberration.

Thus, there is no unique focal point for a spherical reflecting surface, unless one restricts the reflection from the spherical surface to a small central portion. When this is true, the central part of the spherical surface approximates the exact curvature of the paraboloidal surface. It is desirable to consider spherical surfaces in preference to the paraboloidal surface because of the mathematical simplification thus afforded. In the following discussions, therefore, references to spherical surfaces must be understood to refer to centrally limited spherical surfaces.

In many situations, one does not have a beam of light rays parallel to the axis of a surface upon which the beam is incident. For non-parallel rays, the analysis is complicated by the appearance of trigonometric functions. To avoid this difficulty, one generally considers situations for which incident and reflected (or refracted) rays all make

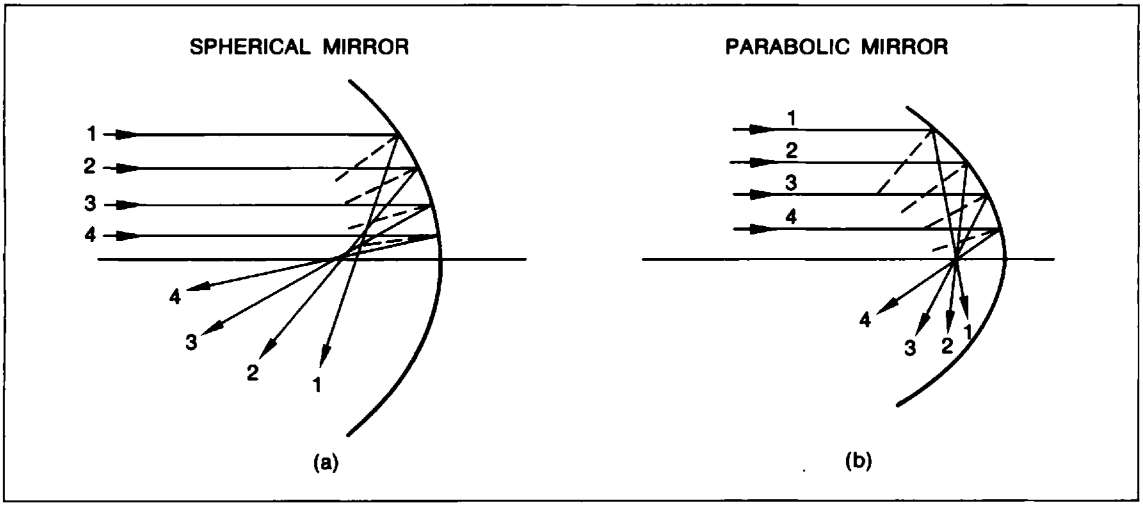


Figure 7-2 Reflections of parallel light rays.

angles relative to the axis of the surface which are small enough to permit use of the approximate relations†

$$\sin \theta \approx \theta$$
$$\tan \theta \approx \theta,$$

where θ is measured in radians. For angles less than 0.2 radians ($\approx$ 12°), these relations are in error by less than 1%; thus, relatively large angles ($\lesssim$ 12°) can be considered. Rays satisfying these relations are called *paraxial* rays since they are only approximately parallel to the axis of the surface. In the analytical solutions that follow, we shall see that the use of paraxial rays makes possible greatly simplified discussions of reflection and refraction from curved surfaces.

7-2 REFLECTION FROM A SPHERICAL SURFACE—RAY TRACING METHOD

It was stated in the last chapter that to analyze any reflection it is sufficient to break up the beam of light into individual rays and apply the laws of reflection to each of them. The ray tracing method does this for a set of rays that have properties lending themselves to easy graphical construction. The paths of three such rays are shown in Figure 7-3. Constructing normals to the spherical surface (radii) at the point of intersection of the light rays with the surface (neglecting spherical aberration), and using the laws of reflection, establishes the fact that these rays have the following properties:

1. A ray parallel to the axis upon reflection from a converging surface crosses the axis at the focal point, which is at a point halfway between the center of curvature and the center of the surface, along the axis, or appears to come from the focal point for a diverging surface.
2. A ray passing through the center of curvature (or directed toward the center of curvature for a diverging surface) is reflected back upon itself since it is incident normally upon the surface.
3. A ray passing through the focal point (or directed toward the focal point for a diverging surface) is reflected parallel to the axis.
4. A ray striking the center of the surface will be reflected at an equal angle on the opposite side of the optic axis.

Notice that this construction yields not only the location of the image but also its size, orientation, and (by virtue of its location) whether it is real or virtual. As the figures indicate, two of the rays above are sufficient to determine the characteristics of the image. Use of the third ray or fourth ray, however, provides a desirable check on the accuracy of the construction.

†See the tables of Natural Trigonometric Functions in the Appendix. (Note that 2π radians = 360°.)

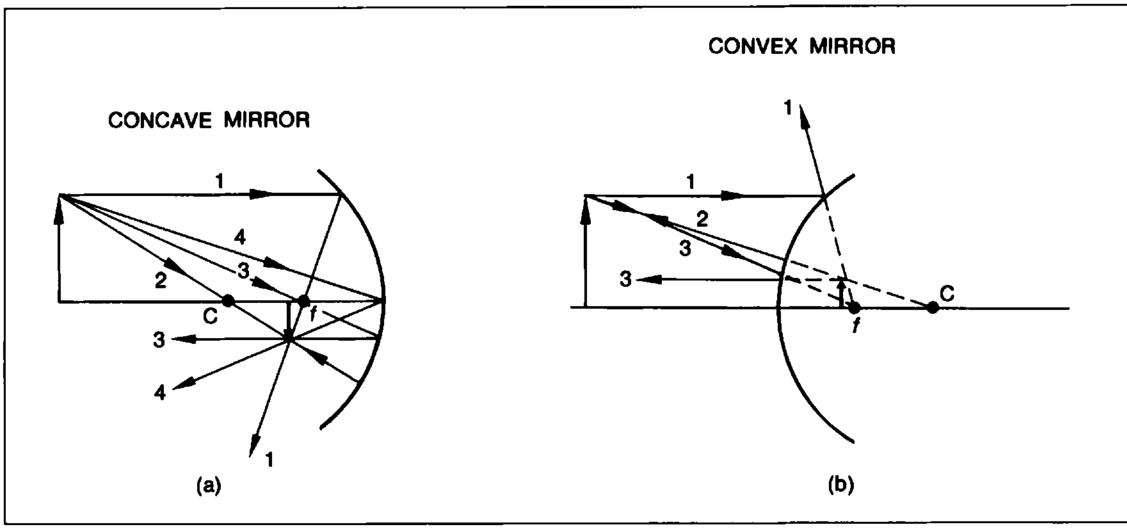


Figure 7-3 Graphical constructions for reflection.

7-3 REFLECTION FROM A SPHERICAL SURFACE—ANALYTICAL METHOD

Consider a point source on the axis of a concave (converging) mirror as in Figure 7-4. The center of curvature is at C, the object and image points are at P and Q, respectively, and O is called the vertex of the optical surface.

From the figure and the fact that the sum of the angles of a triangle is 180°, we see that

$$\beta = \alpha + \delta$$

$$\gamma = \beta + \delta,$$

from which

$$\gamma + \alpha = 2\beta. \tag{7-1}$$

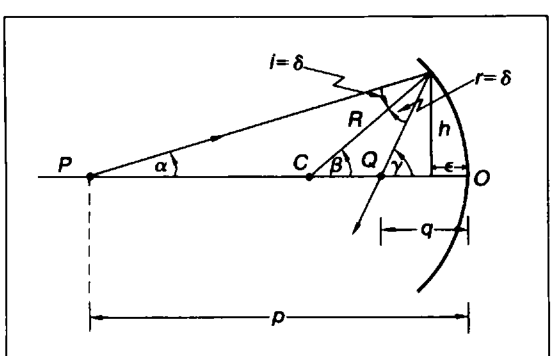


Figure 7-4 Geometry for the law of reflection.

Applying our sign convention to the figure, we obtain

$$\tan\alpha = \frac{h}{p-\epsilon}, \quad \tan\beta = \frac{h}{R-\epsilon}, \quad \tan\gamma = \frac{h}{q-\epsilon}.$$

As asserted earlier, these equations do not admit a simple solution unless the angles are small (paraxial rays). For such a case, we can equate the tangents to the angles and neglect the small distance ϵ, so that Eq. (7-1) becomes (for paraxial rays)

$$\frac{h}{p} + \frac{h}{q} = \frac{2h}{R}$$

or

$$\frac{1}{p} + \frac{1}{q} = \frac{2}{R} = \frac{1}{f}. \tag{7-2}$$

Notice that the definition of the focal point (the value of q for $p = \infty$) leads by means of Eq. (7-2) to the result that the focal length has the value $R/2$ as asserted earlier. Equation (7-2) also includes the special case of the plane mirror. In this case, $R = \infty$, $q = -p$, as in Problem 6-2.

Next we investigate reflection for an extended object. Figure 7-5 shows that unless the object height y is small, the distance from the top of the object to the surface will differ from the distance from the bottom of the object to the surface. This will produce a curved image, and this effect is another aberration of spherical surfaces, called *curvature of field.* We assume here that the object is limited in height so that this curvature is effectively

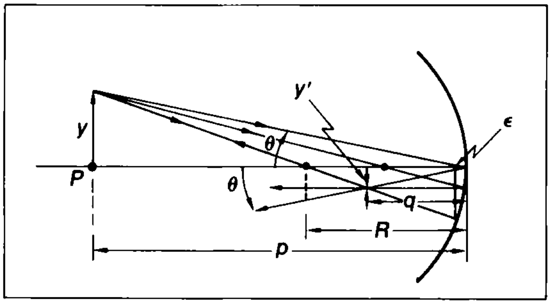


Figure 7-5 Reflection of an extended object.

absent. Since the angles of incidence and reflection are equal, we can write

$$\tan\theta = \frac{y}{p} = \frac{-y'}{q},$$

where the image height is negative by virtue of our sign convention. Defining the magnification m as the ratio of image to object height, we obtain

$$m = \frac{y'}{y} = \frac{-q}{p}. \qquad (7\text{-}3)$$

Notice that for this case (a real image) the magnification is negative since the object and image distances are positive, as is the object height, while the image height is negative. For a virtual image, the image distance would be negative and the magnification would therefore be positive. In this way, we see that the magnification gives information not only as to the increase or decrease in size of the image relative to the object, but also indicates its orientation. The point here is that the use of the sign convention together with Eqs. (7-2) and (7-3) are sufficient to analyze any reflection problem without recourse to a graphical construction.

7-4 REFRACTION AT A SPHERICAL SURFACE

Let us now consider the case of refraction at a spherical surface, illustrated in Figure 7-6. An object at point P in a medium of refractive index n has an image at point Q in a medium of refractive index n'. The radius of curvature of the optical surface is R. From the triangles PIC and CIQ, we have

$$\alpha + \beta = \phi$$
$$\beta = \phi' + \gamma. \qquad (7\text{-}4)$$

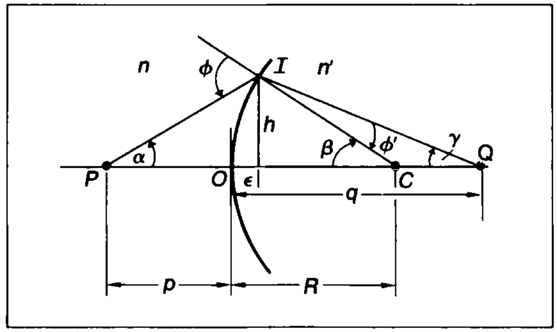


Figure 7-6 Refraction at a spherical surface.

Snell's law in this case is

$$n \sin\phi = n' \sin\phi'.$$

In addition, we can write

$$\tan\alpha = \frac{h}{p+\epsilon}, \quad \tan\beta = \frac{h}{R-\epsilon}, \quad \tan\gamma = \frac{h}{q-\epsilon}.$$

Again, restricting the discussion to paraxial rays permits use of the relations

$$\sin\theta \approx \tan\theta \approx \theta,$$

and the small distance ϵ can be neglected. Then Snell's law becomes

$$n\phi = n'\phi',$$

and the tangent relations become (approximately)

$$\alpha = \frac{h}{p}, \quad \beta = \frac{h}{R}, \quad \gamma = \frac{h}{q}.$$

Using these results in Eqs. (7-4),

$$\alpha + \beta = \frac{n'\phi'}{n} = \frac{n'}{n}(\beta - \gamma)$$

or

$$n\alpha + n'\gamma = (n' - n)\beta$$

and, finally,

$$\frac{n}{p} + \frac{n'}{q} = \frac{(n'-n)}{R} \qquad (7\text{-}5)$$

or

$$\frac{1}{p} + \frac{\frac{n'}{n}}{q} = \left(\frac{n'}{n} - 1\right)\frac{1}{R}. \qquad (7\text{-}6)$$

For a given object distance, therefore, we see that the image distance depends not only on the radius of curvature of the optical surface, as in reflection, but also upon the indices of refraction of the two media involved. We can also investigate the question of a focal point and a related focal length for a refracting surface. In this case, one can consider an infinitely distant object with an image at a focal point or an infinitely distant image with the object at a focal point. The magnitudes of the focal lengths in each case follow from Eq. (7-5), and it is likewise clear that they are on opposite sides of the surface.

To determine the magnification of an extended object (assuming small angles as before), we refer to Figure 7-7. Here we have

$$\tan \phi = \frac{y}{p}, \quad \tan \phi' = \frac{-y'}{q},$$

and

$$n \sin \phi = n' \sin \phi'.$$

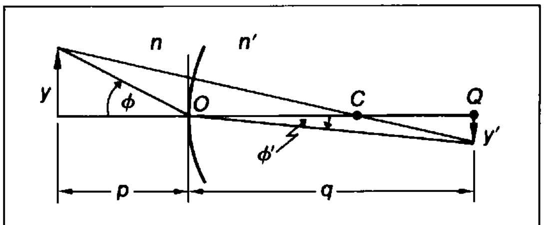


Figure 7-7 Refraction for an extended object by a spherical surface.

For small angles,

$$\phi = \frac{y}{p} = \frac{n'}{n}\phi', \quad \phi' = \frac{-y'}{q}$$

or

$$\frac{ny}{n'p} = \frac{-y'}{q}.$$

Therefore,

$$m = \frac{y'}{y} = \frac{-nq}{n'p}. \tag{7-7}$$

Once again, Eqs. (7-5) and (7-7) with our sign convention are sufficient for a complete analysis of single surface refraction (for paraxial rays). Notice that Eq. (7-6) has the same form as Eq. (7-2) for reflection. This is in part due to the choice of sign convention adopted.

Although we do not do so, the reader may demonstrate that if the two focal lengths are known from experiment—or by calculation from Eq. (7-6)—one can apply the ray tracing method of Section 7-2 without change to situations involving refraction.

7-5 COMBINATIONS OF REFLECTING AND REFRACTING SURFACES

In this section, we consider situations involving more than one optical surface, either reflecting or refracting. As the examples below illustrate, no new concepts are required. Instead, one simply solves the problem serially. That is, the image produced by the first surface becomes the object for the second surface, the new image is then used as the object for the third surface, and so forth. Therefore, adherence to the sign convention and application of Eqs. (7-2), (7-3), (7-6), and (7-7) to the various surfaces for any number of surfaces is sufficient. It must be stressed, however, that application must be made in the order that they affect the light from the original object.

Example 1. Let us describe the image formed for an object placed a distance of $5R$ to the left of the center of a glass sphere (index $n' = 1.50$) of radius R, as shown in Figure 7-8. The height of the object is $R/3$.

SOLUTION

Before we analyze the problem quantitatively, it can be inferred from Snell's law that the final image will be nearer the center of the sphere than is the object, and it will be reduced in size. This is true because the front (left) surface causes a convergence of light rays from the object ($n' > n$). Then, the already convergent rays are converged even more by the second (right) surface since the light rays are now entering a less dense medium and therefore bend away from the normal at the surface. This produces rays that cross the axis nearer the center of the sphere as asserted.

From Eq. (7-5), we have

$$\frac{1}{4R} + \frac{3}{2q_1} = \left(\frac{3}{2} - 1\right)\frac{1}{R}$$

for the first surface, or

$$q_1 = 6R.$$

Thus, with respect to the second surface, $p_2 = -(6R - 2R)$. Why? Equation (7-6) now yields

$$\frac{3}{2}\left(-\frac{1}{4R}\right) + \frac{1}{q_2} = \left(1 - \frac{3}{2}\right)\left(-\frac{1}{R}\right)$$

or

$$q_2 = \frac{8}{7}R.$$

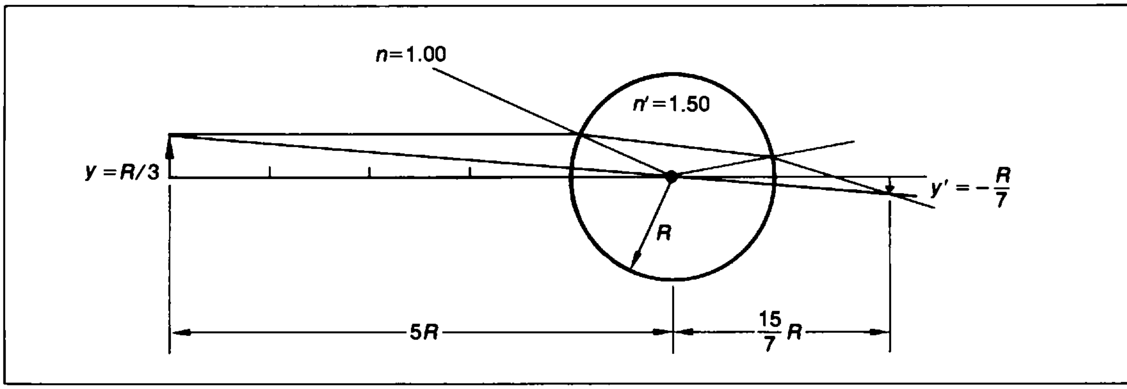


Figure 7-8 Refraction by a sphere.

Therefore, the final image is located at a distance $\frac{15}{7}R$ to the right of the center of the sphere. Since it is outside the sphere, it will be a real image. The total magnification will be the product of that produced by the first surface and that produced by the second surface. Therefore, using Eq. (7-7) twice gives

$$m_{\text{total}} = m_1 m_2 = \left(\frac{-q_1}{\frac{3}{2}p_1}\right)\left(\frac{-\frac{3}{2}q_2}{p_2}\right)$$
$$= \left(-\frac{6R}{\frac{3}{2}\times 4R}\right)\left(\frac{-\frac{3}{2}\times\frac{8}{7}R}{-4R}\right) = -\frac{3}{7}.$$

The real image is therefore inverted and reduced from a height $y = R/3$ to a height

$$y' = m_{\text{total}} = \frac{-3}{7}\times\frac{R}{3} = \frac{-R}{7}.$$

In the figure, a ray from the object through the center of the sphere passes through undeviated, whereas a ray parallel to the axis from the original object has been properly located by means of Snell's law at each surface to yield the result obtained above analytically.

Example 2. Repeat the analysis of Example 1 for an object of height $y = R/3$ located a distance $5R$ to the left of the center of a hemisphere (index $n' = 1.50$) of radius R if the plane surface of the hemisphere is silvered (totally reflecting).

SOLUTION

The situation is illustrated in Figure 7-9. As the rays sketched indicate, the image will be real, inverted, and near the front (left) surface. Using Eq. (7-5),

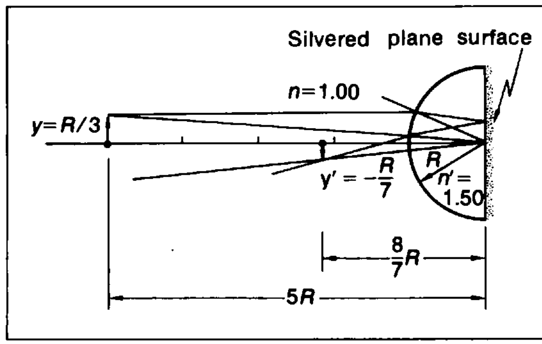


Figure 7-9 Refraction and reflection by a silvered hemisphere.

$$\frac{1}{4R} + \frac{3}{2q_1} = \left(\frac{3}{2} - 1\right)\frac{1}{R}$$

yields $q_1 = 6R$ so that $p_2 = -(6R - R) = -5R$. Equation (7-2) applies next for the reflection at the plane surface, giving

$$-\frac{1}{5R} + \frac{1}{q_2} = 0$$

or

$$q_2 = 5R.$$

This requires that $p_3 = -(5R - R) = -4R$ for the second refraction at the curved surface. Now Eq. (7-5) gives

$$\left(\frac{3}{2}\right)\left(-\frac{1}{4R}\right) + \frac{1}{q_3} = \left(1 - \frac{3}{2}\right)\left(-\frac{1}{R}\right),$$

and $q_3 = \frac{8}{7}R$ as shown. Here, the magnification is

$$m_{\text{total}} = m_1 m_2 m_3 = \overset{\text{refraction}}{\left(\frac{-nq_1}{n'p'}\right)} \times \overset{\text{reflection}}{\left(\frac{-n'q_2}{n'p_2}\right)} \times \overset{\text{refraction}}{\left(\frac{-n'q_3}{np_3}\right)}$$

$$= -\left(\frac{q_1 q_2 q_3}{p_1 p_2 p_3}\right) = -\frac{(6R)(5R)(\frac{8}{7}R)}{(4R)(-5R)(-4R)} = -\frac{6}{14}.$$

Thus, the final image in this case is real, inverted, and has a height

$$y' = m_{\text{total}}\, y = -\frac{6}{14}\frac{R}{3} = -\frac{R}{7}.$$

In the next chapter, we shall apply the considerations of this chapter to situations involving two refracting surfaces separated either by finite or negligible distances compared to the radii of curvature of the two surfaces. Such situations are known as thick and thin lenses, respectively, and their application, either singly or in combination, lead to such optical instruments as the microscope, the telescope, and the camera.

PROBLEMS†

1. It was asserted in Section 7-1 that there is no unique focal point for a spherical surface. Demonstrate this for reflection by making a scale diagram of a concave spherical mirror with a radius of curvature of 10 cm. With a straight edge, draw incident rays parallel to the axis at distances of 1, 2, 3, 4, and 5 cm from the axis. Using a protractor, construct the reflected rays and determine the points at which they cross the axis.
2. Show by ray tracing methods that the focal length f of a concave mirror with radius of curvature R is given by the relation $f = R/2$.
3. Construct by ray tracing methods the image produced by an object located in front of a convex mirror at a distance greater than the focal length. Also determine the magnification and whether the image is real or virtual. Is the image erect or inverted?
4. A student finds a spherical metal bowl which is shiny inside and out. Looking into the concave side, he sees his inverted image 4 cm from the bottom of the bowl. Turning the bowl over (and keeping the distance from himself to the bowl surface the same), he sees his erect image at 3 cm from the bowl. What is the radius of curvature of the bowl?
5. A concave mirror has a focal length of 1.5 cm. Where should an object be placed if its image is to be erect and twice the object size?
6. A spherical shaving mirror (concave; why?) has a radius of curvature of 30 cm. What is the magnification produced when the face is 10 cm from the mirror?
7. To determine the height of a tree, a resourceful physics student notes that the image of the tree just covers the length of a 5 cm plane mirror held 30 cm from his eye. If the tree is 90 m from the mirror, what is its height?
8. A spherical fish bowl has a radius R. Neglecting the effect of the glass wall, where would a goldfish actually at the center of the bowl appear to be when viewed from the outside? If the goldfish is 3 cm long, what will be its apparent length? For water $n = 1.33$.
9. A goldfish swimming inside a spherical fish bowl of radius R appears to be at a distance $2R$ from the front of the bowl. Neglecting the effect of the glass wall, where is the goldfish actually located relative to the front of the bowl? If the goldfish is 3 cm long what will be its apparent length? For water $n = 1.33$.
10. When refraction occurs at a spherical surface, Eq. (7-5) can be used to define two focal lengths f_1, f_2. When the object is at infinity, $q = f_1$, and when the image is at infinity, $p = f_2$. Show that

$$f_1 = \frac{n'R}{n'-n}, \quad f_2 = \frac{nR}{n'-n}, \quad \text{and} \quad f_1 - f_2 = R.$$

†Except for Problem 1, it is assumed that all light rays are paraxial.

11. An object is placed in front of a glass hemisphere of radius R, index of refraction $n = 1.50$ as shown in Figure 7-10. If the object is at a distance $3R$ from the spherical surface and has a height $y = R/3$, find the location, orientation, and size of the final image, and indicate whether it is real or virtual.

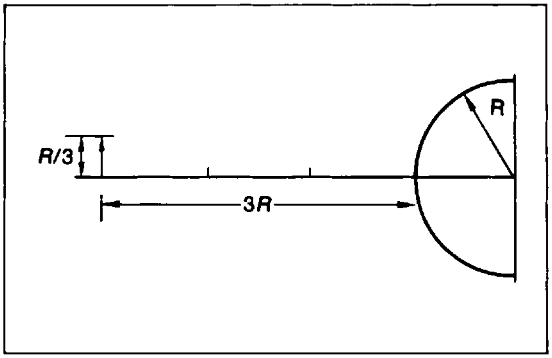


Figure 7-10

12. For the hemisphere of Problem 11, find the final image position when the object is an infinite distance from the spherical surface. Then repeat the calculation when the hemisphere is reversed (so that rays from the object now strike the plane surface first).

8 Lenses and Optical Instruments

8-1 THICK AND THIN LENSES

A lens may be described as a material of given index of refraction which is bounded by two spherical surfaces. If the separation of the two surfaces at the optical axis is not negligible compared to focal lengths and/or object and image distances, the lens is said to be a *thick* lens. Conversely, if the separation is negligible, it is said to be a *thin* lens. In either case, the method of Section 7-5 can be applied to situations involving one or more lenses, as was asserted there. (In fact Example 1 of Section 7-5 is a particular case of a thick lens whose radii of curvature are equal in magnitude.)

For thin lenses, it is possible to simplify the mathematical analysis of a given situation and also to develop a simple ray tracing technique analogous to the method of Section 7-2 for reflection. The ray tracing technique is also applicable to thick lenses provided the concept of principal planes is properly developed. Since such a discussion is more appropriate to an advanced study of optical systems, we will restrict ourselves to the case of thin lenses only.

Consider a double convex (converging) lens with radii of curvature $R_1 - R_2$ as in Figure 8-1, with an object located at a distance p_1 in front of it. From

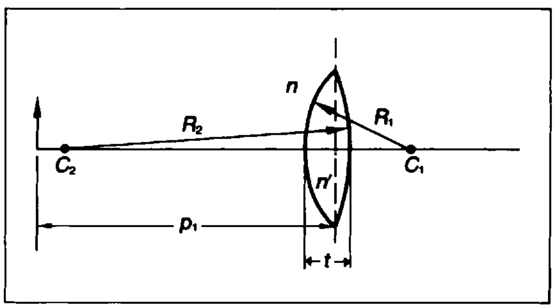


Figure 8-1 Geometry of an object and a thin lens of thickness *t*.

Section 7-5, we can write

$$\frac{n}{p_1} + \frac{n'}{q_1} = \frac{(n'-n)}{R_1} \tag{8-1}$$

for the first surface. Thus, the object distance p_2 for the second surface will be given by

$$\begin{aligned} p_2 &= -(q_1 - t) \\ &\approx -q_1 \quad \text{(since } t \text{ is assumed negligible for thin lenses).} \end{aligned}$$

Therefore,

$$\frac{n'}{p_2} + \frac{n}{q_2} = (n - n')\left(-\frac{1}{R_2}\right) = -\frac{n'}{q_1} + \frac{n}{q_2}. \tag{8-2}$$

Combining Eqs. (8-1) and (8-2),

$$\frac{n}{p_1}+\frac{n}{q_2}=(n'-n)\left(\frac{1}{R_1}+\frac{1}{R_2}\right). \qquad (8\text{-}3)$$

We have previously defined focal lengths for both reflection and refraction. If that concept is applied here we see that when $q=\infty$,

$$p=f_1=\left[(n'-n)\left(\frac{1}{R_1}+\frac{1}{R_2}\right)\right]^{-1},$$

and when $p=\infty$,

$$q=f_2=\left[(n'-n)\left(\frac{1}{R_1}+\frac{1}{R_2}\right)\right]^{-1}. \qquad (8\text{-}4)$$

Thus, the magnitudes of the two focal lengths of a thin lens are the same, as could be checked experimentally by reversing the lens and noting that this has no effect on the image produced. Equation (8-4) is known as the lens maker's equation. The net effect of the lens on the light rays from an object can be written simply as

$$\frac{n}{p}+\frac{n}{q}=\frac{1}{f}, \qquad (8\text{-}5)$$

where the subscripts 1 and 2 are no longer necessary if we understand p to be the original object distance, q to be the final image distance, and note that the two focal lengths are identical in magnitude. If the lens is in air ($n=1$)† then Eq. (8-5) becomes

$$\frac{1}{p}+\frac{1}{q}=\frac{1}{f},$$

which is identical in form to Eq. (7-2) for single surface reflection, and is known as the Gaussian form of the thin lens equation. Similarly, the magnification for a thin lens is given by Eq. (7-3). (See Problem 8-5.) It should be noted that by our use of Eq. (7-5) we are restricted here to situations involving paraxial rays.

For thin lenses, one can regard the refraction of individual light rays as taking place at a plane perpendicular to the axis and through the center of the lens rather than considering separate effects at the two surfaces. The resulting refraction is essentially an average of the effects at either surface, and for thin lenses introduces very little error. With this simplification, it is possible to analyze any thin lens graphically by means of three rays similar to the case for reflection. The three rays chosen in this case are illustrated in Figure 8-2:

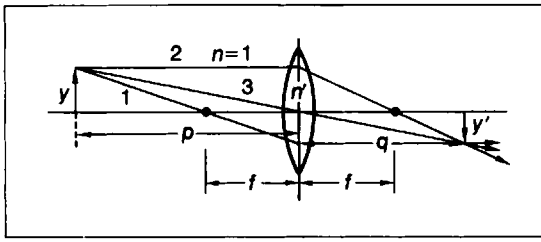


Figure 8-2 Refraction by a converging lens—image location by ray tracing.

1. A ray from the object passing through (or toward) the first focal point will travel parallel to the axis upon refraction.
2. A ray parallel to the axis will be refracted so as to pass through (or appear to come from) the second focal point.
3. A ray through the center of the lens will be essentially undeviated.

The third ray is practically undeviated because the thin lens for a central ray approximates a parallel plate for which there is displacement without deviation (Section 6-5). Since the displacement is proportional to the thickness (Problem 6-5), it is clear that the displacement is negligible for a thin lens. For example, in Figure 8-2 an object is placed a distance p in front of a double convex lens of focal length f. The image is located as shown by constructing the rays indicated in the previous paragraph. Figure 8-3 illustrates a similar graphical analysis for a double

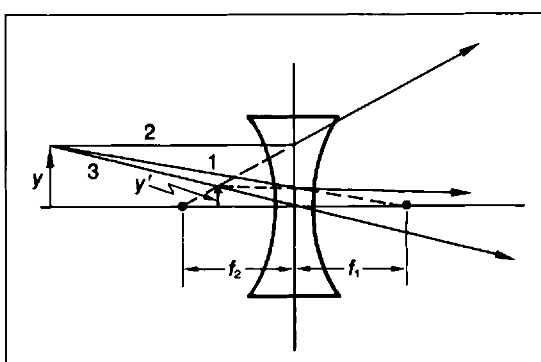


Figure 8-3 Refraction by a diverging lens—image location by ray tracing.

†Since air is essentially a vacuum; see Table 6-1.

concave (diverging) lens. It should be noted here that the first and second focal lengths are reversed compared to a converging lens. This is because of the signs of the radii of curvature in Eq. (8-4).

8-2 LENS COMBINATIONS

The effects of a combination of thin lenses are analyzed in the same manner as was refraction for a combination of surfaces. That is, the image formed by the first of a series of lenses is determined by Eqs. (8-4) and (8-5) or by a ray tracing technique. This image then becomes the object for the second lens (with proper account being taken of distances and signs relative to the second lens). This process is repeated for each successive lens in the combination.

Example 1. Show that if two thin lenses of focal lengths f_1 and f_2, respectively, are placed in contact, the effective focal length f of the combination is given by the expression

$$\frac{1}{f} \approx \frac{1}{f_1} + \frac{1}{f_2}.$$

SOLUTION

Referring to Figure 8-4, Eq. (8-5) yields

$$\frac{1}{p} + \frac{1}{q} = \frac{1}{f_1}$$

or

$$\frac{1}{q} = \frac{1}{f_1} - \frac{1}{p}.$$

But $p' \approx -q$ (since the lenses are thin). Hence,

$$\frac{1}{p'} + \frac{1}{q'} = \frac{1}{f_2}$$

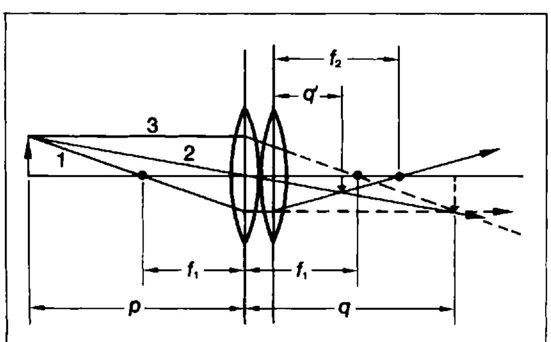


Figure 8-4 Refraction by two lenses in contact.

becomes

$$\frac{1}{p} + \frac{1}{q} \approx \frac{1}{f_1} + \frac{1}{f_2}.$$

But this is equivalent to a single lens of focal length given by

$$\frac{1}{f} \approx \frac{1}{f_1} + \frac{1}{f_2}. \qquad \text{Q.E.D.}$$

In Figure 8-4, the rays for lens 1 are dashed as is the image for lens 1 alone. The ray parallel to the axis (ray 1) is refracted by lens 2 so as to pass through the second focal length, while the ray through the center of lens 1 (ray 2) is essentially undeviated by the second lens, thus locating the image.

Example 2. A lens of focal length f_1 is placed at a point which is a distance d to the left of a second lens of focal length f_2. An object of height y is located at a distance $p_1 = 2f_1$ to the left of the first lens. Describe the final image. Assume $f_1 = f_2 = f > d$.

SOLUTION

Applying Eq. (8-5) for the first lens,

$$\frac{1}{2f} + \frac{1}{q_1} = \frac{1}{f},$$

so that

$$q_1 = 2f.$$

Now, applying Eq. (8-5) for the second lens,

$$p_2 = -(q_1 - d) = d - 2f$$

and

$$\frac{1}{d - 2f} + \frac{1}{q_2} = \frac{1}{f}$$

or

$$q_2 = \frac{fd - 2f^2}{d - 3f} = \frac{f(2f - d)}{(3f - d)} > 0.$$

Thus, the final image is real and inverted, features shown in Figure 8-5. The magnification is $m_{total} = m_1 m_2 = (-q_1/p_1) \times (-q_2/p_2)$.

$$m_{total} = \left(\frac{-2f}{2f}\right)\left(\frac{-f(2f - d)}{(3f - d)}\right)\left(\frac{1}{-(2f - d)}\right) = -\frac{f}{3f - d}.$$

Therefore,

$$y_{final} = m_{total}\, y = -\frac{yf}{3f - d}.$$

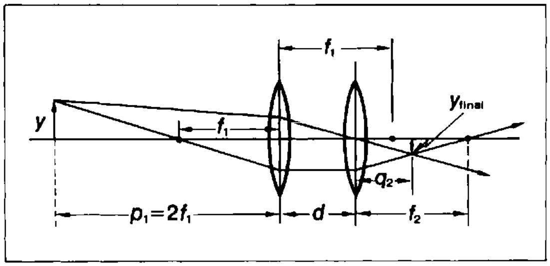


Figure 8-5 Refraction by two lenses not in contact.

8-3 THE EYE

As an optical instrument, the eye possesses the features shown in Figure 8-6. Although the physiology of the eye is extremely complicated, the functions of the various parts of the eye can be broadly outlined. Light enters the eye through the transparent, curved shell called the cornea, *C*, and passes through the liquid substance (aqueous humor—*A*) immediately behind the cornea. It then enters the pupil, a circular aperture in the colored portion of the eye called the iris. The expansion or contraction of the pupil is controlled by the eye muscles in response to the brightness of the entering light. The beam of light then enters the crystalline lens, *L*, which is actually rather plastic and whose surface curvature can be changed by muscle action of the eye in order to achieve an appropriate focal length for viewing objects at various distances. The interior portion of the eyeball contains a jelly-like material, the vitreous humor, *V*, while the inner lining of the back of the eyeball is a light sensitive lining called the retina, *R*. It is made up of millions of light receptors (called rods and cones because of their shapes) which are arranged normal to the retinal surface and which transmit signals through the optic nerve, *O*, when light is incident upon them. Thus, for proper functioning, light entering the eye should come to a focus on the retinal surface, which is spherical in shape. Although the major part of the focusing action occurs at the cornea, all of the transparent portions of the eye are appropriately regarded as part of the lens system of the eye, as Table 8-1 of approximate indices of refraction for these portions indicates.

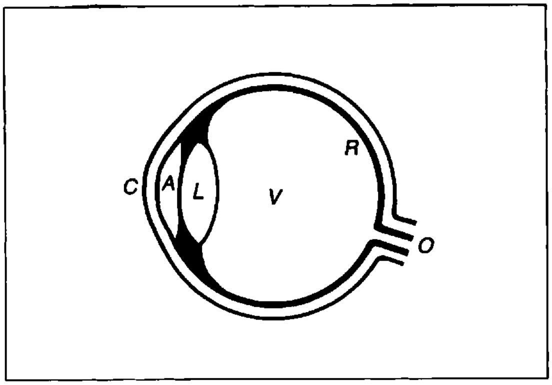


Figure 8-6 The human eye.

Table 8-1 Indices of Refraction for Portions of the Eye.

Cornea	1.351
Aqueous humor	1.336
Crystalline lens	1.437
Vitreous humor	1.336

Because of the adaptability of the eye, objects can be viewed distinctly over a range of distances, the extremes of which are known as the far point and the near point. For a relaxed normal eye, the far point is at infinity. *Myopia*, or *nearsightedness*, is the name given to the case for which the far point is at some finite distance, a situation that can be corrected by means of a suitable diverging lens.

The near point of the eye is determined by the flexibility of the crystalline lens and the consequent ease with which the eye muscles can increase the lens curvature for close distance viewing. This process of *accommodation*, as it is called, is a function of age in that the crystalline lens loses flexibility with age. As a result, the near point is smallest for children and increases with age, going from about 7 cm at the age of ten years to over 40 cm after 50 years. This condition is called *presbyopia*, and is a normal change in the eye rather than a defect of vision. The near point for the normal adult eye for many years can be taken to be 25 cm. An eye for which the near point is beyond 25 cm during the adult years is said to be *hypermetropic*, or far sighted, and correction is obtained by use of a suitable converging lens. A third common type of defect is *astigmatism*, a condition indicated by distorted images. It occurs when one or more of the elements of the eye (such as the cornea or crystalline lens) are not perfectly spherical. Correction is obtained by means of a suitably placed cylindrical lens to correct the asymmetry of the eye elements.

Example 3. A very nearsighted person cannot see clearly objects that are more than 30 cm away. What is the focal length of the lens required for this person to be able to see distant objects clearly?

SOLUTION

The required lens must form a virtual image at 30 cm in front of the eye for an object at infinity. Thus, $p=\infty$, $q=-30$ cm, and Eq. (8-5) yields

$$\frac{1}{\infty}-\frac{1}{30}=\frac{1}{f}$$

or $f=-30$ cm.

8-4 THE SIMPLE MICROSCOPE OR MAGNIFIER

When one views an object with the unaided eye, the detail that can be seen is dependent upon the size of the retinal image, which in turn is dependent upon the angle subtended by the object at the eye. Thus, the nearer one places the object to the eye, the larger will be the angle subtended. However, the eye cannot focus sharply on an object closer than the near point, so that the maximum angle subtended by a clearly focused object will occur when the object is at the near point, which we assume as above to be 25 cm. To increase the angle subtended, one places a converging lens in front of the eye to act as a magnifier or simple microscope. The lens thus used forms a virtual image of an object placed closer to the eye than the near point, and the eye then sees this enlarged virtual image which can be seen clearly anywhere between the near point and the far point. Most relaxed viewing would occur if the image were to appear at infinity (the normal far point); we assume this to be the case.

The magnifying power M of an optical instrument is defined to be the ratio of the angle subtended by the object when placed at the focal point of the lens (enlarged virtual image at infinity) θ' to the angle subtended by the object at the near point (unaided eye) θ. Thus,

$$M=\frac{\theta'}{\theta}\approx\frac{y/f}{y/25}=\frac{25}{f}, \tag{8-6}$$

where y, the object height, and f are both in centimeters. (Why?) The influence of aberrations for a simple double convex lens limits M to about 2 or 3, although a corrected lens can have M values as high as 20.

8-5 THE COMPOUND MICROSCOPE

When one wishes to examine very small objects with great magnifying power, it is necessary to use a combination of two lenses each of which contributes to the total magnifying power. Such a combination is called a compound microscope. It is shown schematically in Figure 8-7. As the figure indicates, the object to be examined is placed at a point slightly greater than the focal length of the objective lens, which forms an enlarged or laterally magnified real image at a point just within the first focal point of the eyepiece or ocular lens.† The eyepiece lens then forms an angularly magnified image at a distance that can be anywhere between the near and far points of the eye.

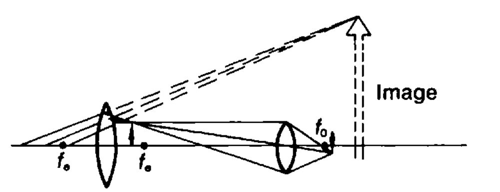


Figure 8-7 The compound microscope.

The total magnification in this case is the product of the enlargement lateral magnification m_0 produced by the objective and the magnifying power M_e of the eyepiece. Thus,

$$M_{\text{total}}=m_0M_e\approx\frac{-q}{f_0}\times\frac{25}{f_e}, \quad \text{all distances in cm.} \tag{8-7}$$

8-6 TELESCOPES

The telescope, like the compound microscope, consists of an objective and an ocular. However, since the telescope is commonly used for examining very distant objects, the image from the objective will be formed at the second focal point of the objective. If the final (virtual) image is to be at infinity, the first image must be at the first focal point of the ocular. As a result, the length of the telescope or,

†For brevity we shall use objective and eyepiece—or ocular—when discussing the objective lens and the eyepiece—or ocular—lens.

more accurately, the distance between objective and ocular, is the sum of the focal lengths, $f_0 + f_e$.

The magnifying power for the telescope is defined as the ratio of the angle subtended by the final image (θ_E) to the angle subtended by the object at the unaided eye. Figure 8-8 illustrates the situation. Thus, the essentially parallel rays from the distant object subtend the same angle θ_0 with the objective as they would for the naked eye. The ray from the object through the first focal point of the objective proceeds parallel to the axis of the telescope, through the ocular and its second focal point to the eye, which is slightly to the right of this second focal point of the ocular. Therefore, it is clear that

$$\tan\theta_0 = \frac{-h}{f_0}, \quad \tan\theta_E = \frac{h}{f_e}.$$

Because of the great distances involved, the angles and their tangents are essentially equal, so that

$$M_{\text{total}} = \frac{\theta_e}{\theta_0} = \frac{-f_0}{f_E}. \tag{8-8}$$

The minus sign in Eq. (8-8) indicates an inverted image, which is not a serious disadvantage for nonterrestrial observations. For terrestrial telescopes, an erect image is to be preferred. This can be accomplished by the insertion of a third (erecting) lens between the objective and the ocular. The reader can show that if the erecting lens is located at a distance of twice its focal length from the image of the objective, and the ocular is located so that the image due to the erecting lens is at its first focal point, the total magnifying power is unchanged. What would then be the length of the telescope?

If the increase in length is undesirable, the erection can be accomplished by means of a pair of 45°-45°-90° prisms, which totally reflect the light four times (from each pair of the inclined faces of the prisms) between the objective and the ocular. It is left as an exercise to show that the image is erect after the reflections described. This method of image erection is utilized in the prism binocular.

8-7 THE CAMERA

The photographic camera is a device for producing a real image of an object on a light sensitive film so that a permanent record of the object photographed may be obtained by appropriate chemical treatment of the exposed film. The camera is similar to the eye in many respects. Thus, the lens of the camera must properly focus light from the object on the film in the rear of the camera, just as the lens system of the eye must focus light from an object on the retina. Also, most cameras are equipped with a diaphragm allowing light to pass through an aperture which is adjustable depending upon light brightness, just as the pupil in the iris does for the eye. If the camera is to be used for both distant and close objects, a movable lens system must be provided to achieve the accommodation that muscular control of the crystalline lens provides in the eye. Automatic cameras use a photocell detector to control the aperture size continuously.

If the photographic record is to be satisfactory,

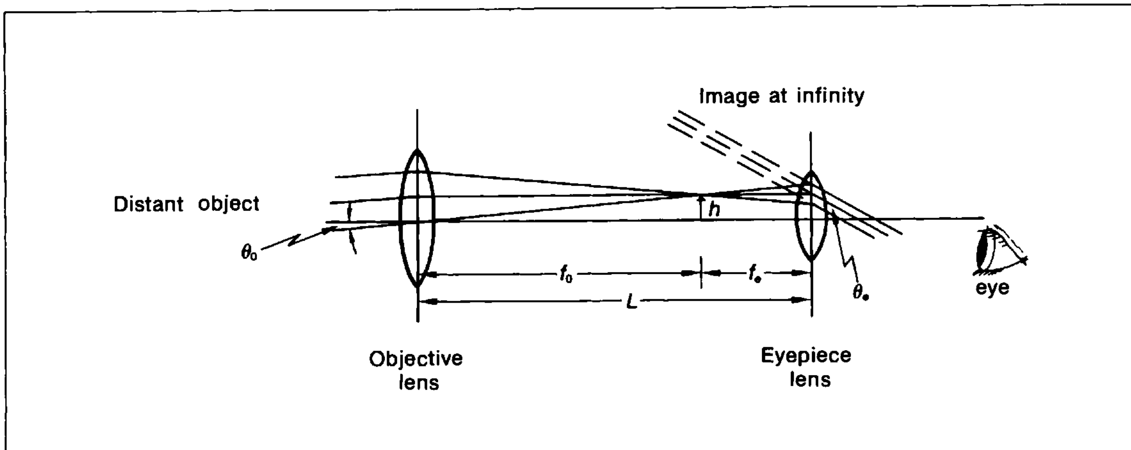


Figure 8-8 The telescope.

the exposure of the film must fall within rather well-defined limits. Greater exposure will result in a record that is too dark, while underexposure will produce a faded, washed-out result. The light-gathering ability of a camera is determined by the open area of the lens and the focal length of the lens. That is, the light that is brought to a focus on the film is equal to the light from the object times the solid angle subtended by the open area of the lens at a point on the film. From Figure 8-9, we see that the solid angle Ω is approximately equal to $\pi d^2/4f^2$. Thus, the light gathered by the lens system is proportional to $(d/f)^2$. The f-value of the lens is written as $f/(N)$, where N is the ratio of the lens focal length f to the lens opening diameter d. It is a measure of the speed of the lens system. The smaller the f-value, the greater is the amount of light gathered by the lens system. As a result, the exposure time can be reduced compared to a "slower" (higher f-value) lens system. The exposure time is thus proportional to the square of the lens opening diameter for a lens of given focal length. For convenience, many camera lens systems have diaphragm settings for which exposure times differ by factors of two. Thus, the exposure time for $f/6.3$ compared to $f/4.5$ is given by

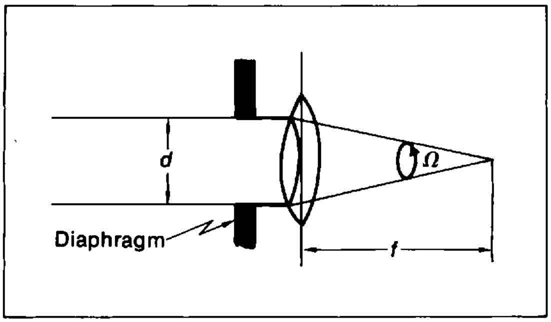


Figure 8-9 Light gathering of a camera lens system.

$$\frac{(6.3)^2}{(4.5)^2} \approx \frac{39.7}{20.2} \approx 1.96 \approx 2.$$

Camera lenses commonly available have f-values ranging from $f/1.5$ to $f/16$. For the low f-value lenses the paraxial rays approximation is no longer valid, and the effects of chromatic aberration and various other lens aberrations must be removed by appropriate combinations of lens materials and lens configurations.

PROBLEMS

1. An object is placed a distance $3R$ from the surface of a glass sphere, radius R and index 1.50. Determine the position of the final image. Then repeat the calculation using the (invalid here!) thin lens equation. This shows to some extent the errors that improper use of relationships can incur.
2. An object is located (a) 10 cm, (b) 20 cm from a double convex lens of focal length 15 cm. Find by ray tracing and by calculation the position, orientation, and magnification of the image formed in the two cases. Tell whether the image is real in each case.
3. A concave lens forms a virtual image which is 10 cm from the object. The magnification is $\frac{2}{3}$. What is the focal length of the lens?
4. (a) Two thin watch glasses, radius of curvature R, are used to form a convex lens that is filled with water ($n = 1.33$). If an object of height $R/3$ is located a distance R to the left of the lens, describe the image formed. Neglect the effect of the glass.
   (b) Suppose now that the water in the lens is emptied and the air lens is immersed in a tank of water. Describe the image formed for the same object and location as in (a).
5. Show that Eq. (7-3) correctly gives the magnification for a thin lens.
6. A lens produces an image on a screen 25 cm from the object. The image is 4 times larger than the object. Find the focal length of the lens.
7. Calculate the possible focal lengths for a lens of index 1.50 that is to be made using two surfaces with radii of curvature of 10 and 15 cm.

8. A convex and a concave lens, with focal lengths which both have of a magnitude of 10 cm, are arranged coaxially and separated by 3 cm. Find the distance between an object and its image when the object is on the axis and 15 cm from:
   (a) the convex lens,
   (b) the concave lens.
   Note: There are *four* possible answers.
9. An object is located 200 cm from a converging lens of focal length 40 cm. A diverging lens of focal length − 60 cm is located 20 cm behind the converging lens. Find the position of the final image both by calculation and by ray tracing. Is the image real or virtual, erect or inverted?
10. (a) Where is the near point of eyes for which eye glasses of focal length + 50 cm are required? These glasses shift the near point to 25 cm from the eyes.
    (b) What focal length eye glasses are required for eyes with a near point of 25 cm if objects at 25 cm are to be seen clearly?
11. (a) The far point for a myopic eye is at 75 cm. What is the focal length of a lens that will allow distant objects to be seen clearly?
    (b) Where is the far point for an eye for which a diverging lens of focal length − 100 cm is required for viewing distant objects?
12. A jeweler, using an eyepiece to examine a watch, holds the watch about 5 cm from the lens. What is the magnifying power of the eyepiece?
13. A thin lens of focal length 10 cm is used as a simple magnifier. What magnifying power is obtainable with the lens? What is the closest distance that an object may be brought to the eye when using the lens?
14. A compound microscope has objective and eyepiece lenses of focal lengths 1 and 3 cm, respectively. An object placed 1.2 cm from the objective leads to a virtual image produced by the eyepiece that is 25 cm from the eye. What is the total magnifying power of the microscope, and what is the separation distance between the lenses?
15. The image from a terrestrial telescope is to be made erect by insertion of an erecting lens as discussed in Section 8-6. Show that the total magnifying power is unchanged if the erecting lens is located at a distance of twice its focal length from the image from the objective and the ocular is located so that the image from the erecting lens is at its first focal point. What is the length of the telescope now?
16. If a fixed amount of light requires 1/50 second exposure for a camera setting of $f/4$, what will be the exposure time for the same light using a setting of $f/2.8$?
17. A camera lens forms the image of a distant point source of light on a screen at *A*. (See Figure 8-10.) When the screen is moved backwards a distance of 3 cm to *B*, the image on the screen becomes a circle of light with a diameter of 7.5 cm. What is the f-value of the lens?

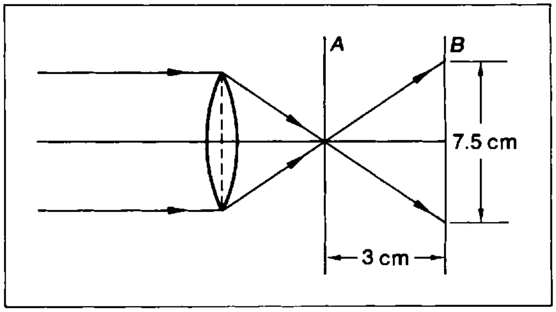


Figure 8-10

9 The Nature of Light

9-1 INTRODUCTION

In our discussion of optical phenomena, we have thus far assumed only rectilinear propagation, since this is in apparent agreement with the experimental facts presented up to this point. Using this assumption, relationships have been developed making quantitative predictions possible for reflection and refraction phenomena. This success does not, however, provide insight as regards the nature of light. To progress further in our understanding, we must make additional assumptions. Such assumptions are known formally as hypotheses or postulates. A set of such hypotheses together with any logically derived consequences constitute a physical theory. Such a theory is then judged on the basis of its success in correlating a variety of known phenomena and in predicting the outcome of additional experiments not yet performed.

It should be stressed that a physical theory is not a static system. Thus, if it has been successful in explaining a number of situations but fails to explain still another, the usual procedure is to seek modifications of the theory to include the additional situation rather than to discard the partially successful theory in favor of an entirely new one. Thus, for example, the laws developed by Newton that represent the theory of classical mechanics (see Chapter 14 et seq.) are entirely adequate for discussing the motion of everyday objects such as automobiles, hockey pucks, or even satellites. They are not, however, suitable for motion involving minute objects such as mesons or accelerated electrons that are moving with speeds approaching that of light (which is extremely high as we shall see in Section 9-4). For the latter category of phenomena, the theory of relativity (see Chapter 15) as developed principally by Einstein is required. This situation does not invalidate the classical or Newtonian theory for slow moving objects. In fact, for speeds that are small compared to that of light, the relativistic theory contains the classical theory as a limiting or special case. In this sense, relativity theory is a generalization of the classical theory rather than a total replacement.

Another situation similar to this is encountered when we wish to discuss systems on an atomic or subatomic scale, where the "graininess" of the material world becomes more apparent. It is perhaps not surprising that in this case the classical theory must be replaced by a quantum theory to properly account for the discrete nature of matter. Quantum mechanics provides a satisfactory theory of the material world, but the theory embodies many concepts and consequences that are at variance with our observations of the everyday (or macroscopic) world. Again, however, this quantum mechanical formalism leads correctly to the classical (large scale) results when it is applied to classical situations.

It should be emphasized that the role of a physical theory is to provide a basis for understanding the results of experiments rather than a search for *the* true or correct theory of nature. The successful prediction of the outcome of a given experiment is a measure, not of the correctness of a theory, but of its adequacy. To state that a theory is true or correct is to imply that it will be successful in predicting outcomes for any and all experiments, and (except in a negative sense) this is not subject to complete verification.

It is in fact frequently the case in physics that alternative theories have been proposed to correlate or explain the same group of phenomena. In such situations, further research and discussion are directed toward: (1) finding which of the alternative theories is most successful for the widest range of phenomena, and (2) finding the modifications of the most successful theory (or theories) that extend the applicability without unnecessarily increasing the number of basic postulates.

The development of a physical theory is thus a process primarily of evolution, involving a continuous interplay between theory and experiment. (We remark here that an introduction to the study of physics is incomplete if it does not include some opportunity to become acquainted with some of the equipment and techniques that the experimental physicist employs in testing and extending the predictions of the theoretical physicist.)

Let us now return to the question of the nature of light. The phenomena presented thus far—rectilinear propagation, reflection, refraction, dispersion, and the related existence of the color spectrum—had all been discovered or studied by 1670. Other phenomena (to be discussed in Chapters 11 and 12), including diffraction, double refraction, and the colors of thin films, were known by the beginning of the eighteenth century. At roughly the same point in time, two theories existed which were claimed (by their proponents) to be capable of explaining the behavior of light. In the following sections, we shall present an outline of these theories and compare them as to their adequacy. In the final section, we will present briefly some methods used for the determination of the speed of light and the values obtained as a result.

9-2 A PARTICLE THEORY FOR LIGHT

It is Sir Isaac Newton that we associate with the corpuscular theory of light. This theory proposes that light consists of a stream of particles or corpuscles that move at constant speed in a uniform medium, suffer elastic collisions at a reflecting surface, and are attracted by an optically more dense substance as they approach it. Because light can be reflected and refracted simultaneously, he also proposed the existence throughout space of "an ethereal medium, much of the same constitution as air but far rarer, subtiler, and more strongly elastic." This "luminiferous ether" at a boundary between media is "less pliant and yielding than in other places," so that the light particles impinging upon the boundary set this ether into vibration in a manner analogous to the action of a stone cast into a pool of water. The rapid vibrations thus provide intervals of easy reflection and intervals of easy transmission, which occur at such a high frequency as to create the impression of continuous reflection and refraction.

This theory provides an optical analogue to the behavior of material particles as described by the laws of classical mechanics. Thus, it explains rectilinear propagation in a uniform medium as well as reflection and refraction. Snell's law can be derived by means of this theory; the relative index of refraction can be shown to be equal to the ratio of the speed of the light particles in the second medium to that in the first medium. In this way, Newton concluded that the speed of light was greater in an optically more dense medium. This conclusion was not satisfactorily examined experimentally for over one hundred and fifty years, at which time it was shown to be incorrect. However, it was the earlier failure of this theory to predict either the bending of a beam of light about an obstacle (diffraction) or the interference effects of two or more beams of light that made it obsolete. It was discarded in favor of a wave theory when it became clear that no satisfactory modification of the particle theory could be found that incorporated these phenomena discovered after its initial formulation. Thus, it was a wave theory of light which survived an unfortunately persistent, often bitter, and intensely personal controversy extending through the eighteenth century, even into the nineteenth century.

9-3 A WAVE THEORY FOR LIGHT

The wave theory was developed initially by Christian Huygens, a Dutch contemporary of Newton. Like the particle theory, it assumes the existence of an ether in which the light waves propagate similar to water waves propagating on the surface of a pond. In this theory, the light waves propagate according to a rule known as Huygens' principle. (See Chapter 10.) Since reflection and refraction occur simultaneously for any wave motion, the laws of reflection and refraction are explained satisfactorily. In addition, one can conclude that light travels more slowly in an optically more dense medium, which proves to be the case as implied above in Section 9-2. More important is the fact that this theory is completely successful in also predicting interference and diffraction effects. As we will see in Chapter 10, however, Huygens' principle would not seem to predict rectilinear propagation of light. This was in fact the principal reason Newton advocated the particle theory, since nonrectilinear propagation of light is not readily observed in experiments involving reflection and refraction.

In subsequent chapters, we shall discuss the properties of waves. As a result, it will be clear that the wave theory of light merits adoption as an adequate theory regarding the nature of light.

9-4 THE SPEED OF LIGHT AND ITS DETERMINATION

When one attempts to determine the speed of an object, two quantities must in some way be measured; the distance traveled by the object and the time elapsed while the movement takes place. Racing cars or aircraft seeking to establish new speed records must traverse a carefully measured course of specified length, and are timed with highly sophisticated precision timing instruments.

In principle, the determination of the speed of light would be expected to be equally straightforward. Thus, Galileo and an assistant stood on separate hills in Rome, a known distance apart, equipped with shuttered lanterns. The experiment was quite simple: Galileo opened the shutter of his lantern allowing light to proceed to the assistant who opened his shutter as soon as he saw the signal from the first lantern. Galileo determined the length of time between opening his shutter and detecting the assistant's signal. This should have provided the necessary data were it not for the extremely high speed of light. Thus, the crudeness of the timing apparatus, together with reaction time limitations, limited the results to a conclusion that the speed of light was either instantaneous or at least immeasurably great. It is this great speed that makes it necessary to resort to light paths of astronomical magnitude if one wishes to measure time intervals directly. Any other technique necessarily will be less direct if accuracy is desired.

The earliest astronomical method of measurement was explained in 1676 by Olaf Römer, a Danish astronomer, who was then at Paris making measurements of the times of revolution of the satellites of the planet Jupiter. The orbits of these satellites lie in essentially the same plane as the orbits of Jupiter and the Earth about the Sun. Consequently, revolution times can be determined by measuring the time interval between successive emergences of the satellites from behind the planet's disk.

From a lengthy series of observations of the innermost satellite, Römer found that the time of revolution was more or less than the average result when the Earth was receding from or approaching Jupiter. He correctly concluded that the differences were due to the variable distance from the Earth to Jupiter, and the finite time required by the light reflected from the satellite to traverse this distance. The time required for Jupiter to make one revolution about the Sun is about 12 times greater than the time (1 year) required by the Earth. Thus, while the Earth moves through an appreciable portion of its orbit, Jupiter is essentially at rest, as illustrated in Figure 9-1, which shows the two planets at successively numbered corresponding positions. Any change in the time duration of eclipses will be due to a change in the Earth-Jupiter separation distance. In particular, the differences in eclipse periods occurring when the Earth is at positions 1 and 3 will be due to a change in separation distance that is satisfactorily approximated by the diameter of the Earth's orbit about the Sun. The speed of light is then found by dividing the orbital diameter of the Earth by the difference in eclipse times mentioned above. In Römer's time, the Earth's orbital diameter was not accurately known, and the time difference measured by him was about 22 minutes (compared to the present value of nearly 17 minutes). Thus, the speed derivable from his work was

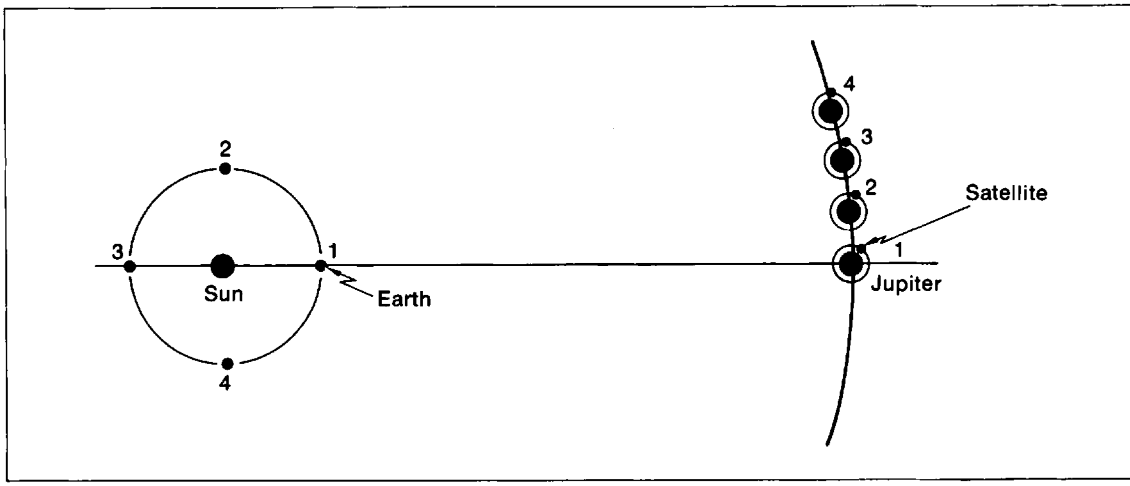


Figure 9-1 Römer's method for the speed of light (schematic).

of interest mainly because it established that it was indeed finite, though extremely large (about 2×10^8 m/sec). Another astronomical determination made by James Bradley some 50 years later, involving an independent phenomenon (stellar aberration), resulted in a value of 3×10^8 m/sec, thus confirming the order of magnitude indicated by Römer.

Successful terrestrial measurements of the speed of light were first performed in the mid-nineteenth century by the French physicists Hippolyte Fizeau and Jean Foucault, contemporaries who collaborated for a time using a technique devised by Fizeau and improved and modified in later independent work by Foucault. In the Fizeau experiment, a properly focused beam of light directed toward a distant mirror passes through a toothed wheel. When the wheel is not rotating, the light reflected from the distant mirror passes through the same notch in the wheel and continues to the observer (see Figure 9-2). Now if the wheel is set in motion at a rotational speed that is slowly increased, the light will first be observed when the reflected light arrives back at the toothed wheel, just as the second

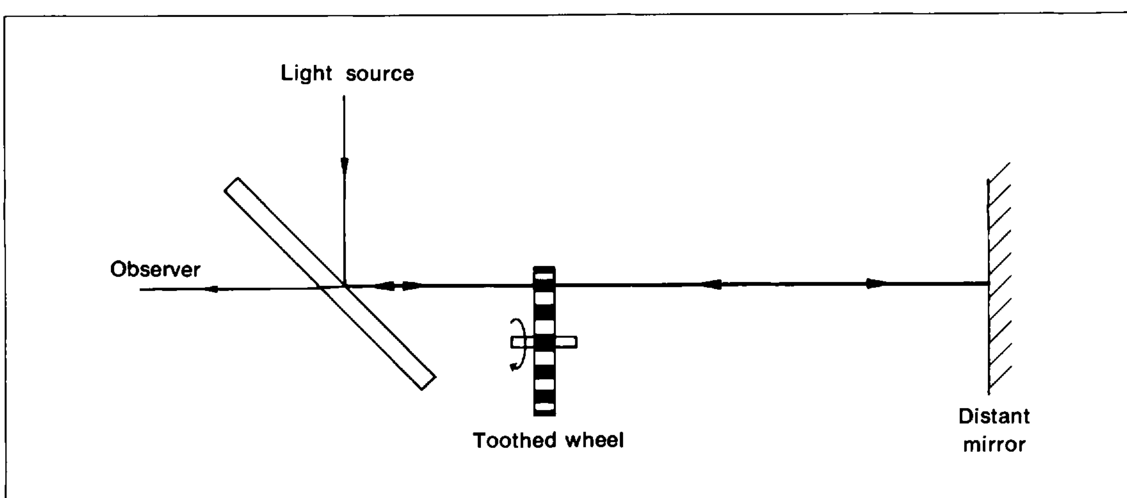


Figure 9-2 Fizeau's toothed wheel experiment (schematic).

notch has arrived at the position of the initial notch for the original passage of the light. At rotational speeds such that the third, fourth, etc. notch can rotate into position while the beam of light travels to the distant mirror and back, the light will again be visible to the observer.

Thus, if one can accurately measure the rate of revolution of a wheel with equally spaced notches, a series of rotation speeds that are integer multiples of a lowest speed is obtained. The speed of light is then found from the relation

$$\text{speed of light} = \left(\frac{2 \times \text{wheel to mirror distance}}{\dfrac{\text{number of notches which shifts to initial position}}{\text{total number of notches on wheel}}}\right)\left(\frac{\text{revolutions of wheel}}{\text{seconds}}\right). \quad (9\text{-}1)$$

In Fizeau's experiment, the wheel had 720 teeth or notches, the distant mirror was 8.63 km away, and the slowest rotation rate for letting light pass was found to be 25.2 rev/sec, leading to a result of 3.13×10^8 m/sec. Foucault later replaced the rotating wheel by a rotating plane mirror, a modification which made possible the use of laboratory distances, and obtained a result of 3.01×10^8 m/sec in air, (See Problem 9-3.) He also determined by this method that the speed of light in water was less than this value, as predicted by the wave theory. The American physicist Albert Michelson refined the experiments of Fizeau and Foucault in a series of investigations beginning in 1931. The result of his efforts was a value for the speed of light which is known to be accurate to within 1 part in 10^5.

Other techniques that are based upon the fact that light waves are electromagnetic in nature (as are radar and radio waves—see Chapter 36) have been developed and refined over the past 50 years.† As a result of these efforts it was recently reported‡ that the speed of light propagating in a vacuum (this speed is designated by the letter c) has the value

$$c = 299{,}792.5 \pm 0.3 \text{ km/sec}.$$

In subsequent discussions, we will use the approximation (satisfactory for slide rule calculations) that

$$c = 3 \times 10^8 \text{ m/sec},$$

a figure in error by less than 1%.

†See "The Speed of Light," *Scientific American*, August 1955, p. 62.

‡Report to the Commission on Nuclidic Masses and Related Atomic Constants of the I.U.P.A.P. by E. R. Cohen and J. W. M. DuMond, June 24, 1963.

PROBLEMS

1. Suppose that the lanterns used for signals by Galileo and his assistant were 6 km apart. If the error in this distance were negligible and the speed of light was to be accurate to within 10% of the accepted value, what is the maximum error permissible in the determination of the time between the opening of the two lantern shutters?
2. (a) The Crab Nebula is a luminous remnant of a supernova (exploding star) event observed by Chinese astronomers in A.D. 1054. Assuming the exploding star was originally a sphere of negligible diameter and that it has been expanding uniformly at 10^3 km/sec since the event was observed, estimate its diameter today.
   (b) The light year is the distance that light traveling in a vacuum could cover in one year. Find the distance in (a) in light years.

3. In 1862, Foucault employed a rotating plane mirror instead of a toothed wheel in determining the speed of light. With the speed of rotation at 800 rev/sec, a ray of light that is reflected from the rotating mirror to travel 15 m to a distant plane mirror and be reflected back to the rotating mirror again encounters the altered position of the rotating mirror shown dashed in Figure 9-3. This results in an angular displacement of 3.45 minutes of arc for the returning ray. Find the speed of light from this data. Hint: A reflected ray of light experiences an angular displacement 2θ when the plane reflecting surface experiences an angular displacement θ.

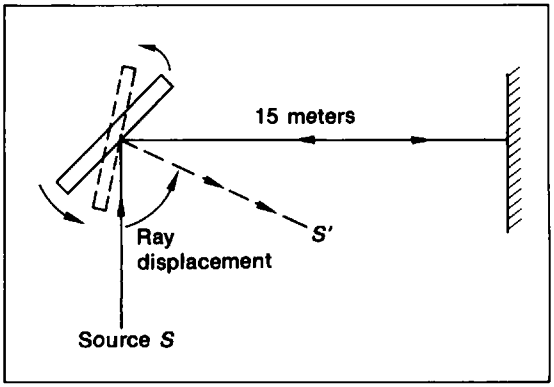


Figure 9-3

4. The Sun is on the average 1.49×10^{11} m from the Earth. How long does light take to travel this distance?
5. In 1925, Michelson modified the rotating mirror method by using an 8-sided mirror that could be rotated at high speed. Light from a source was reflected from one face to a distant plane mirror, then reflected back to another face, and from there reflected to a telescope for observation. (See Figure 9-4.) A series of light pulses from the source hit face 1 of the rotating mirror. If the rotation speed of the mirror is adjusted so that face 2 has moved to replace face 3 while the light was going down to and back from the distant mirror, then the light pulses will appear in the telescope. If the distant mirror was 37.46 km away and the rotational speed of the mirror was 500 rev/sec, what value was obtained for the speed of light?

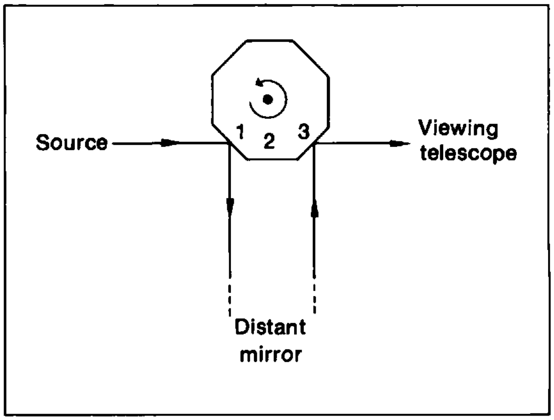


Figure 9-4

10 Waves and Wave Motion

10-1 THE NATURE OF A WAVE

We present here a discussion of the properties or characteristics of waves or wave motion, using examples taken from our everyday experience. This will help us evaluate the assertion of the previous chapter that light can be satisfactorily understood as a wave phenomenon. In addition, the concepts developed here will be useful in studying other examples of wave phenomena in later chapters.

Common examples of wave motion are easy to find: the ripples on the surface of a lake or pond produced by wind action or by casting stones into them, pulses traveling along a long string (or along helical springs such as the Slinky toy) fastened at one end, and the sound pulsations detected by the ear as the result of the ringing of a large bell. While these examples differ in detail, they all have certain features in common. In each case, there is a medium (the water, string or spring, or the air, respectively) through which a disturbance (the water ripples, the pulses, or the sound pulsations) travels. While the disturbance is passing a given point in the medium, the particles of the medium in the neighborhood of the point execute small displacements about their original undisturbed positions. After the disturbance has passed, however, the medium returns to its undisturbed state. Thus, no net displacement or alteration of the medium results from the passage or transmission of a propagating wave. Because of this fact, it is possible to direct our attention to a mathematical description of a general propagating wave without having to consider in detail its specific physical nature or the medium in which it propagates.

Note carefully that this is *not* equivalent to saying all wave phenomena are similar in detail. In particular, we mention here two distinct types of wave motion (to which we shall return in Chapter 13)—namely, longitudinal waves and transverse waves. A longitudinal wave is one in which the motion of the particles of the medium is parallel to the direction of propagation of the wave, such as a sound wave traveling along an air column. A transverse wave is one in which the particles of the medium execute motion in directions perpendicular to the direction of propagation of the wave, for example, the vibrating string along which a wave propagates.

10-2 A MATHEMATICAL DESCRIPTION OF A WAVE

In order to avoid mathematical complications, we assume a periodic (regularly repeating) disturbance of the medium (for visualization, consider a surface wave on water). Specifically, we assume that a side view of the surface wave can be described mathematically by the relation

$$y = A \sin \theta. \qquad (10\text{-}1)$$

Here θ is the phase angle of the wave and depends generally upon both the time of observation of the wave and the location in the medium. For a given phase angle, y represents the actual displacement of the water surface from the undisturbed condition, and A, the amplitude of the disturbance, represents the maximum displacement from the equilibrium or undisturbed condition. A real water wave would be a more complicated function of the variable θ than the sine function, but it is a remarkable fact (discovered by the French physicist Joseph Fourier) that the true wave form can be obtained by the appropriate addition of a series of different sine and cosine terms of the type indicated in Eq. (10-1). We therefore concentrate on this simple wave, since it can be used as the basis for describing more complex cases.

Figure 10-1 is a graph of the assumed wave form as a function of the angle θ in radians. The distance between successive identical displacements is called the *wavelength* and is designated by λ. Suppose that the wave is traveling to the right with a speed v, and that it takes a time interval τ, called the *period* of the wave, for one complete wavelength of the wave to pass a given point. This means that the medium at that point will have experienced, during a time interval τ, all possible values of the displacement that the wave can cause. Then the speed of propagation v is related to τ and λ by the equation

$$v = \lambda/\tau. \qquad (10\text{-}2)$$

That is,

$$v\ (\text{m/sec}) = \lambda\left(\frac{\text{meters}}{\text{wave}}\right)\Big/\tau\left(\frac{\text{seconds}}{\text{wave}}\right).$$

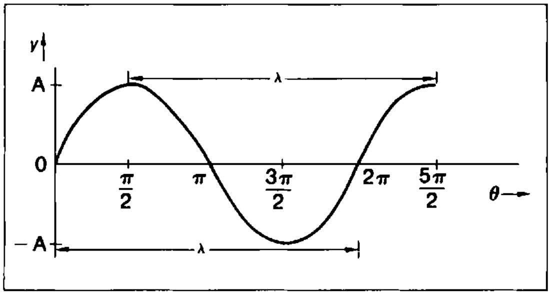


Figure 10-1 Sinusoidal wave form.

If we define the number of complete wavelengths passing a given point per unit time as the *frequency* f, it is clear that it is necessarily equal to $1/\tau$. Therefore, Eq. (10-2) may be rewritten as

$$v = \lambda f. \qquad (10\text{-}3)$$

The speed of propagation depends upon the physical properties of the medium.

The displacement y of a given particle of the water surface will be a function of both the time of observation and the location in a direction along the surface x. That is, $y = y(x, t)$. Let us now choose an origin for the x coordinate by requiring that $y = 0$ for $x = 0$ and $t = 0$. From Figure 10-1 it can be seen that when $x = \lambda/2$, $y = 0$ and when $x = \lambda$, $y = 0$. This suggests that we write $\theta(x, 0) = 2\pi x/\lambda$; the desired sinusoidal wave form is thereby attained. Thus, a sinusoidal wave at $t = 0$ will have the equation

$$y(x, 0) = A \sin\left(\frac{2\pi x}{\lambda}\right). \qquad (10\text{-}4)$$

Now, at a later time t, the wave form will have traveled to the right a distance vt. The particle displacement at an arbitrary point x at time t will necessarily be the same as that for a particle located at the same position x' at $t = 0$ if x is related to x' by the equation

$$x' = x - vt.$$

Therefore,

$$y(x', 0) = A \sin\frac{2\pi x'}{\lambda} = y(x, t),$$

so that

$$\begin{aligned} y(x, t) &= A \sin\frac{2\pi}{\lambda}(x - vt) \\ &= A \sin 2\pi\left(\frac{x}{\lambda} - ft\right) \\ &= A \sin 2\pi\left(\frac{x}{\lambda} - \frac{t}{\tau}\right) \qquad (10\text{-}5) \end{aligned}$$

is the equation for a transverse wave traveling to the right. The angle $\theta(x, t) = 2\pi/\lambda\ (x - vt)$ is, as noted after Eq. (10-1), called the phase angle of the wave. For a sinusoidal wave traveling to the left, a similar analysis yields the same expressions with v replaced by $(-v)$.

10-3 SUPERPOSITION OF WAVES

Let us now consider the consequences of subjecting the same medium to two (or more) sinusoidal traveling waves at the same time. Experiments for many types of wave phenomena (but not all) indicate that for small wave amplitudes the net disturbance of the medium at any point is obtained simply by adding algebraically the individual disturbances at the point. This useful principle, which in any situation should be subject to experimental verification, is known as the principle of linear superposition. Assuming that linear superposition is valid, we can write the following expression for the net disturbance at a point x in the medium at a time t by two different sinusoidal traveling waves

$$\begin{aligned} y_{net} &= y_1 + y_2 \\ &= A_1 \sin 2\pi\left(\frac{x}{\lambda_1} - f_1 t\right) + A_2 \sin 2\pi\left(\frac{x}{\lambda_2} - f_2 t\right), \end{aligned} \tag{10-6}$$

where individual disturbances have wavelengths λ_i and frequencies f_i. For N separate waves,

$$y_{net} = \sum_{n=1}^{N} A_n \sin \theta_n,$$

where

$$\theta_n = 2\pi\left(\frac{x}{\lambda_n} - f_n t\right). \tag{10-7}$$

Consider as an example two waves of identical frequency ($f_1 = f_2 = f$). Because the speed of propagation depends on the medium $v_1 = v_2 = v$, so $\lambda_1 = \lambda_2 = \lambda$. For the first wave, Eq. (10-7) yields

$$y_1 = A_1 \sin \theta_1 = A_1 \sin \frac{2\pi}{\lambda}(x - vt).$$

Let us also assume that the expression for y_2 can be replaced by

$$y_2 = A_2 \sin\left[\frac{2\pi}{\lambda}(x - vt) + \phi\right]. \tag{10-8}$$

In Eq. (10-8) the quantity ϕ is known as the phase angle difference between the two waves at the point x at time t, or, in terms of the phase angles, $\phi = \theta_2 - \theta_1$. Substitution of Eq. (10-8) in Eq. (10-6) together with $\lambda_1 = \lambda_2 = \lambda$ and $v_1 = v_2 = v$ yields

$$y_{net} = A_1 \sin \frac{2\pi}{\lambda}(x - vt) + A_2 \sin\left[\frac{2\pi}{\lambda}(x - vt) + \phi\right]. \tag{10-9}$$

Suppose now that ϕ is an integer multiple of 2π. Since $\sin(\alpha \pm 2\pi n) = \sin \alpha$ for integer values of n, we obtain

$$y_{net} = (A_1 + A_2) \sin \frac{2\pi}{\lambda}(x - vt), \tag{10-10}$$

and the amplitudes add in magnitude. The waves in this case are said to be *in phase* and *constructive interference* results. On the other hand, if the phase difference is an odd integer multiple of π, then $\sin[\theta \pm (2n - 1)\pi] = -\sin\theta$; we obtain

$$y_{net} = (A_1 - A_2) \sin \frac{2\pi}{\lambda}(x - vt), \tag{10-11}$$

and the amplitudes combine subtractively in magnitude. In this case, the waves are said to be *out of phase* and *destructive interference* results, producing a reduced amplitude. If $A_1 = A_2$ for this case, there is a complete cancellation and no net disturbance results. For values of the phase differences other than the cases considered, Eq. (10-9) will yield a result intermediate between the constructive and the destructive limiting cases.

At this point we might well ask what factors determine the phase angle difference ϕ for two given sinusoidal waves observed at (x, t). To answer this question, we must consider the sources that produce the two waves. If the sources are said to be *in phase*, this means that they emit the wave disturbances simultaneously or in synchronization. On the other hand, if they are *out of phase*, we understand this to mean that there is a delay of one half a cycle (or wavelength) between the emission of a wave by one source and the subsequent emission of a wave by the second. Suppose the two sources are located at the same point. Then the net disturbance produced by them at any other point will be the sum of the amplitudes of the two disturbances for sources in phase. It will be the difference between the amplitudes of the two disturbances if they are out of phase, since in this case the crest of one wave will occur at the same time and place as a trough of a wave from the other wave, which is either a half cycle ahead of or behind the first. More generally, one can specify this phase relationship between the two sources by specifying the lag angle δ by which the wave disturbance from the one source lags behind the other. Then the expression *in phase* means $\delta = 0$, while *out of phase* means $\delta = \pi$.

A second factor producing a phase angle difference between two waves is related to the locations of the two sources relative to the point at which we observe the net disturbance. To see this, consider first two sources that are in phase ($\delta = 0$), that are respectively at distances x_1 and x_2 from the point in space. For identical sources (equal wavelengths λ

and frequencies f), one would expect an addition of amplitudes to occur if the difference in distances from the sources to the point in space is equal to an integer number of wavelengths. That is, if the difference in path length traversed by the waves can be written as

$$\Delta x = x_2 - x_1 = n\lambda, \quad n = 0, 1, 2, \ldots, \qquad (10\text{-}12)$$

we can expect an addition of amplitudes. On the other hand, when

$$\Delta x = (2n+1)\frac{\lambda}{2}, \quad n = 0, 1, 2, \ldots, \qquad (10\text{-}13)$$

the amplitudes will subtract.

Conversely, when $\delta = \pi$ (sources out of phase), the wave amplitudes will add when

$$\Delta x = (2n+1)\frac{\lambda}{2}, \qquad (10\text{-}14)$$

and subtract when

$$\Delta x = n\lambda, \quad n = 0, 1, 2, \ldots. \qquad (10\text{-}15)$$

(Why?)

We can express the phase angle difference that is due to this difference in path length by the quantity $2\pi\Delta x/\lambda$. The total phase angle difference between the two sources discussed in the examples above is then given by the relation

$$\phi = \frac{2\pi\Delta x}{\lambda} + \delta. \qquad (10\text{-}16)$$

It is not difficult to see that the effect produced by a given path length difference can be cancelled out by an appropriate choice of lag angle δ, and conversely. In any case, then the interference of two or more sinusoidal waves at a point in space will depend upon both the synchronization of the sources (lag angle δ) and the difference in path length from the sources to the point in question $2\pi\Delta x/\lambda$. When these factors are known, the nature of the interference can be predicted. Conversely, if the interference is observed and the location of the sources is known, one can deduce the value of the lag angle δ.

As a further example, consider two waves of equal amplitude but slightly different frequencies. If they are impressed simultaneously on the same medium, we can say that at some point (for convenience let us choose $x = 0$) the total amplitude is

$$y_{net} = A\,[\sin 2\pi f_1 t + \sin 2\pi f_2 t]. \qquad (10\text{-}17)$$

It is useful here to employ the identity

$$\sin\alpha + \sin\beta = 2\left[\sin\left(\frac{\alpha+\beta}{2}\right)\right]\left[\cos\left(\frac{\alpha-\beta}{2}\right)\right].$$

This yields

$$y_{net} = 2A\sin\left[2\pi\left(\frac{f_1+f_2}{2}\right)t\right]\cos\left[2\pi\left(\frac{f_1-f_2}{2}\right)t\right]. \qquad (10\text{-}18)$$

We can interpret this as a wave disturbance whose frequency is the mean of the two separate frequencies and whose amplitude varies with time as

$$2A\cos\left[2\pi\left(\frac{f_1-f_2}{2}\right)t\right].$$

Thus, there will be a maximum disturbance whenever the cosine term takes the values ± 1. There will, therefore, be two such maxima per cycle of amplitude variation. Such maxima are known as beats, and the beat frequency is thus twice the frequency of amplitude variation.

$$\text{beat frequency} = 2\left[\frac{f_1-f_2}{2}\right] = f_1 - f_2. \qquad (10\text{-}19)$$

For audible sound waves, the unaided ear can discern beat frequencies up to approximately 10 cycles/sec.

When two waves of equal amplitude and frequency travel in opposite directions in a medium, the resulting disturbance is called a standing wave. Mathematically stated,

$$y_{net} = A\left\{\sin\left[2\pi\left(\frac{x}{\lambda} - ft\right)\right] + \sin\left[2\pi\left(\frac{x}{\lambda} + ft\right)\right]\right\}. \qquad (10\text{-}20)$$

Since $\sin(\alpha \pm \beta) = \sin\alpha\cos\beta \pm \cos\alpha\sin\beta$, the expression can be written

$$y_{net} = [2A\cos 2\pi ft]\sin\frac{2\pi x}{\lambda}. \qquad (10\text{-}21)$$

We see that, for all values of the time, the net disturbance is zero for $x = n\lambda/2$, where n is any integer. For all other values of x, the amplitude varies sinusoidally with the time through the factor in square brackets.

Consider as an example the case of a stretched string of length L with fixed ends. When the ends are fixed, y_{net} must vanish at $x = 0$ and at $x = L$, for any value of t. Equation (10-21) then shows that the only wavelengths allowed are those for which the equation

$$\lambda_n = \frac{2L}{n}, \quad n \text{ any integer}, \qquad (10\text{-}22)$$

is satisfied. The corresponding frequencies are given by the equation

$$f_n = \frac{nv}{2L} = nf_1, \qquad (10\text{-}23)$$

where $f_1 = v/2L$ is called the fundamental frequency or first harmonic, while the nth frequency is called the nth harmonic. Another type of notation refers to higher frequencies as overtones of the fundamental. In this notation, the second harmonic is the first overtone, etc. The speed of propagation v arises through Eq. (10-3), $v = f\lambda$.

10-4 REFLECTION AND REFRACTION OF WATER WAVES

We turn now to a physical discussion of what happens when waves encounter complete or partial barriers or pass from one medium into another. For convenience of discussion, we choose water waves as observed by means of a ripple tank, a glass bottom tray of water illuminated from above. When a pattern of waves is generated in the tray, the crests of the waves tend to focus the incident light, while the troughs diverge it. As a result, a viewing screen placed below the tray will exhibit an alternating array of bright and dark bands, which are the images of the crests and troughs of the waves being studied. Provision must be made for the absorption of the waves at the edges of the tray if secondary influences are to be avoided. We assume that this has been done for our discussion.†

As an initial example, consider a plane wave train. Such a pattern of disturbance can be generated by periodically dipping a straight edge into the water, producing the result shown schematically in Figure 10-2. If a second plane wave train is developed simultaneously at right angles to the first, the resultant pattern will appear as indicated in Figure 10-3. This is what one would predict by means of the superposition principle discussed in the previous section.

Now consider the effect on a single plane wave train produced by placing a solid barrier diagonally in its path as in Figure 10-4(a). Experiments with ripple tanks show that the waves are reflected as

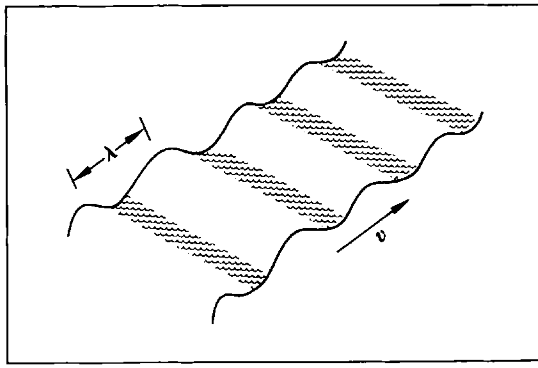


Figure 10-2 Ripple tank plane wave (schematic).

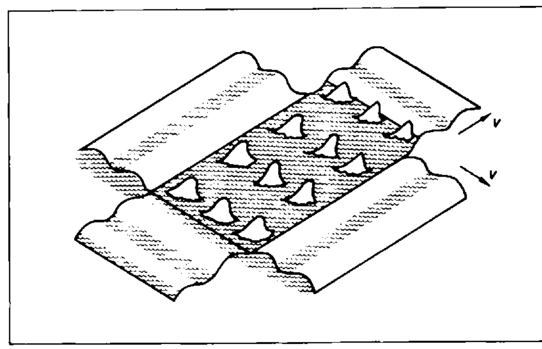


Figure 10-3 Ripple tank superposition of perpendicular plane waves (schematic).

shown in Figure 10-4(b). The angles θ'_i and θ'_r are observed to be equal. As a result, the angles θ_i and θ_r, which are measured with respect to the normal NN' to the barrier, are also equal. This is of course the law of reflection introduced in Chapter 6.

To illustrate refraction with a ripple tank, it is necessary that there be two or more regions of the water in the tray in which the speed of propagation of the water waves assumes different values. This is accomplished by using the fact that for shallow water depths the speed of propagation varies in proportion to the square root of the depth. Thus, if one has a tank filled with water at two different depths, the propagation of water waves in this tank will be analogous to the propagation of light in a system involving two different media. In Figure 10-5(a), we see the effect of a varying depth on a plane water wave propagating in a direction parallel to that of decreasing depth. Since the frequency of occurrence of wave crests is constant, and the speed of propagation decreases in shallower water,

†An excellent article on the construction and use of such a tank appears in the Amateur Scientist section of the October 1962 issue of *Scientific American*.

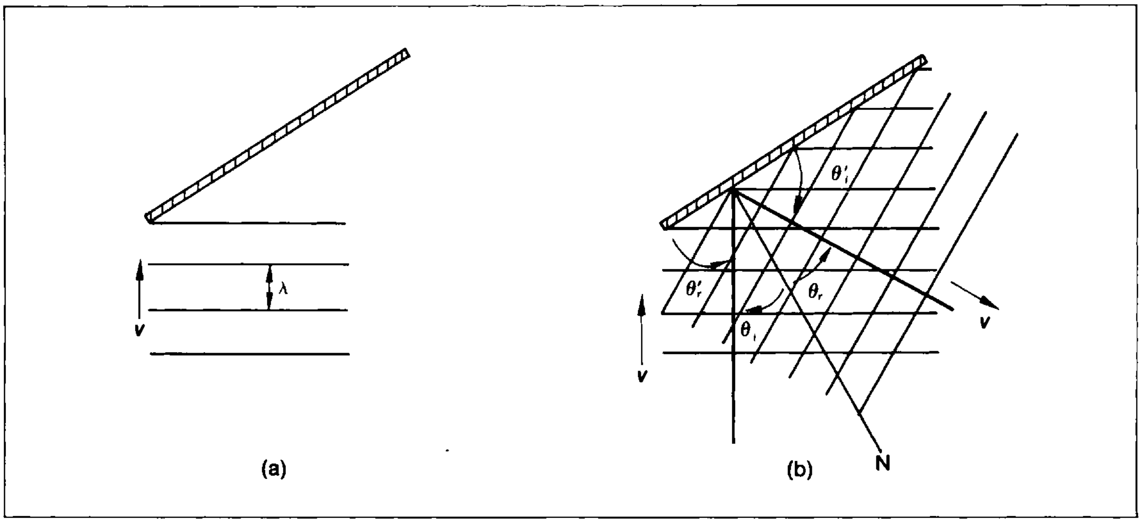


Figure 10-4 Reflected waves in a ripple tank diagonal barrier (schematic).

the wavelength must necessarily decrease in the direction of propagation, as the figure indicates.

In Figure 10-5(b), the plane wave direction of propagation is at an angle with respect to the regions of different depth. It is clear that the direction of propagation is altered as the wave proceeds into the second "medium," so that refraction of the water waves occurs. By constructions similar to that in Figure 10-4(b), we can establish that Snell's law applies here as it does for light. Figure 10-6 represents successive wave crests in the vicinity of the interface with medium 1, the medium of greatest speed of propagation. Notice that the leading crest consists of two segments which, because they are in different media, propagate at different speeds. Since the plane wave crests are lines of constant phase angle, the time required for the second wave crest to travel to the location of the first is a constant independent of position along the wave crest. Thus, the time of propagation for a segment of the wave crest to proceed from P to O is identical to that for a segment proceeding from O' to P'.

Let the speed of propagation in medium 1 be v_1 and that in medium 2 be v_2. The time to traverse the

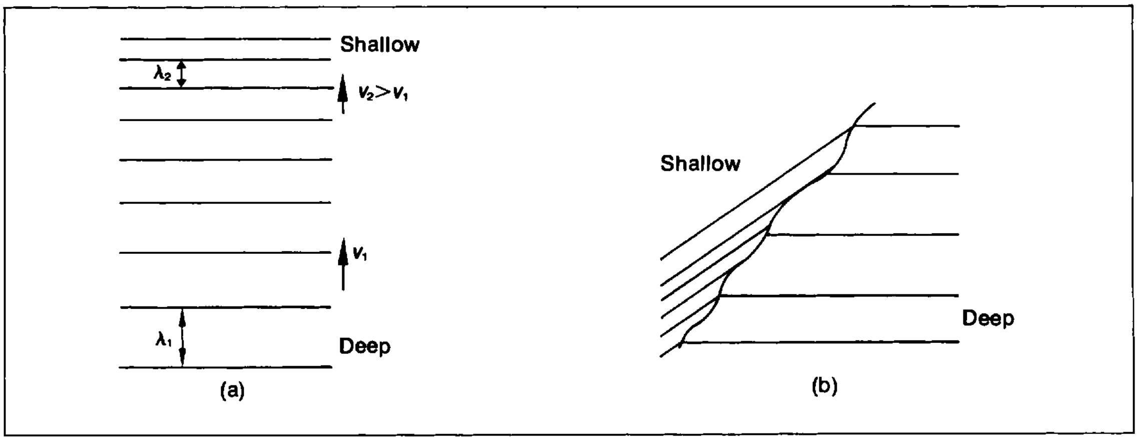


Figure 10-5 (a) Effect of varying depth on wavelength. (b) Refraction of water waves in ripple tank due to varying depth.

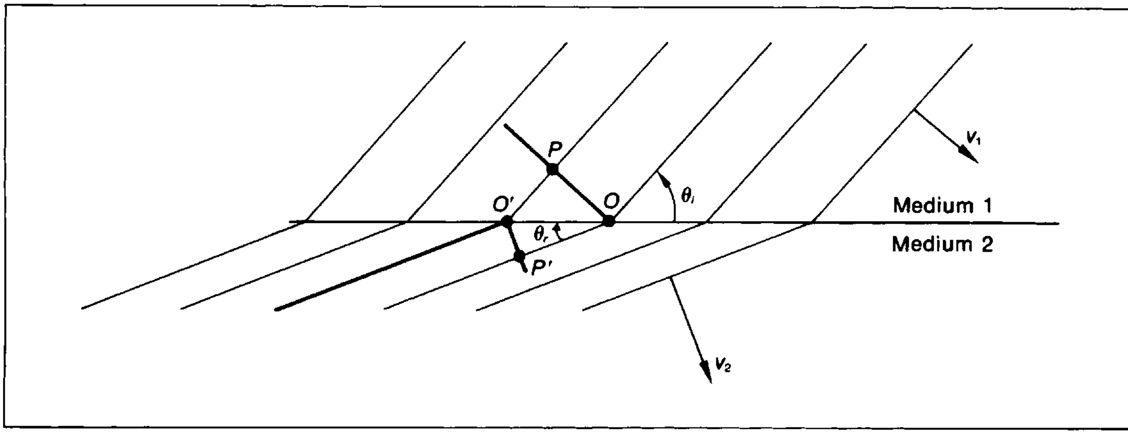


Figure 10-6 Wave crest refraction at a boundary between two media.

distances OP and $O'P'$ can then be written

$$t = \frac{OP}{v_1} = \frac{O'P'}{v_2}. \tag{10-24}$$

But from Figure 10-6,

$$\frac{OP}{O'P'} = \frac{OO' \sin \theta_i}{OO' \sin \theta_r}. \tag{10-25}$$

Substituting Eq. (10-25) into Eq. (10-24), and rearranging, we obtain

$$\frac{\sin \theta_i}{\sin \theta_r} = \frac{v_1}{v_2} = n_{21}. \tag{10-26}$$

We therefore obtain not only Snell's law but find in addition that the index of refraction is given by the ratio of the speeds of propagation in the two media. By analogy, we can infer from these wave theory arguments that if medium 1 were air, medium 2 water, and the wave disturbance were light waves, then the speed of propagation of light in water would be less than that in air since $n_{H_2O\ air} \approx 4/3$. This assertion was verified by Foucault in the mid-nineteenth century using a modification of the Fizeau toothed-wheel apparatus discussed in Chapter 9. This is the reverse of the particle theory prediction for light, and is therefore one of the strong points favoring the wave theory.

10-5 HUYGENS' PRINCIPLE AND DIFFRACTION OF WAVES

In his study of the wave theory of light, Huygens developed a method for constructing by geometric means the behavior of a given wave disturbance for which the initial shape of a wave front is known and for which the variation of the speed of propagation with position is known. The principle upon which the method is based, known as Huygens' principle, can be expressed quite simply. The principle states that each point on the initial wave front can be regarded as a source of spherical wavelets (propagating radially in the forward directions) which travel with the speed of propagation of the wave disturbance in the medium. The envelope of these wavelet wave fronts at any subsequent time represents the wave front resulting from the initial wave front at the later time. In situations for which the speed of propagation varies with position, it is important that successive time intervals be chosen for which the propagation speed does not vary appreciably in the space interval between wavelet envelopes. Figure 10-7 illustrates the application of

Figure 10-7 Application of Huygens' principle for a propagating wave.

Huygens' principle for a wave front of arbitrary shape propagating in a uniform medium.

If we apply Huygens' principle to the case of plane water waves striking a barrier with an aperture, one finds (for wavelengths comparable to the width of the aperture) that the plane wave fronts exhibit a bending around into the "shadow" of the aperture, a situation described as diffraction in the case of light waves. This result, in contrast to the observed rectilinear propagation of light through slits with widths of 1 mm or more; was an apparent defect of the wave theory that led Newton to consider it an untenable model for light. To resolve this difficulty, consider the series of ripple tank photographs in Figure 10-8(a–c). In this series, the aperture width is held constant while the wavelength of the wave fronts is reduced from a value six-tenths that of the aperture width d [Fig. 10-8(a)] to a value three-tenths of d [Fig. 10-8(b)] to the last case for which the wavelength is one-tenth d [Fig. 10-8(c)]. It is clear that for wavelengths that are small compared to the aperture width the diffraction effects become negligible.

The resolution of our puzzle in the case of light, therefore, requires the assumption that the wavelengths of light are small compared to the dimensions of 1 mm or more. For such a situation, one would not expect to observe optical diffraction effects unless the apertures were reduced sharply in width. In the next two chapters, we will consider in more detail the interference and diffraction of light waves. In Chapter 36, it is indicated that visible light represents a rather narrow band of wavelengths in the spectrum of electromagnetic waves extending from below gamma rays of very short ($\sim 10^{-10}$ m) wavelength to beyond radio waves of long wavelength ($\sim 10^{3}$ m). In this spectrum, visible light possesses wavelengths ranging from $\sim 7 \times 10^{-5}$ m (red) to $\sim 4 \times 10^{-5}$ m (violet). Accepting for now the accuracy of this assertion, it is easy to understand how apertures of ordinary dimensions failed to exhibit diffraction effects.

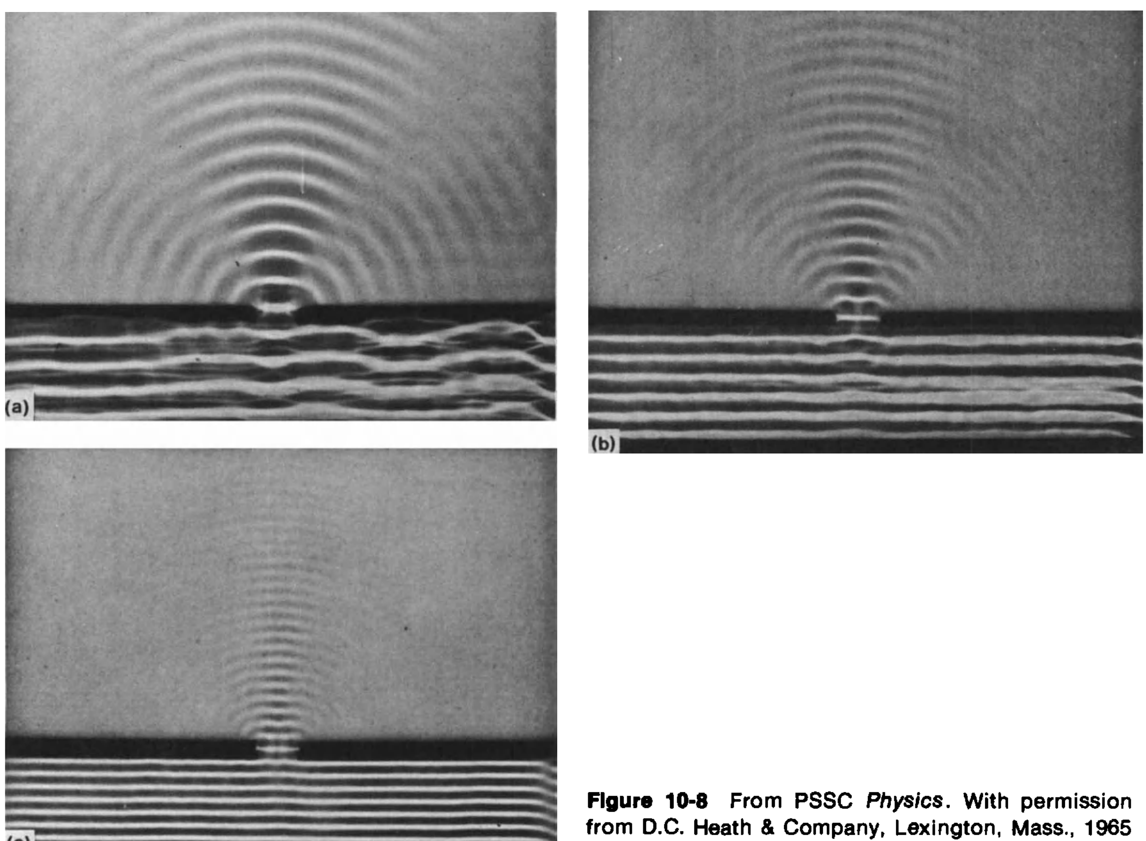


Figure 10-8 From PSSC *Physics*. With permission from D.C. Heath & Company, Lexington, Mass., 1965 and Education Development Center, Newton, Mass.

PROBLEMS

1. A sawtooth wave form is in effect a linearized sine curve. (See Figure 10-9.) Using methods developed by Fourier, this wave form can be expressed as an infinite series of sine terms as follows:

$$y = \frac{4}{\pi}\sum_{n=0}^{\infty}\frac{(-1)^n \sin(2n+1)\theta}{(2n+1)^2} = \frac{4}{\pi}\left[\sin\theta - \frac{1}{3^2}\sin 3\theta + \frac{1}{5^2}\sin 5\theta - \cdots\right].$$

To appreciate how few terms of the series are needed for a reasonably good fit, plot
(a) the first term,
(b) the sum of the first two terms, and
(c) the sum of the first three terms of the series for values of θ from 0 to 2π in steps of $\pi/4$.

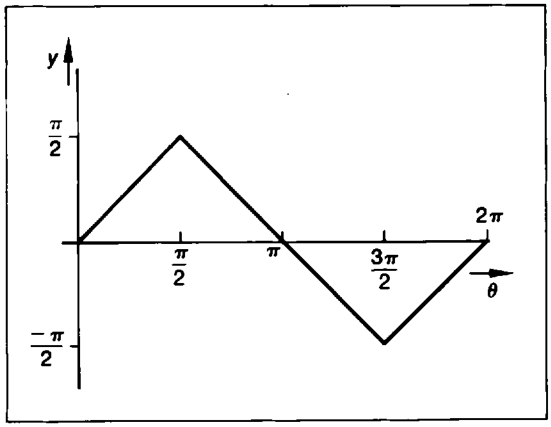


Figure 10-9

2. A sinusoidal water wave has an amplitude of 20 cm and a wavelength of 35 cm. If a photograph of the wave profile is made, what will be the distance between a given crest and a point farther along the wave where the phase angle has increased by 745°?
3. An organ pipe emits a sound wave with a frequency of 20 cycles/sec. The velocity of sound is 344 m/sec. If the pipe is $\frac{1}{4}$ of a wavelength long, what is its length in meters?
4. A sound wave traveling in a given medium is described by the relation

$$y = 10^{-3}\sin 2\pi\,(1.6x - 480t),$$

where y represents the variation in density of the medium (expressed in gm/cm^3) as the wave passes. What are the amplitude, frequency, wavelength, and speed of propagation of the wave? Time is in seconds, distance is in meters.
5. A wave traveling to the left along a string has an amplitude of 2 cm, a wavelength of 5 m, and a speed of propagation of 20 m/sec. Write the equation of the wave, and determine the frequency of the wave.
6. Find the amplitude of a wave made by superposing two waves of the same frequency and wavelength, traveling in the same direction. The first wave has an amplitude A. The second has an amplitude $2A$, and lags the first by $\delta = \pi/3$.
7. Show that Eq. (10-9) can be written in the form

$$y_{net} = A_{net}\sin\left[\frac{2\pi}{\lambda}(x - vt) + \beta\right],$$

where

$$A_{net} = [A_1^2 + 2A_1A_2\cos\phi + A_2^2]^{1/2},$$

$$\tan\beta = \frac{A_2\sin\phi}{A_1 + A_2\cos\phi}.$$

Then show that Eqs. (10-10) and (10-11) follow readily from this result.

8. Two sinusoidal waves of equal amplitude, frequency, and wavelength are traveling in the same direction in a medium but have a phase difference of $\pi/2$. Find an expression for the disturbance assuming linear superposition is valid.
9. Two tuning forks with the same frequency are sounded simultaneously (in phase and equal amplitude). The frequency of the forks is 112 cycles/sec and the wavelength of the waves is 10 ft.
   (a) What will be the phase angle difference between the waves at a point 40 ft from one fork and 45 ft from the other?
   (b) What will be the sound amplitude at this point?
   (c) What will be the phase angle difference at a point equidistant from the two forks?
   (d) What will be the phase angle difference at a point 2.5 ft closer to one fork than the other?
10. An oscillator generates waves of 4 m. A second oscillator is 10 m north of the first oscillator. It generates waves of the same wavelength and amplitude as the first, but it is 180° out of phase with the first. Measuring east from the first oscillator, find the locations where the resultant sound amplitude is zero.
11. A vibrating tuning fork produces 11 beats/sec when placed near a standard fork of frequency 512 cycles/sec. What is the frequency of the first fork?
12. A stretched string $\frac{1}{2}$ m in length has fixed ends. If the speed of propagation of a wave on this string is 250 m/sec, find:
   (a) the fundamental frequency,
   (b) the frequencies for the first 3 overtones that exhibit nodes (zero disturbance) at the midpoint of the string, and
   (c) the beat frequency that would result if the 3rd harmonic frequency were compared to a tuning fork with a frequency of 768 cycles/sec.
13. Suppose waves are excited on a string that has been formed into a circle of radius r. If the wavelength of the waves is λ, find the relation that must be satisfied if standing waves are to be formed on the string.
14. The speed of propagation of waves on the surface of a certain tank of liquid is given by

$$v = 1.5 + 1r^2,$$

   where v is in m/sec and r is the perpendicular distance (in meters) along the surface from a reference line of the surface. Using a sheet of graph paper, use Huygens' principle to trace successive wave fronts for a plane wave front that is initially perpendicular to the reference line.
15. Above a heated surface (such as a road bed or a desert heated by the sun), the speed of propagation of light is greater than in cooler air. Assume that this variation in propagation speed with distance from the surface is gradual rather than abrupt. Use this information to explain the mirage. This is a phenomenon in which light rays from an extended object travel to the observer directly from the upper portion of the object (through the cooler upper air) and also by curving path downward toward the surface and upward again to the observer. Thus, the observer sees both the actual object and an inverted image (why?) below it as though a reflecting surface lay between the two. In the desert, this apparent reflecting surface is interpreted by the thirsty traveler as a body of water.

11 Interference

11-1 INTERFERENCE OF LIGHT WAVES

In Section 10-3, we discussed conditions leading to constructive or destructive interference when two or more sinusoidal waves simultaneously propagate in a given medium. It was assumed in that discussion that the lag angle, δ, between any two sources is constant during the period of observation. Sources satisfying this condition are called coherent sources. If the lag angle, δ, and the amplitude of the net disturbance vary randomly, the sources are said to be incoherent. In the case of sound waves, two coherent sources could be obtained by driving two loudspeakers simultaneously by a single audio oscillator (a device for generating audible sine waves). If, on the other hand, one of the loudspeakers were switched on and off in a completely random fashion, it is possible to show that the superposition of the two loudspeaker outputs will produce a net disturbance exhibiting a frequency identical to that of the driving audio oscillator. Now, however, the amplitude and the phase angle difference will be seen to vary in a completely random fashion; that is, the sources are incoherent.

The relevance of this discussion to the case of light waves becomes clear when it is noted that light waves are produced when atoms that have been suitably excited (by heating in a flame, for example) make transitions to less excited states by the emission of light. The time during which a typical atom radiates light in such a transition is about 10^{-8} seconds. Although this would seem to be a very short time interval, it is in fact an interval long enough for a light wave to undergo about 10^{7} oscillations. (This follows from the fact that the frequency of light waves is about 10^{15} cycles/sec as determined from the value of the speed of light and the measured values of the wavelength for light waves.) Since individual atoms will be making transitions at random times if they have been excited thermally, it is clear that the light radiated by two such atoms will be incoherent in character and interference effects will not be observed. Similarly, two incandescent light bulbs will be incoherent light sources since each is a collection of atoms that are thermally excited by the passage of an electric current through a metallic filament (such as tungsten).

Highly coherent sources of light can be produced by a different means of excitation, which was developed by Charles Townes and Arthur Schawlow in 1960. This means of excitation is employed in the device called a laser, a name which derives from the expression *l*ight *a*mplification by *s*timulated *e*mission of *r*adiation. A complete discussion of laser characteristics involves the quantum theory of matter, which is beyond the scope of the present discussion. It is possible, however, to make several general statements regarding laser operation. First, the atoms constituting the active part of the laser substance (such as chromium atoms in a ruby crystal) are raised to an excited state, not by thermal means, but by a process known as optical pumping. In optical pumping, light of a higher frequency than that which will be subsequently emitted is absorbed by the active laser material. Now when light of the frequency to be amplified by laser action enters the laser device, the excited atoms are stimulated to emit light of the same frequency as they make transitions to a less excited state. The incident light and the emitted light are coherent, and the net disturbance represents an amplification of the incident light. If this coherent signal is made to travel repeatedly through the laser material (by positioning the laser material between perfectly parallel reflecting surfaces), the signal becomes enormously magnified and remains very nearly monochromatic due to the nature of the excited states of the active atoms. Finally, if one of the reflecting surfaces is actually only partially reflecting, a substantial portion of the amplified signal will be transmitted in a well-defined direction, providing an intense source of coherent radiation. Two such laser beams can be used to provide striking demonstrations of interference effects.†

To exhibit interference effects for light waves without employing laser sources, one might divide the spherical waves propagating from a point source into two or more beams. This can be done, for example, by placing an opaque screen containing two narrow, parallel slits in the path of the light waves. The beams transmitted by the slits are necessarily coherent at the plane of the screen since they are both portions of the same wave front from a single source (so that $\delta = 0$). Therefore, the effects due to superposition of the two beams at any given point beyond the slits would be expected to follow the predictions discussed in Section 10-3. This double slit experiment, first performed by Thomas Young in 1801, was in fact the earliest experimental evidence favoring the wave theory of light.

Figure 11-1 is a schematic illustration of the double slit arrangement. A point source S is located symmetrically with respect to the narrow slits S_1 and S_2, which are separated by a distance d. The light emanating from the coherent "sources" S_1 and

†The reader seeking more details about lasers is referred to articles by Schawlow appearing in the June 1961 and September 1968 issues of *Scientific American.*

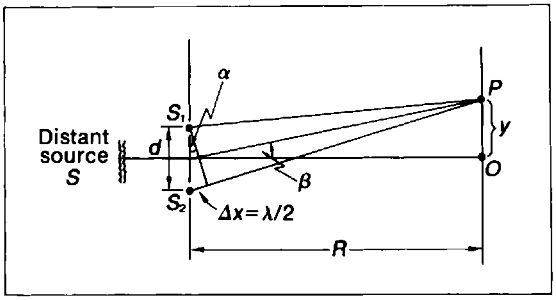


Figure 11-1 Interference at a double slit.

S_2 is then allowed to strike a viewing screen (or photographic plate) placed parallel to the screen containing the two slits at a distance R (much greater than d) from it. The point O is along an axis from S, which bisects the distance d between S_1 and S_2. Thus, at point O, the path difference $\Delta x = 0$ for the two beams. Therefore, the two waves will interfere constructively at O to give a bright line image of either slit. Now suppose that the distance y from O to P represents the distance from this central bright line to the center of the nearest dark line that represents destructive interference. This requires that the distance $S_2P - S_1P = \lambda/2$. From the diagram, we deduce the relations

$$\sin\alpha = \frac{S_2P - S_1P}{d} = \frac{\lambda}{2d} \qquad (11\text{-}1)$$

and

$$\tan\beta \approx \frac{y}{R}. \qquad (11\text{-}2)$$

Since $y/R \ll 1$, the angles α, β will be quite small, and we can assume $\alpha \approx \sin\alpha \approx \tan\beta \approx \beta$. As a result,

$$\frac{y}{R} \approx \frac{\lambda}{2d}, \qquad (11\text{-}3)$$

and, in general,

$$\frac{y}{R} \approx \frac{(2n+1)\lambda}{2d}, \quad n = 0, 1, 2, \ldots, \qquad (11\text{-}4)$$

for dark lines. Similarly, the relation satisfied by the series of bright lines is found to be

$$\frac{y}{R} \approx \frac{n\lambda}{d}, \quad n = 0, 1, 2, \ldots. \qquad (11\text{-}5)$$

If a source of white light is separated by means of a prism into the colors of the spectrum, and the colors are individually used in this experiment, the conditions of Eqs. (11-4) and (11-5) lead to the observation that the wavelengths of visible light fall in a rather narrow range of values between about 4×10^{-7} m at the violet end and about 7×10^{-7} m at the red end. (The values cited are approximate since the limits of visibility depend upon the individual eye involved.) As asserted above, it follows from the speed of light and these wavelength values that the frequency limits for visible light lie between about $\frac{3}{4} \times 10^{15}$ and $\frac{3}{7} \times 10^{15}$ cycles/sec for violet and red light, respectively. Typical dimensions for the experiment are: slit widths of ≈ 0.1 mm, slit separation $d \approx 0.2$ mm and screen position $R \approx 1$ m.

(We note that more convenient units for light wavelengths have been used over the years. Thus, the Angstrom—Å—is equal to 10^{-10} m, so that a wavelength of 4×10^{-7} m becomes 4000 Å. Currently, a more commonly used unit, the nanometer—nm—which equals 10^{-9} m, leads to the equivalence 4×10^{-7} m = 4000 Å = 400 nm.)

There is another feature of light wave propagation that must be taken into account if a completely satisfactory interpretation of interference phenomena is to be obtained in all cases. This is the fact that a light wave reflected from an optically more dense medium will undergo a phase shift of 180° (corresponding to a shift of one-half a wavelength). On the other hand, a light wave reflected from an optically less dense medium is observed to undergo no phase shift. This is easily demonstrated by arranging a point source of light, a viewing screen, and a flat reflecting surface (with index of refraction greater than that of air) as shown in Figure 11-2. By a suitable positioning of the source relative to the reflecting surface, the path difference between a light wave traveling directly from the source to a point P on the screen and one which is first reflected before striking the screen can be made to approach zero. When this is done, destructive interference is observed, leading to the conclusion that a phase shift of one-half a wavelength occurred upon reflection.

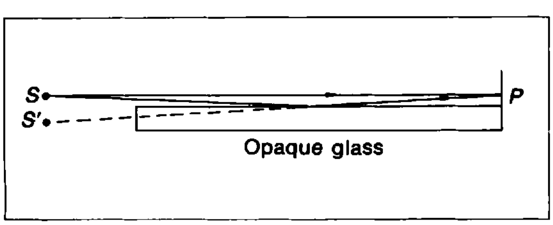


Figure 11-2 Lloyd's mirror experiment.

11-2 APPLICATIONS OF INTERFERENCE

The phase shift upon reflection from an optically more dense medium has practical application in reducing the reflection of light from lens surfaces in optical instruments such as cameras and binoculars. In a camera, for example, one would like to have all light incident on the lens system transmitted to the photographic plate, providing maximum contrast, and at the same time minimizing "ghost effects" produced by stray reflections from the lens surfaces onto the plate. For a given wavelength, this can be accomplished if the surface of the lens is coated with a thin film that has an index of refraction less than the glass of the lens, yet greater than that of air. In addition, the thickness of the film should be equal to a quarter wavelength of the light involved *as measured in the film*, λ_f. The reasons for these requirements are easily explained. First, there will be a phase shift of 180° at both surfaces of the film since at each surface the light proceeds toward a medium of greater refractive index. Therefore, a light ray reflected from the film-glass interface will differ in optical path length by $2 \times \lambda_f/4 = \lambda_f/2$ compared to a ray reflected at the air-film interface, or a 180° phase difference results. In other words, a quarter wavelength thickness of the film creates a favorable situation for transmission rather than reflection, as desired. However, since light entering a camera contains many wavelengths, a compromise must be made. Since the eye and many types of photographic plate are most sensitive to green light, the thickness condition is met for a wavelength in the green range of the spectrum. As a result, any reflection that does occur is a combination of blue and red so that "coated optics," as they are called, usually have a purplish appearance. Reflected light for coated optics is less than 1%, while the value for uncoated optics is from 4% to 5%.

To determine the value of λ_f for a given λ_{air}, we must recall that in our discussion of the nature of light (Chapter 9) it was stated that for a wave theory of light the relative index of refraction is given by the ratio of the speed of light in the less dense medium to the speed in the more dense medium. In our case, using this fact and Eq. (10-3), we obtain

$$n_{f\,\text{air}} = \frac{v_{\text{air}}}{v_{\text{film}}} = \frac{\lambda_{\text{air}} f}{\lambda_f f} = \frac{\lambda_{\text{air}}}{\lambda_f}, \tag{11-6}$$

since the frequency remains constant in proceeding from one medium to another. As a result, the thickness of film required is given by the expression

$$t = \frac{\lambda_f}{4} = \frac{1}{4}\frac{\lambda_{\mathrm{air}}}{n_{f\ \mathrm{air}}}. \qquad (11\text{-}7)$$

Example 1. Magnesium fluoride (MgF_2) has proved to be suitable for coating lenses to reduce reflection. If $n = 1.38$ for $\lambda_{\mathrm{air}} = 5500$ Å, what thickness is required?

SOLUTION

From Eq. (11-7),

$$t = \frac{5500\ \text{Å}}{4 \times 1.38} \approx 1000\ \text{Å} = 10^{-5}\ \text{m}.$$

Another practical application of interference effects is related to the phenomenon known as Newton's rings. If a plano-convex lens with a radius of curvature R is placed on a plate of glass that is accurately flat (Figure 11-3), there will be a thin film of air between the plate and the lens whose thickness is a function of the radial distance r from the point of contact. If this system is illuminated from above by light of wavelength λ, a pattern of interference rings is observed, as in Figure 11-4. In this situation, the interference effects are due to the air film of thickness t at radial distance r. Light reflected from the bottom of the air film undergoes a phase shift of 180°, while there is no shift at the top of the air film. Thus, at the center of the air film ($r = 0$) where the thickness t is essentially zero the net phase shift is 180° and destructive interference occurs, as seen in Figure 11-4. On the other hand, if the condition

$$2t = \left(m + \frac{1}{2}\right)\lambda_{\mathrm{air}}; \quad m = 0, 1, 2, \ldots, \qquad (11\text{-}8)$$

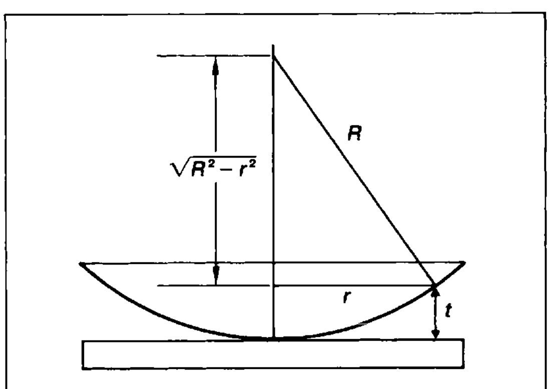


Figure 11-3 Newton's rings geometry.

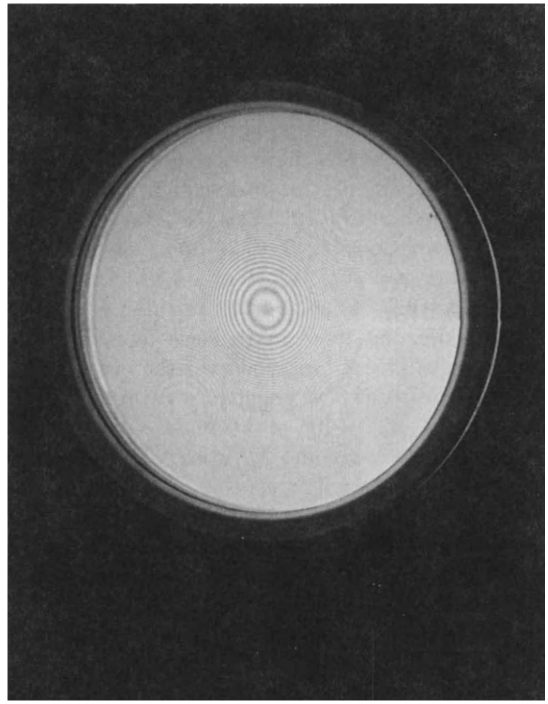


Figure 11-4 Newton's rings pattern viewed by reflection (courtesy of Martin Drexhage).

is satisfied, then constructive interference will occur. From Figure 11-3,

$$t = R - \sqrt{R^2 - r^2} = R\left[1 - \sqrt{1 - \left(\frac{r}{R}\right)^2}\right]. \qquad (11\text{-}9)$$

If $r/R \ll 1$, the binomial expansion yields

$$t \approx \frac{r^2}{2R}, \qquad (11\text{-}10)$$

so that the bright rings are seen to have radii given by

$$r \approx \sqrt{\left(m + \frac{1}{2}\right)\lambda_{\mathrm{air}} R}. \qquad (11\text{-}11)$$

The observation of Newton's rings provides a very sensitive test as to whether or not the lens is symmetrically ground, since any departure from circular rings is an indication of high or low spots on the curved surface of the lens. In a similar way, a flat plate can be tested for flatness by placing it on a known flat plate and placing a thin slip of paper under one edge to provide a thin wedge shaped film of air. It is left as a problem to discuss the pattern of interference fringes in this case.

11-3 THE MICHELSON INTERFEROMETER

As a final example of interference phenomena, we now discuss the interferometer, invented by Michelson and used by him in some of the most sensitive measurements in the entire history of physics. These include the determination of the length of the international standard meter in terms of the red line of the spectrum of cadmium† and the unsuccessful attempt to detect the luminiferous ether (see Chapter 15), which led to the theory of relativity. Figure 11-5 is a schematic diagram of the device. A beam of light from source S strikes a plate of glass A, whose upper surface has been lightly silvered. As a result, the original beam is split into two beams. One beam is reflected and proceeds to mirror M_2 where it is reflected from the front surface back through the plate A to the observer at O. (The glass plate B is the same thickness and is made of the same glass as A and inserted to insure that the two beams experience the same path length in glass. The reader should satisfy himself that this is accomplished.) The second beam passes through plate A toward mirror M_1 and is reflected from the front surface back to A, where it undergoes another reflection at the upper surface, and emerges parallel to the first beam traveling toward O.

If $l_1 = l_2$ (equal optical paths) one would observe a bright spot at O due to the fact that both beams have undergone the same phase shifts upon reflection. Now if mirror M_1 is moved backwards (as shown by the dashed lines in Figure 11-5) a distance $\lambda/4$, the bright spot at O will become a dark spot since the first beam will have experienced an optical path $2(\lambda/4) = \lambda/2$ greater than the second beam, and destructive interference must occur. If the motion of M_1 is increased to $\lambda/2$, then a bright spot will reoccur. The apparatus is designed so that this distance can be measured directly. It is clear, therefore, that by counting a large number of alternations of bright and dark spots at O and measuring the distance M_1 is moved, one can obtain wavelength measurements that are accurate to a fraction of a wavelength. We have considered only the central ray of light from S_1, whereas the light beam incident upon A actually will be cone-shaped.

†Today the meter is defined to be 1650763.73 wavelengths of the orange red light of the Krypton-86 isotope as discussed in *Scientific American*, December 1960, p. 75.

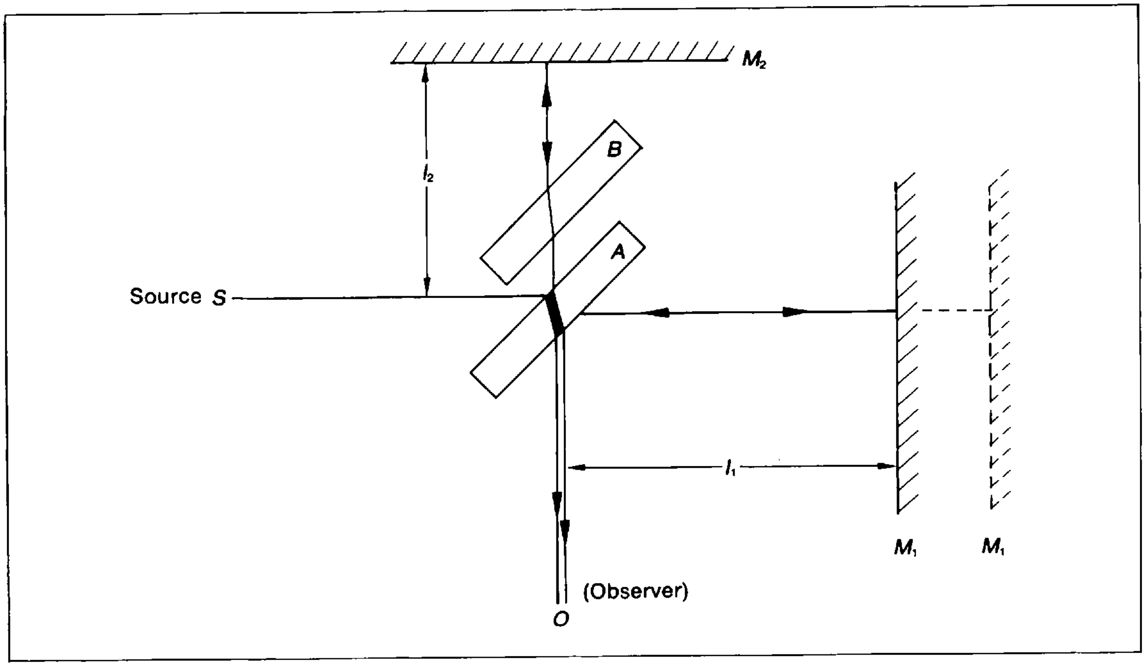


Figure 11-5 The Michelson interferometer.

As a result, rather than a single spot at O, one would expect a series of circular fringes analogous to those seen in the Newton's rings phenomenon. By suitable adjustment of the mirrors M_1 and M_2 and the source S, one can obtain a series of alternating linear bright and dark fringes instead of rings.

As an example of the type of measurement that is possible, suppose that a transparent material with index of refraction $n(> 1)$ and unknown thickness t is inserted in the path of the beam that is reflected at mirror M_2. Then the beam will traverse in the film an optical path length

$$2t = N_t\lambda_f = \frac{N_f\lambda_{air}}{n}, \qquad (11\text{-}12)$$

where N_t is some number. In the absence of the film, a thickness t of air (we assume $n_{air} \approx 1$) is equivalent to an optical path length

$$2t = N_a\lambda_{air}, \qquad (11\text{-}13)$$

where N_a is also some number. Therefore, if the material is placed so that one can observe half the beam after traversing the material and half after traversing the same thickness of air, one would expect a shift of the fringe pattern that is equal to

$$(N_t - N_a)\lambda_{air} = 2t(n-1). \qquad (11\text{-}14)$$

For thin films $(N_t - N_a) < 1$, so that fractional fringe shifts are observed. Knowing $(N_t - N_a)$, λ_a and n, the value of t is easily found. Figure 11-6 is a sketch of such a shift of fringes for a collodion film. In one case, $\lambda_a = 5461$ Å, $n = 1.18$, and the fringe shift $(N_t - N_a)$ was estimated to be about $\frac{1}{5}$ of a wavelength. From Eq. (11-14), it follows that

$$t \approx 3 \times 10^{-5}\,\text{cm}.$$

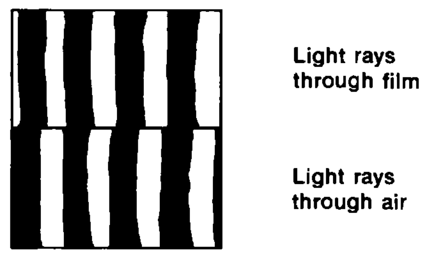


Figure 11-6 Fringe shift in collodion film.

PROBLEMS

1. A monochromatic (single wavelength) point source of light is used to produce light that is incident upon two narrow slits separated by 0.5 mm. Adjacent bright fringes are 1.36 mm apart when observed on a screen 1 m from the slits. What is the wavelength of the incident light?
2. What is the separation distance between the central bright fringe and the second bright fringe on either side when light of wavelength 5500 Å passes through two slits 1.5 mm apart and is incident on a screen 180 cm away?
3. A thin layer of water ($n \approx \frac{4}{3}$) on the surface of a layer of oil ($n < \frac{4}{3}$) shows interference colors. For normally incident light of wavelength 4000 Å, what is the thickness of the water layer?
4. Two pieces of flat glass are separated at one end by a piece of paper. When illuminated by a vertical beam of light of wavelength 7000 Å, 60 dark fringes are observed across the upper plate from one edge to the other. What is the thickness of the paper?
5. Explain why the interference pattern of bright and dark fringes observed in the transmitted light is just the opposite of that observed in reflected light in the Newton's rings phenomenon.
6. A beam of white light is incident normal to a glass film ($n = 1.50$) that is 4×10^{-6} m thick. What wavelengths in the visible spectrum will be most pronounced in the reflected beam? The film is held in air.

7. A beam of light containing two wavelengths, 4500 Å and 6000 Å, passes through two narrow slits separated by 0.5 mm. A viewing screen is placed 3 m from the slits.
   (a) At what distance from the central bright fringe (of both wavelengths) will the bright fringe of one wavelength fall exactly on the dark fringe of the other?
   (b) What will be the orders of these overlapping fringes? The order of the interference is given by the values of n in Eqs. (11-4) and (11-5).
8. A system of Newton's rings is formed when a plano-convex lens resting on a flat glass plane is illuminated from above. The radius of the 25th bright ring is 1 cm when light of wavelength 6000 Å is used. Find the thickness of the air film at the 25th ring and the radius of curvature of the lens surface.
9. A system of Newton's rings are observed when a plano-convex lens resting on a flat glass plane is illuminated from above by light of 4800 Å. If the radius of curvature of the lens is 10 m, and the diameter of the lens is 4 cm, how many dark rings are visible?
10. A thin film of transparent material with an index of refraction $n = 1.50$ is placed in one of the beams of a Michelson interferometer. It is observed that when light of wavelength 5000 Å is used the insertion of the film produces a shift of 30 fringes. What is the film thickness?
11. A Michelson interferometer can also be used to determine the index of refraction of a gas such as air. This is done by putting a cell with plane glass windows in one of the beams and adjusting the interferometer to get a fringe pattern. Then the gas is pumped out and the fringe shifts are counted. If the index of refraction for air is 1.000293, and the cell is 8 cm long, how many fringe shifts would be counted using 5000 Å light?
12. A precision micrometer gauge is to be checked using a Michelson interferometer. Using red light from cadmium ($\lambda = 6439$ Å), an observer counts 4640 fringes while one mirror is moved a distance of 1.500 mm *as measured by the gauge.* What is the actual distance moved? What is the percentage error in the distance as given by the gauge?

12 Diffraction

12-1 INTRODUCTION

When the wave front of a propagating light wave passes by the edge of an obstacle it is observed that:

1. within the geometrical shadow of the obstacle the light intensity is not zero, although it decreases rapidly with increasing distance into the shadow; and
2. bands of unequal light intensity are observed in the region near the edge of the obstacle.

This combination of effects is called a diffraction pattern and is due to the removal of a section of the incident wave front by the obstacle. The first explanation of diffraction phenomena is due to Fresnel, who used Huygens' principle (with modifications) to show that there is no fundamental distinction between diffraction and interference. In the case of diffraction, the amplitude at any point on a propagating wave front can be determined by regarding every point on a previous wave front as a source of secondary wavelets. Then, at the point on the new wave front the amplitudes of each arriving wavelet (with differences in phase properly considered) are added. It is however necessary (and reasonable) to assume that wavelets propagating to the point in directions that are increasingly oblique to the forward direction of the wave front will make correspondingly smaller contributions to the sum of amplitudes.

Thus, diffraction emerges not as a process basically different from interference, but as an interference situation involving an indefinitely large number of sources (every point on a wave front) instead of the cases involving small numbers of sources that were considered in Chapter 11. The main distinction that does arise in the analysis of diffraction is that because the number of sources is increased the mathematics involved can be rather complicated. We can, however, exhibit the essential features of diffraction phenomena without a full mathematical analysis.

In Section 12-2, we discuss the diffraction patterns observed along a plane beyond, but not greatly distant from, obstacles placed perpendicular to a plane wave front. Diffraction patterns created when the source or observing screen or both are near an obstacle are a type known as Fresnel diffraction.

If the source of light and the pattern observed are both at great distances from the obstacle, there are significant differences in the features observed compared to Fresnel diffraction. This second type is called Fraunhofer diffraction, and is discussed in Section 12-3.

In Section 12-4, we consider the Fraunhofer diffraction pattern due to a double slit. In the final section, Fraunhofer diffraction by an array of a great number of slits very close together is given. This latter situation describes the diffraction grating, a device of great value for the analysis of a beam of light containing more than one wavelength.

12-2 FRESNEL DIFFRACTION

Figure 12-1 illustrates the connection between the amplitude at a point P on a new plane wave front BB' due to contributions coming from all points on an original plane wave front AA'. When the point P is a distance R from the point O, the wave front AA' is divided into concentric circular regions that are called Fresnel zones. These zones have radii $r_1, r_2, r_3, \ldots, r_n$, where the subscript n specifies the respective zones. The radii are related to the distance R by the fact that the outer boundary of each zone is one-half a wavelength ($\lambda/2$) greater than the distance from P to the inner boundary of the same zone. From the geometry of the figure, we can write for the nth radius

$$r_n^2 = \left(R + \frac{n\lambda}{2}\right)^2 - R^2 = n\lambda R + \left(\frac{n\lambda}{2}\right)^2 \approx n\lambda R, \qquad (12\text{-}1)$$

since $\lambda/R \ll 1$. The area of the nth zone is, therefore, given by

$$A_n = \pi r_n^2 - \pi r_{n-1}^2 \approx \pi[n\lambda R - (n-1)\lambda R] \approx \pi\lambda R, \qquad (12\text{-}2)$$

so that the areas of all the zones are essentially the same. The significance of these relationships lies in the fact that by this construction we have obtained a series of adjacent zones containing corresponding points from which light of wavelength λ arriving at P will be 180° out of phase. That is, zones of even index will all be in phase, as will zones of odd index, but odd and even zones will differ in phase by 180° or $\lambda/2$.

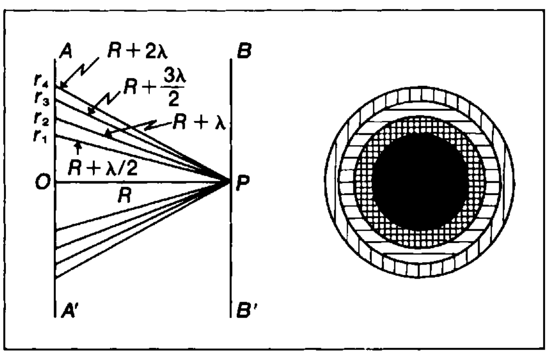


Figure 12-1 Fresnel zones for a plane wave front.

Now let the amplitude at point P due to the wavelets of zone 1 be represented by a_1, that due to the wavelets of zone 2—by a_2, etc. Because of the 180° phase difference between adjacent zones, it follows that the total amplitude at point P due to the entire wave front AA' is

$$a_{\text{total}} = a_1 - a_2 + a_3 - a_4 + \cdots \qquad (12\text{-}3)$$

which can be rearranged to the form

$$a_{\text{total}} = \frac{a_1}{2} + \left(\frac{a_1}{2} - a_2 + \frac{a_3}{2}\right) + \left(\frac{a_3}{2} - a_4 + \frac{a_5}{2}\right) + \cdots \qquad (12\text{-}4)$$

Because of the increasing obliqueness of wavelets arriving from successive zones,

$$a_1 > a_2 > a_3 > \cdots > a_n > \cdots.$$

Thus, the three terms in each parenthesis will contribute almost nothing. As a result, the total amplitude at point P from the wave front at AA' is no more than would be obtained if only one-half of the first zone were present! In Section 13-4, it is stated that the intensity of a wave is proportional to the square of the wave amplitude. Therefore, in the present situation, we must conclude that interference effects from adjacent zones reduce the total intensity at point P to a value one-fourth as great as that which would be observed from the central zone alone. Paradoxical as this may seem, it has been demonstrated, not only with visible light, but also with radar waves whose longer wavelengths permit the use of more easily constructed Fresnel zones.

When a screen with a circular opening (aperture) is placed perpendicular to a plane wave front, the pattern observed beyond the opening depends on the size of the aperture, the distance beyond the opening to the point of observation, and the location of the observation point on a plane parallel to the blocking screen. As an example, suppose the observation point lies on a line normal to the screen through the center of the aperture, so that the Fresnel zones are concentric with the aperture of radius r. If the observation point P is at a distance R such that $\pi r^2 = \pi R\lambda$, only the first zone will be exposed, giving an intensity that is (as we have already seen) four times that due to the entire unobstructed wave front. If the observation point is brought closer to the screen until $\pi r^2 = 2\pi\lambda R$, then two zones will pass through the aperture, giving darkness at P (since the amplitude there is $a_1 - a_2$). The same result could be obtained by keeping the observation point at the original distance R and increasing the radius of the aperture until the relation $\pi r^2 = 2\pi\lambda R$ is again satisfied. By continuing this argument, it becomes clear that when any odd number of zones are passed by the aperture a centrally located observation point P will be bright, while for an even number of zones passed the point will be dark.

Finally, consider the effect of shifting the observation point P along a line parallel to the blocking screen so that it is now no longer centrally located behind the aperture. The center of the Fresnel zones will also be shifted to lie along the same line normal to the aperture plane at P. As a result, the zones passing through the aperture are no longer symmetric with the aperture, and their contributions to the intensity at P depend strongly upon the location of P. Omitting further details, we state that the net result is a diffraction pattern made up of rings alternating in brightness and concentric with the center line normal to the aperture. Whether the center of the pattern is bright or dark depends upon the number of zones passed. Figure 12-2 shows the variation of intensity on a screen placed parallel to and behind a circular aperture of radius r with center at 0.

Now suppose that the screen with a circular aperture is replaced by a circular disk (or a spherical object). In this case, the pattern observed will be similar except that the central region of the pattern will always be bright. The reason for this is that the first of the zones not obscured by the disk will make a positive contribution to the intensity that is reduced to a total value one-fourth as great by the interference of the remaining zones, exactly as in the aperture case. This bright central region was predicted by Poisson, who deduced its existence using Fresnel's modifications of Huygens' principle. It is historically interesting that the possibility of such a situation seemed so absurd to Poisson that he regarded it as proof of the inapplicability of the wave theory of light. The subsequent demonstration of its reality by Arago (the bright spot is called the Arago spot by some, and by others—the Poisson spot!) and Fresnel essentially ended the controversy that had shrouded the wave theory first proposed by Young.

As a final example of Fresnel diffraction, consider a plane wave front incident upon a straight edge that blocks part of the wave front. If the observation point P is directly behind the edge of the obstacle, as shown in Figure 12-3, it is clear from our previous cases that only the upper half of all the Fresnel zones will contribute to the amplitude at P. Since the total amplitude for the entire wave front was found to be $a_{1/2}$, we see that at this point it becomes $a_{1/2}$. At points behind the obstacle (such as P_1), more than the lower half of the zones will be blocked, so that the total amplitude will steadily decrease to zero with a shift of the observation point more deeply into the shadow of the obstacle. Conversely, shifting out from the shadow toward points

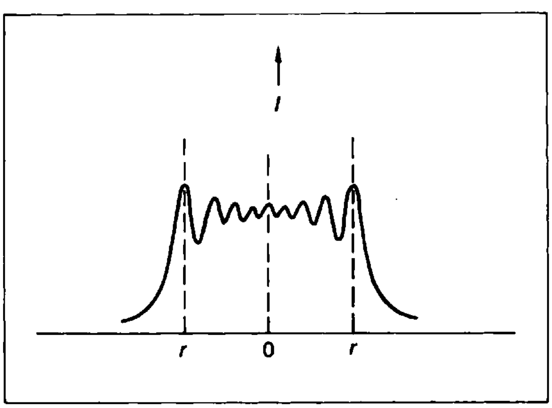


Figure 12-2 Fresnel diffraction pattern for a circular aperture.

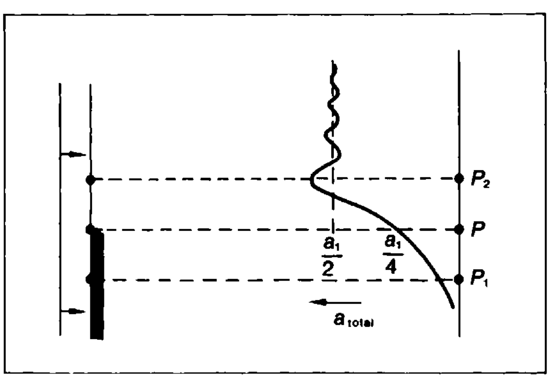


Figure 12-3 Fresnel diffraction-amplitude variation produced by a straight edge obstacle.

such as P_2 will expose additional portions of the lower half of the Fresnel zones, resulting in a total amplitude that is alternately greater or less than the unobstructed plane wave value of $a_1/2$. As the figure indicates, when the observation point becomes increasingly distant from the edge of the obstacle, the difference in amplitude at successive locations becomes smaller and smaller; the total amplitude for points sufficiently distant assumes the value of $a_1/2$, which is to be expected when the obstacle is so far distant that only outlying zones of negligible importance are affected by it.

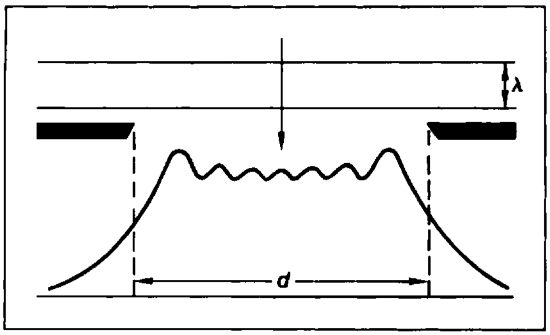


Figure 12-4 Fresnel diffraction at a single long rectangular slit.

It is not hard to understand that the Fresnel diffraction pattern due to a single slit should have the features illustrated in Figure 12-4, since it is the result of bringing together the patterns due to two long parallel straight edges that were originally widely separated. Close to a slit of width d, which is large compared to the wavelength λ, we see that except for positions very near the edges of the slit the amplitude is essentially constant. In addition, there is nearly total darkness within the geometric shadow, as would be expected from ray optics. For smaller slit widths, however, fewer zones remain unobstructed, and the diffraction features become more pronounced. When the viewing screen is moved to a greater distance behind the slit, an increase in the number of zones that are obstructed results. The pattern that is then observed will show marked intensity variations even behind the center of the slit in analogy to the case of the circular aperture. For example, distances behind the slit can be found for which the central position on the screen is totally dark or a maximum brightness.

12-3 FRAUNHOFER DIFFRACTION FROM A SINGLE SLIT

Let us continue our discussion of the diffraction pattern due to a long narrow slit. In particular, what features will the pattern exhibit when the secondary wavelets (which originate from a plane wave incident normally on the slit) leave the slit parallel to each other? This is an example of Fraunhofer diffraction, a simpler mathematical case, even though there is no change in the physical situation. The experimental difficulty of determining light intensities at great distances from the obstacle can be avoided by placing a converging lens behind the obstacle and observing the light intensity at the focal plane of the lens. Similarly, the source need not be far removed from the obstacle if it is placed at the first focal point of a converging lens, so that the wave fronts striking the obstacle become plane wave fronts. Figure 12-5 shows a plane wave incident upon a single slit of width d. Diffracted waves leave the slit at an angle θ relative to the direction of the incident wave. The convergent lens indicated is used to bring the parallel wavelets leaving the slit to a focus at point P in the focal plane of the lens. The path difference for the wavelets traveling to P from the upper edge of the slit compared to those leaving the lower edge is equal to $d \sin \theta$, from the figure. Now, if

$$d \sin \theta = n\lambda, \quad n = 1, 2, 3, \ldots, \tag{12-5}$$

then destructive interference will take place, and point P will be dark. To see this, consider the case $n = 1$, so that the upper wavelets must travel a distance equal to one wavelength further than those at the bottom of the slit. When this is true, a

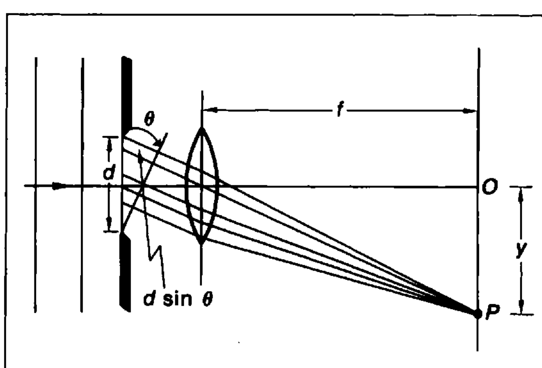


Figure 12-5 Fraunhofer diffraction for a single slit.

wavelet leaving the midpoint of the slit is one-half a wavelength out of phase with one leaving the bottom of the slit when they arrive at P. As a result, they interfere destructively with each other. Proceeding from the lowest point of the slit, all wavelets leaving the slit can be paired (one from each half of the slit) so that their path difference remains $\lambda/2$, so that complete destructive interference exists at P. For $n = 2$, the slit must be divided into four portions for each pair of which a path difference of $\lambda/2$ exists. This creates a system for which two pairs of points each give destructive interference. Since all points along the slit can be accounted for, total destructive interference occurs. The proof of the validity of the assertion for subsequent integers is left to the reader.

At point O, the angle θ is zero, and there is no path difference. As a result, all wavelets interfere constructively, giving a bright section in this central position in the pattern. Between each dark portion of the pattern there will be a bright section, whose intensity decreases with increasing distance from O on the screen. Although the proof is too advanced to be a part of this discussion, it is noted here that the variation in amplitude of the superimposed wavelets at a point P is given by

$$A_P = A_0 \left[\frac{\sin\left(\frac{\pi d \sin\theta}{\lambda}\right)}{\frac{\pi d \sin\theta}{\lambda}} \right], \qquad (12\text{-}6)$$

where A_0 is the amplitude observed at 0 in Figure 12-6. In addition, the intensity at P is given by

$$I = I_0 \left[\frac{\sin\left(\frac{\pi d \sin\theta}{\lambda}\right)}{\frac{\pi d \sin\theta}{\lambda}} \right]^2. \qquad (12\text{-}7)$$

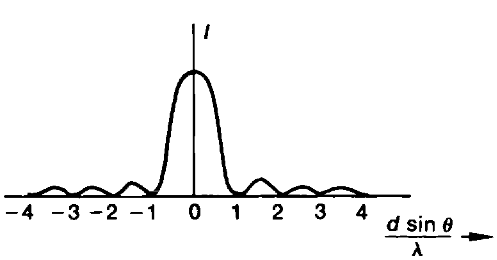


Figure 12-6 Intensity variation with diffraction angle for a single slit.

Figure 12-6 is a sketch illustrating the variation of intensity predicted by Eq. (12-7).

Example 1. Light of wavelength $\lambda = 5 \times 10^{-7}$ m is incident as plane waves on a slit of width $d = \frac{1}{5}$ mm, as shown in Figure 12-5.

(a) Find the angular width of the central maximum in the pattern.

(b) Calculate I/I_0 for a diffraction angle of $(1/\pi)°$.

SOLUTION

(a) From Eq. (12-5),

$$\sin\theta = \frac{\lambda}{d} = \frac{5 \times 10^{-7}\ \text{m}}{\frac{1}{5} \times 10^{-3}\ \text{m}} = 25 \times 10^{-4}.$$

Thus, $\theta = 0.143°$, and the angular width of the central maximum is

$$2\theta = 0.286°.$$

(b) From Eq. (12-7),

$$\frac{I}{I_0} = \left[\frac{\sin\left[\frac{\pi \times \frac{1}{5} \times 10^{-3}\ \text{m}}{5 \times 10^{-7}\ \text{m}} \times \sin\left(\frac{1}{\pi}\right)^\circ\right]}{\frac{\pi \times 10^4}{25} \sin\left(\frac{1}{\pi}\right)^\circ} \right]^2.$$

Since $\sin\alpha \approx \alpha$ (in radians) for α less than 10°, we can write

$$\sin\left(\frac{1}{\pi}\right)^\circ \approx \left(\frac{1}{\pi}\right)^\circ \times \frac{\pi\ \text{rad}}{180°} = \frac{1}{180}\ \text{rad}.$$

Therefore,

$$\frac{I}{I_0} \approx \left[\frac{\sin\left(\frac{400\pi}{180}\right)}{\frac{400\pi}{180}} \right]^2 = \left[\frac{\sin\left[\frac{20\pi}{9}(\text{rad}) \times \frac{180°}{\pi}\ \text{rad}\right]}{\frac{20\pi}{9}} \right]^2$$

$$\approx \left[\frac{9 \sin 400°}{20\pi}\right]^2 = \left(\frac{81}{400\pi^2}\right) \times \left(\frac{3}{2}\right) = \frac{243}{8\pi^2} \times 10^{-2}$$

$$\approx 3.1 \times 10^{-2} \text{ or about } 3\%.$$

12-4 FRAUNHOFER DIFFRACTION FROM A DOUBLE SLIT

It is logical to consider next the diffraction produced by two long parallel slits of equal width d, separated by a distance a, as illustrated in Figure

12-7. Each slit creates a diffraction pattern that exhibits the features discussed in the last section. At a given observation point, in the present case, there will be a superposition of these separate diffraction patterns. As a result, the intensity observed will now be due to the interference of the two diffraction patterns.

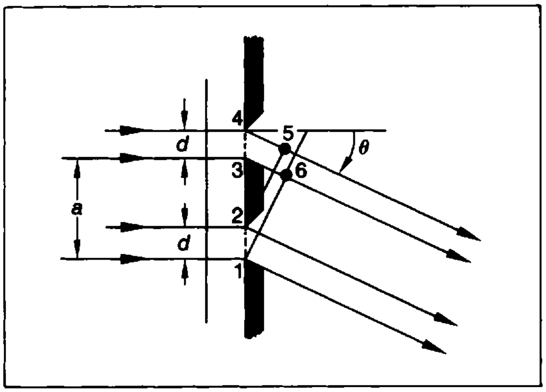


Figure 12-7 Plane waves diffracted by a double slit.

For a given angle θ, the amplitude due to either slit is given by Eq. (12-6). There will be a phase difference between the two, however, which corresponds to the path difference from point 3 to point 6 or, equivalently, from point 4 to point 5 in Figure 12-7. It therefore follows that the observation point will be dark if this path difference is equal to an odd integer number of half wavelengths:

$$\text{path difference}_{3\text{-}6} = (2n+1)\frac{\lambda}{2}, \quad n = 0, 1, 2, \ldots. \tag{12-8}$$

From Figure 12-7,

$$\text{path difference}_{3\text{-}6} = a \sin\theta,$$

and our destructive interference condition becomes

$$a \sin\theta = (2n+1)\frac{\lambda}{2} \quad \text{(dark fringe condition)}. \tag{12-9}$$

On the other hand, when

$$a \sin\theta = 2n\frac{\lambda}{2}, \quad n = 0, 1, 2, \ldots, \tag{12-10}$$

there will be constructive interference and bright fringes result.

The resulting amplitude at the observation point is given by the expression (which is presented without proof):

$$A = 2\underbrace{\left[A_0 \frac{\sin\left(\frac{\pi d \sin\theta}{\lambda}\right)}{\frac{\pi d \sin\theta}{\lambda}}\right]}_{\text{amplitude from a single slit}} \cos\left(\frac{\pi a \sin\theta}{\lambda}\right), \tag{12-11}$$

and the intensity at the same point is

$$I = 4I_0\left(\frac{\sin\left(\frac{\pi d \sin\theta}{\lambda}\right)}{\frac{\pi d \sin\theta}{\lambda}}\right)^2 \cos^2\left(\frac{\pi a \sin\theta}{\lambda}\right), \tag{12-12}$$

where I_0 is the intensity at $\theta = 0$ for a single slit as given by Eq. (12-7). We see that the effect of interference between the two single slit diffraction patterns is one of modulation. In other words, the rapidly varying terms $\cos^2(\pi a \sin\theta/\lambda)$ produces variations in intensity due to interference, which are superimposed on the intensity variations due to the two single slit diffraction patterns. The resulting intensity distribution is shown in Figure 12-8.

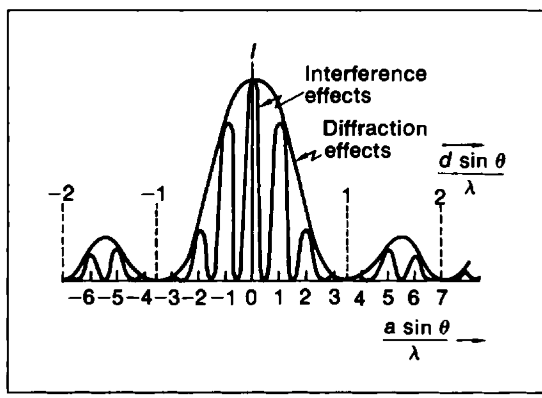


Figure 12-8 Intensity distribution for double slit diffraction.

Example 2. In a double slit diffraction pattern, the third principal maximum is missing because that interference maximum coincides with the first diffraction minimum (zero). Find the ratio a/d.

SOLUTION

Since the third interference maximum corresponds to $n = 2$ in Eq. (12-10),

$$a \sin\theta = 2\lambda,$$

the first diffraction zero corresponds to $n = 1$ in Eq. (12-5),

$$d \sin \theta = \lambda.$$

Dividing the two relations, we obtain

$$\frac{a}{d} = 2.$$

It is left as a problem to use this result and Eq. (12-12) to sketch the resulting intensity distribution.

12-5 THE DIFFRACTION GRATING

What sort of diffraction pattern should we expect from a diffraction grating, which consists of a very large number of extremely closely spaced long parallel slits of equal width? To answer the question analytically, we could proceed along the lines suggested in Sections 12-3 and 12-4. However, to avoid the mathematical complexities, we present instead Figure 12-9, which illustrates the intensity distribution and the changes that take place as the number of slits is increased. (In the figure, the ratio of slit width to slit spacing is kept constant.) We can make several observations about the patterns as the number of slits is increased:

1. subsidiary maxima appear, with intensities that are much weaker than those of the principal maxima;
2. the intensities of the subsidiary maxima become progressively weaker; and
3. the widths of the principal maxima become progressively narrower.

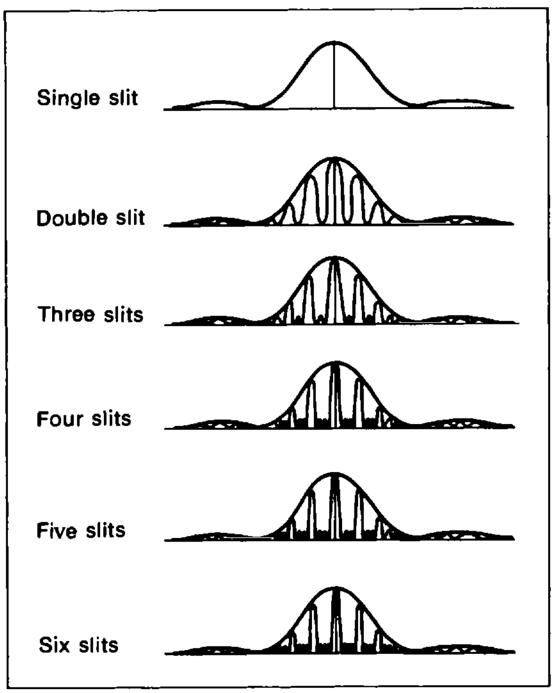


Figure 12-9 Effect of multiple slits on diffraction patterns—constant slit width to separation ratio.

If the number of slits is made very large (ruling engines have been developed that routinely produce several thousand lines or slit spacings per centimeter of ruled surface), it is observed that the pattern reduces to a series of extremely sharp lines, with no surviving subsidiary maxima. In addition, if the light incident on the slit system contains many wavelengths, there is a separation of the various wavelengths into a well-resolved series of spectral lines. This pattern is repeated at larger angles. In the pattern repetitions (called diffraction orders) overlapping of various colors from different orders occurs. The degree of overlapping increases in the higher orders.

We can obtain a quantitative expression of these latter facts if we refer to Figure 12-10, which shows a plane wavefront of wavelength λ incident normally on a series of narrow parallel slits separated by a distance d. It is assumed that $\lambda \ll d$. As in the earlier discussion of Fraunhofer diffraction, we are interested in the phase relationship between wavelets arriving from the various slits at the focal plane of a converging lens (as in Figure 12-5). From Figure 12-1, we see that constructive interference between wavelets from all of the slits will occur at the screen whenever the angle θ is such that the path lengths traveled by the wavelets from adjacent slits differ by an integer number of wavelengths.

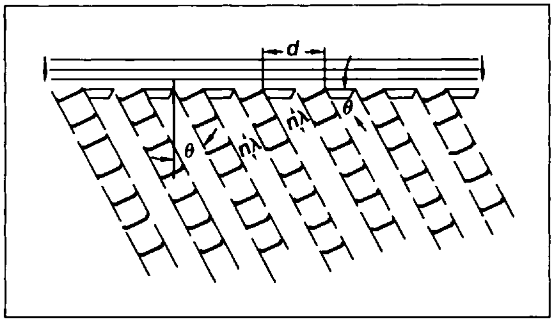


Figure 12-10 A diffraction grating.

When this happens, the following relation holds:

$$\sin\theta = n\frac{\lambda}{d}. \qquad (12\text{-}13)$$

The integer n specifies the order of the spectrum that is produced by the grating when a non-monochromatic beam is incident upon it.

Thus, if the grating spacing is known, if experimental provision is made for determining θ (or $\sin\theta$), and if the order n is observed, one can identify the wavelengths that are characteristic of any given light source. Examples of such applications are the spectra produced by flame excitation of various inorganic salts or electric arc excitation of gases such as hydrogen or neon. In the field of chemistry, information of this kind can be used as a basis for either qualitative or quantitative analysis of unknown samples. In physics, the existence of intriguing numerical relationships between the various wavelengths of the excitation spectrum for a given substance was known in the 1880s and led to the development of modern quantum theory.

The equation indicates that for a given order a grating will produce a spectrum that has longer wavelengths (the red end of the spectrum) appearing at larger angles than the shorter wavelengths. The equation can also be used to justify the overlapping of colors from successive orders that was asserted earlier.

Example 3. Using Eq. (12-13), show that, regardless of the grating spacing, the third order violet line will overlap the second order red line for light incident normally on the grating.

$$(\lambda_{\text{violet}} \approx 4\times10^{-7}\text{ m}, \quad \lambda_{\text{red}} \approx 7\times10^{-7}\text{ m})$$

SOLUTION

From Eq. (12-13),

$$\sin\theta_{\text{violet}} = 3\lambda_{\text{violet}}/d,$$

$$\sin\theta_{\text{red}} = 2\lambda_{\text{red}}/d.$$

Therefore,

$$\frac{\sin\theta_{\text{violet}}}{\sin\theta_{\text{red}}} = \frac{3\lambda_{\text{violet}}}{2\lambda_{\text{red}}} \approx \frac{6}{7}.$$

This shows that $\theta_{\text{violet}} < \theta_{\text{red}}$, so that the second order red line overlaps the third order violet line, as asserted.

Example 4. A diffraction grating has a spacing of 5000 lines/cm. Find the angular spread of the second order spectrum, assuming normal incidence.

SOLUTION

In this case,

$$d = 1/5000 = 2\times10^{-4}\text{ cm} = 2\times10^{-6}\text{ m}.$$

Therefore,

$$\sin\theta_{\text{violet}} \approx (2)(4\times10^{-7}\text{ m})/(2\times10^{-6}\text{ m}) = 0.4$$

$$\sin\theta_{\text{red}} \approx (2)(7\times10^{-7}\text{ m})/(2\times10^{-6}\text{ m}) = 0.7.$$

As a result,

$$\theta_{\text{violet}} \approx 24°$$

and

$$\theta_{\text{red}} \approx 45°,$$

so that the angular spread of the second order spectrum is

$$\theta_{\text{red}} - \theta_{\text{violet}} \approx 21°.$$

PROBLEMS

1. Monochromatic plane waves of wavelength λ are incident normally on a screen with a circular aperture of diameter d. Along an axis through the center of the aperture and normal to the screen, the light intensity will pass alternately through maxima and minima. If the distance from the screen along this axis is R, show that the location of these maxima and minima are given by the relation

$$R = \frac{d^2}{4n\lambda},$$

where n is any integer. Show that maxima correspond to n even, minima—to n odd.

2. (a) Referring to the results of Problem 1, what is the largest value of R for which the light intensity is a minimum?
   (b) What is the largest value of R for which the light intensity is a maximum?

3. A Fresnel circular zone plate is a circular array of alternately transparent and opaque rings constructed so that each has approximately the same area. (See Section 12-2.) If the even numbered rings are opaque, the light transmitted by this zone plate will be much more intense than if all the light were transmitted. Explain why this is so.
4. Plane waves of sodium light ($\lambda = 5.893 \times 10^{-7}$ m) are incident normally on a circular aperture 4.60 mm in diameter, and observed at a point on the axis 3 m from the aperture.
   (a) Is the center of the diffraction pattern bright or dark?
   (b) Determine the minimum distance the point of observation should be moved along the axis to reverse the situation of (a).
5. (a) Calculate the radius of a circular zone plate with a "focal length" of $\frac{11}{3}$ m for light of wave length $\lambda = 5.46 \times 10^{-7}$ m. The plate has only 8 zones; zones, 1, 3, 5, and 7 are transparent, and the even numbered zones are opaque.
   (b) Repeat for infrared light, $\lambda = 5.46 \times 10^{-6}$ m.
6. Monochromatic plane waves of wavelength λ are incident normally on a rectangular slit of width d. A viewing screen is placed behind the slit as in Figure 12-6. Sketch the expected diffraction patterns when the screen is
   (a) near the slit,
   (b) somewhat farther from the slit, and
   (c) at a great distance from the slit.
   Note: it is assumed that $\lambda \ll d$.
7. Equation (12-7) can be written in the form

$$\frac{I}{I_0} = \frac{\sin^2 \beta}{\beta^2},$$

where

$$\beta = \frac{\pi d \sin \theta}{\lambda}.$$

   (a) Using this form, show that the minima of the Fraunhofer single slit diffraction pattern occur at values of θ such that $\beta = n\pi$, where $n = 1, 2, 3, \ldots$. (Why is the case $n = 0$ not included?)
   (b) Show that the maxima of the Fraunhofer single slit diffraction pattern occur approximately at values of θ such that $\beta = (2n - 1)\pi/2$, where $n = 1, 2, 3, \ldots$, thus for odd integer multiples of $\pi/2$. Although this is an approximation, it becomes increasingly accurate for large n.
   (c) The exact values of β for which I/I_0 has successive maxima can be shown (by calculus methods) to satisfy the relation

$$\tan \beta = \beta^{-}.$$

   Using trigonometry tables, determine the errors involved in using the approximate values of (b) for the first 4 secondary maxima.
8. Assume the focal length f of the converging lens placed just behind the slit in Figure 12-6 is large compared to the slit width d. Show that the locations of the maxima and minima on the viewing screen placed in the focal plane of the lens are given by

$$y = f \tan \theta.$$

9. (a) Calculate the angles at which the first minimum would occur for the Fraunhofer diffraction pattern of a slit of width 2×10^{-6} m on which plane waves of wavelengths $\lambda_1 = 4 \times 10^{-7}$ m and $\lambda_2 = 7 \times 10^{-7}$ m are incident normally.
   (b) If the focal length f of the converging lens behind the slit is $\frac{1}{2}$ m, how far apart will the two minima be when viewed on screen in the focal plane of the lens?
10. Repeat Problem 9 for a slit of width 2×10^{-4} m.

11. When a single rectangular slit of width d is illuminated by two line sources of wavelength λ, there will be a diffraction pattern for each line source on the viewing screen. Depending upon the angular separation of the line sources, the two patterns will overlap to an extent that can make it difficult to say whether the total pattern is due to one or two sources. The two sources are said to be just resolvable when the first minimum of the one pattern coincides on the screen with the first maximum of the other. Draw a sketch illustrating this situation, and show that this limiting angle of resolution is given by

$$\theta_c = \frac{\lambda}{d}.$$

(It is assumed here that $\lambda/d \ll 1$.)

12. When two point sources of wavelength λ illuminate a circular aperture of *diameter d*, a more involved analysis leads to the resolution condition (known as Rayleigh's criterion)

$$\theta_c \approx 1.22\lambda/d, \quad \text{for } \lambda/d \ll 1.$$

Two point sources of wavelength $\lambda = 5 \times 10^{-7}$ m are 25 cm in front of a circular aperture of diameter $d = 4 \times 10^{-3}$ m. What is their minimum separation if they are to be resolvable?

13. In the double slit diffraction example, it was found that $a/d = 2$. Use result and Eq. (12-12) to construct a plot of

$$\frac{I}{I_0} \text{ versus } \frac{a \sin \theta}{\lambda}.$$

14. Monochromatic light of wavelength $\lambda = 5.46 \times 10^{-7}$ m is incident normally on a double slit. The Fraunhofer diffraction pattern is viewed in the focal plane of a lens of focal length $f = 2$ m. The sixth principal maximum is missing, and the distance between the two minima next to the central maximum is observed to be 1 mm. Find d, the width of the slits, and a, the distance between them.

15. When $a/d = 1$, the two slits are no longer separated (see Figure 12-8). Instead the system reduces to a single slit of width $2a$. Show that Eq. (12-12) reduces to Eq. (12-7) with d replaced by $2a$ and 4 times the intensity I_0 of a single slit.

16. If $a/d = 4$ for a double slit system, which of the interference maxima will be missing?

17. A diffraction grating for which the slit spacing is unknown can still be used for quantitative work. For example, consider a plane wave containing a known wavelength λ_s and an unknown wavelength λ_x incident normally on a diffraction grating. It is observed that the maximum of order n_s for the known wavelength occurs at the same angular position as the maximum of order n_x of the unknown wavelength.

(a) Show that

$$\lambda_x = \left(\frac{n_s}{n_x}\right)\lambda_s.$$

(b) Show also that

$$\lambda_x = \left(\frac{\sin \theta_x}{\sin \theta_s}\right)\lambda_s$$

when the maxima are in the same order.

(c) One of the visible wavelengths associated with the spectrum from atomic hydrogen is known to be $\lambda_s = 6.560 \times 10^{-7}$ m. It is found that the third order maximum of this wavelength occurs at the same position as the fourth order maximum of an unknown wavelength λ_x. Find λ_x.

(d) At what angle will the second order maximum of the unknown wavelength occur if the second order maximum of λ_s occurs at $\theta_s = 30°$?

18. A certain diffraction grating shows a maximum angular separation of visible light (4000 Å to 7000 Å) in the first order.

(a) Calculate the slit separation for the grating.

(b) Calculate the angular spread of visible light for the grating.

(c) If spacing of slits is increased, what changes will occur in the angular separation of visible light and in the number of orders that can be observed?

(d) Repeat (c) for a decrease in slit spacing.

19. In obtaining Eq. (12-13), normal incidence of monochromatic plane waves of wavelength λ was assumed. If, instead, the light is incident at an angle i and the nth order maximum is observed at an angle θ, show that Eq. (12-13) becomes $n\lambda = d(\sin i + \sin\theta)$. (A diagram analogous to Figure 12-11 will be very helpful.) Note that when $i = 0°$, this result reduces as it should to Eq. (12-13).

20. (a) Consider two wavelengths which differ only slightly, λ and $\lambda + \Delta\lambda$, where $\Delta\lambda/\lambda \ll 1$. Show that the angular separation $\Delta\theta$ of these two lines in the nth order is given approximately by

$$\Delta\theta \approx \frac{n\,\Delta\lambda}{d\cos\theta}.$$

Hint: Use the trigonometric identity

$$\sin(A + B) = \sin A\cos B + \cos A\sin B,$$

and note that

$$\cos\Delta\theta \approx 1, \quad \sin\Delta\theta \approx \Delta\theta,$$

when $\Delta\theta$ is small (less than 5°), and is expressed in radians.

(b) If the nth order maximum for wavelength λ is observed at an angle θ, show that

$$\Delta\theta \approx \frac{\Delta\lambda}{\lambda}\tan\theta.$$

(c) The spectrum of sodium has two closely spaced lines, 5890 Å and 5896 Å. If the angular separation of these lines in the second order is 7.42×10^{-4} rad = 0.428°, determine the slit spacing of the grating.

13 Transverse Waves and Polarized Light

13-1 INTRODUCTION

In the last two chapters, it was demonstrated that interference and diffraction phenomena involving light can be understood by assuming light has a wave nature. In Section 10-1, it was stated that two basic wave types exist: longitudinal waves—in which the vibrations take place in a direction parallel to the propagation direction of the wave; and transverse waves—in which the vibrations take place in a plane perpendicular to the propagation direction of the wave. Any given wave motion could therefore be either purely longitudinal, purely transverse, or some combination of the two types. The principle concern of this chapter is the appropriate type of description for light waves and the physical effects associated with it.

As the chapter heading implies, light waves are in fact transverse. We therefore begin with a discussion of transverse waves. For this purpose, we find it useful to introduce the concept of vector quantities, and to provide a brief introduction to the algebra of vectors. (Further details are provided in the Appendix.) We then discuss polarization, and show how polarization effects can be used to demonstrate the transverse nature of light waves. In this way, we can avoid the mathematical complications involved in proving the principle of superposition for light waves. Finally, we consider a number of examples of the interaction of light and various materials that involve polarization.

13-2 TRANSVERSE WAVES

The vibrations associated with a transverse wave propagating along the z-axis must lie in planes parallel to the x-y plane. The simplest example of such a wave is provided by a transverse wave on a string, as described in Chapter 10. For wavelength λ, propagation speed v, and vibrational amplitude A_y, and assuming vibrations occur in the y-direction only, the vibration equation becomes

$$y = A_y \sin\left[\frac{2\pi}{\lambda}(vt - z) + \phi\right], \qquad (13\text{-}1)$$

where ϕ is a constant phase angle which depends upon initial conditions. (For example, if $y = \frac{1}{2}A_y$ when $z = 0$, $t = 0$, then $\phi = \pi/6$.)

If the same wave is related to vibrations in the x-direction only, then Eq. (13-1) becomes

$$x = A_x \sin\left[\frac{2\pi}{\lambda}(vt - z) + \phi\right]. \qquad (13\text{-}2)$$

In either case, when one looks along the string (down the z-axis) toward the source producing the vibrations, one sees a linear disturbance along either the y- or x-axis, respectively. This can be accomplished, for example, by setting one end of a long string in motion sinusoidally along the appropriate axis.

Now suppose that sinusoidal vibrations along the x- and y-axes are simultaneously imposed on the string. What is the configuration of the disturbance that would now be observed by looking along the string toward the source of the disturbance? As in Chapter 10, we can apply the principle of superposition provided we account for the fact that these vibrational displacements are perpendicular. It will not be possible for these disturbances to combine to produce a completely destructive or constructive interference. We present two techniques for determining the configuration of the disturbance—the first, a mathematical manipulation, the second, an equivalent graphical construction.

To simplify the mathematics of our illustration, we introduce the substitution

$$\theta_z = \frac{2\pi}{\lambda}(vt - z).$$

In addition, we devote our attention to the point $z = 0$ so that Eqs. (13-1) and (13-2) can be written

$$\begin{aligned} y &= A_y \sin(\theta_0 + \phi_y) \\ &= A_y \cos\phi_y \sin\theta_0 + A_y \sin\phi_y \cos\theta_0 \end{aligned} \qquad (13\text{-}3)$$

$$\begin{aligned} x &= A_x \sin(\theta_0 + \phi_x) \\ &= A_x \cos\phi_x \sin\theta_0 + A_x \sin\phi_x \cos\theta_0, \end{aligned} \qquad (13\text{-}4)$$

where we have used the identity $\sin(\alpha \pm \beta) = \sin\alpha \cos\beta \pm \cos\alpha \sin\beta$. To determine the vibrational configuration independent of time, it is necessary to eliminate

$$\theta_0 \left(= \frac{2\pi}{\lambda}(vt) = \frac{2\pi}{\lambda}(f\lambda)t = \omega t \right)$$

from the equations.† We do this by the following sequence of operations:

1. Multiply Eq. (13-3) by $A_x \cos\phi_x$, Eq. (13-4) by $A_y \cos\phi_y$, and subtract the resulting equations.
2. Multiply Eq. (13-3) by $A_x \sin\phi_x$, Eq. (13-4) by $A_y \sin\phi_y$, and subtract the resulting equations.
3. Square and add the two equations obtained in steps 1 and 2.
4. Simplify the result by using the identities $\sin^2\alpha + \cos^2\alpha = 1$ and $\cos(\alpha \pm \beta) = \cos\alpha \cos\beta \pm \sin\alpha \sin\beta$.

When these steps have been performed (see Problem 13-2), we obtain the equation

$$A_x^2 y^2 + A_y^2 x^2 - 2(A_x A_y \cos\Delta\phi)xy = A_x^2 A_y^2 \sin^2\Delta\phi, \qquad (13\text{-}5)$$

where $\Delta\phi = (\phi_x - \phi_y)$ is the phase angle difference between the two transverse waves.

This equation may be recognized as the equation for a conic section. For special values of A_x, A_y, and $\Delta\phi$, it leads to several common situations. Thus, if either A_x or A_y is zero, the result is of course a linear disturbance as outlined above, with vibrations along either coordinate axis. On the other hand, if neither A_x nor A_y equals zero and $\Delta\phi = 0$, Eq. (13-5) becomes

$$(A_x y - A_y x)^2 = 0 \qquad (13\text{-}6)$$

or

$$y = \frac{A_y}{A_x} x. \qquad (13\text{-}7)$$

This is a straight line disturbance also, but one which is inclined to the x-axis by an angle ψ, where $\tan\psi = A_y/A_x$. If $A_x = A_y$, $\psi = \pi/4$, and the disturbance is composed of equal amounts along the two coordinate axes.

A circular disturbance results if $A_x = A_y$ and $\Delta\phi = \pm(2n+1)\pi/2$, where $n = 0, 1, 2, 3, \ldots$. Finally, if $A_x \neq A_y$ and $\Delta\phi \neq 0$, the disturbance will be elliptical in form. These results are summarized in Table 13.1.

Now let us consider the second method, a graphical technique, for performing the same superposition of two transverse waves with perpendicular vibrations but identical wavelength and propagation speed. Figure 13-1 is a simultaneous plot of y versus ωt in the upper right part of the diagram, of x versus ωt in the lower left part, and in the upper left the resulting vibrational configuration is obtained by plotting x, y values for identical values of ωt. In this example, $A_x \neq A_y$ and $\Delta\phi = \pi/2$. The result is in agreement with the predictions of Table 13-1. It might be noted that this configuration can be obtained on an oscilloscope if one applies the appropriate sinusoidal voltages on the vertical and horizontal inputs of the oscilloscope. Such figures are known as Lissajous figures, and can be used to determine unknown frequencies if a signal of known frequency is available. (See Problem 13-3.)

†Note that $\omega = 2\pi f$. ω is called the angular frequency. Since 2π is the number of radians per cycle, and f is the number of cycles per second, ω is the number of radians per second.

Table 13-1 Summary of Vibrational Configurations.

Conditions	Configuration of Disturbance
$A_x = 0,\ A_y \neq 0$	Linear, along y-axis
$A_x \neq 0,\ A_y = 0$	Linear, along x-axis
$\Delta\phi = 0,\ \pm 2n\pi$	Linear, at angle $\psi = \tan^{-1}\left(\frac{A_y}{A_x}\right)$ from x-axis
$\Delta\phi = \pm(2n-1)\pi$	Linear, at angle $\psi = \tan^{-1}\left(\frac{-A_y}{A_x}\right)$ from x-axis
$\Delta\phi = \pm(2n-1)\pi/2,\ A_x = A_y$	Circular
$\Delta\phi = \pm(2n-1)\pi/2,\ A_x \neq A_y$ $n = 1, 2, 3, \ldots$ (any integer)	Elliptical, principal axes coincide with x- and y-axes
$A_x \neq A_y$, $\Delta\phi$ any value not given above	Elliptical, principal axes inclined to the x- and y-axes

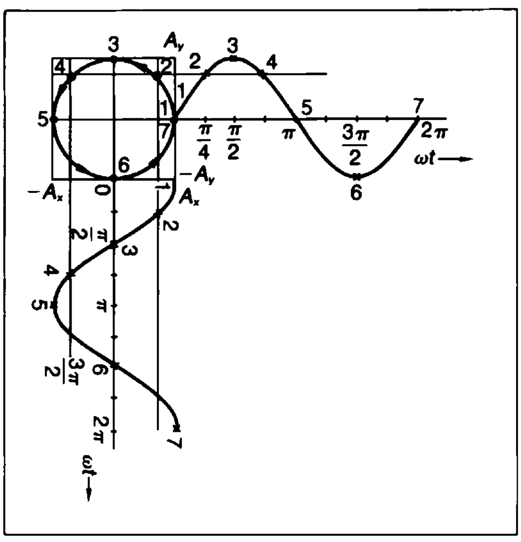


Figure 13-1 Graphical addition of perpendicular sinusoidal vibrations (with equal wavelengths and propagation speeds) transverse to the same propagation direction.

13-3 VECTORS AND VECTOR ALGEBRA

In the previous section, we saw that it is possible to "add" (superpose) disturbances and obtain results that are not consistent with simple algebra. That is, a positive disturbance and a negative disturbance may yield a variety of results that are neither simple sums nor differences. The reason, of course, is that the quantities involved possessed not only a magnitude but also a direction, and both factors must be accounted for in the superposition process. Quantities of this sort, which are of frequent occurrence in physics, are known as *vectors*. These are in contrast to *scalar* quantities (or simply *scalars*) that have magnitude only. For example, the volume of a solid is a scalar, while the displacement of a particle of a string is a vector. By displacement of a particle, we mean the shift in location of the particle (due to some source of disturbance applied to the string, as above).

A vector quantity is illustrated graphically by drawing an arrow from an origin to a point. The length of the arrow is the magnitude of the vector, and its direction specifies the direction of the vector. For example, Figure 13-2 shows the displacement of a given particle of a vibrating string due to a transverse wave propagating along the string in the z-direction at a given instant. For the situation shown, the particle is displaced from $y = 0, x = 0$ to the point $y = A_y$, $x = A_x$. From the Pythagorean theorem, the magnitude of the displacement is $A = (A_x^2 + A_y^2)^{1/2}$ and the direction is given by

$$\psi = \tan^{-1}\left(\frac{A_y}{A_x}\right).$$

Any given vector consists of parts (called components) representing the extent of the vector in three different (usually mutually perpendicular) direc-

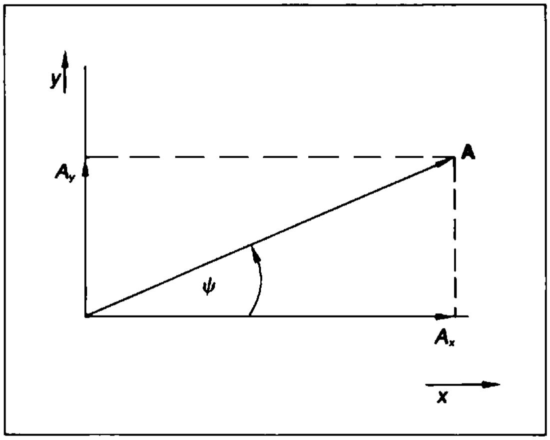


Figure 13-2 Graphical representation of the two-dimensional vector **A**.

tions. Thus, in Figure 13-2, **A** has no z-component while the x- and y-components are A_x and A_y, respectively. We can indicate that the vector **A** is obtained as a *vector* sum of the components if we introduce the concept of unit vectors. A unit vector is a vector of unit length. An arbitrary vector of the same physical kind and in the same direction as the unit vector is then simply a scalar number (in whatever system of physical units is required for the quantity) indicating the number of unit vectors that are equivalent in magnitude to the arbitrary vector. One convention is to denote unit vectors in the x-, y-, and z-directions by **i**, **j**, and **k**, respectively. In this notation, the vector **A** in Figure 13-2 becomes

$$\mathbf{A} = A_x\mathbf{i} + A_y\mathbf{j} + 0\mathbf{k}$$

or, simply,

$$\mathbf{A} = A_x\mathbf{i} + A_y\mathbf{j}. \tag{13-8}$$

How do we add two vectors in general? The procedure is straightforward if the two vectors are written in the form of Eq. (13-8) because the addition of any two x-components (and similarly for y- and z-components) follows regular algebraic rules. Physically, this must be true since we have seen that constructive or destructive interference can occur for superposition of two vibrational displacements in the *same* direction. Thus, if $\mathbf{C} = \mathbf{A} + \mathbf{B}$, we must have

$$C_x\mathbf{i} + C_y\mathbf{j} + C_z\mathbf{k} = (A_x\mathbf{i} + A_y\mathbf{j} + A_z\mathbf{k}) + (B_x\mathbf{i} + B_y\mathbf{j} + B_z\mathbf{k}) \tag{13-9}$$

or

$$C_x = A_x + B_x,$$
$$C_y = A_y + B_y,$$
$$C_z = A_z + B_z. \tag{13-10}$$

From the Pythagorean theorem, it again follows that the magnitude of **C** is

$$|\mathbf{C}| = (C_x^2 + C_y^2 + C_z^2)^{1/2}$$
$$= [(A_x + B_x)^2 + (A_y + B_y)^2 + (A_z + B_z)^2]^{1/2}$$

Example 1. Given vectors $\mathbf{A} = 5\mathbf{i} + 3\mathbf{j}$, $\mathbf{B} = 7\mathbf{i} - 2\mathbf{j}$. Find $\mathbf{C} = \mathbf{A} + \mathbf{B}$ and $\mathbf{D} = \mathbf{A} - \mathbf{B}$.

SOLUTION

From Eqs. (13-9), (13-10), and (13-11),

$$\mathbf{C} = 12\mathbf{i} + \mathbf{j}, \quad |\mathbf{C}| = \sqrt{145},$$
$$\mathbf{D} = -2\mathbf{i} + 5\mathbf{j}, \quad |\mathbf{D}| = \sqrt{29}.$$

The results are also shown graphically in Figures 13-3 and 13-4. Notice that the figures demonstrate that vector addition is commutative, that is,

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

and

$$\mathbf{A} + (-\mathbf{B}) = (-\mathbf{B}) + \mathbf{A} = \mathbf{D}.$$

A knowledge of the mathematics of vectors is essential for quantitative study in physics; the reader is therefore urged to study carefully the material presented in the Appendix. For the remainder

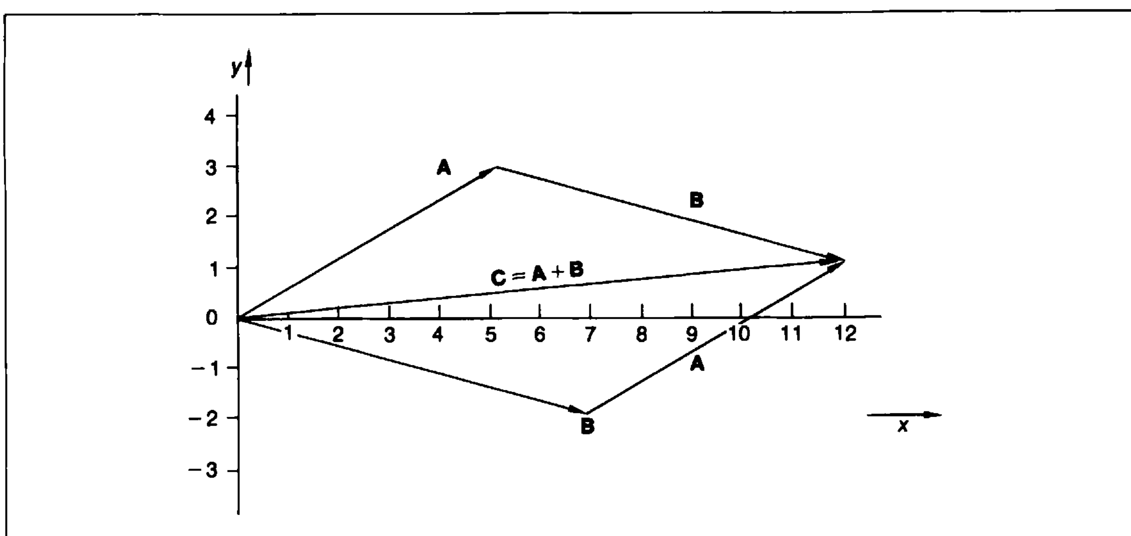


Figure 13-3 Graphical representation of C=A+B.

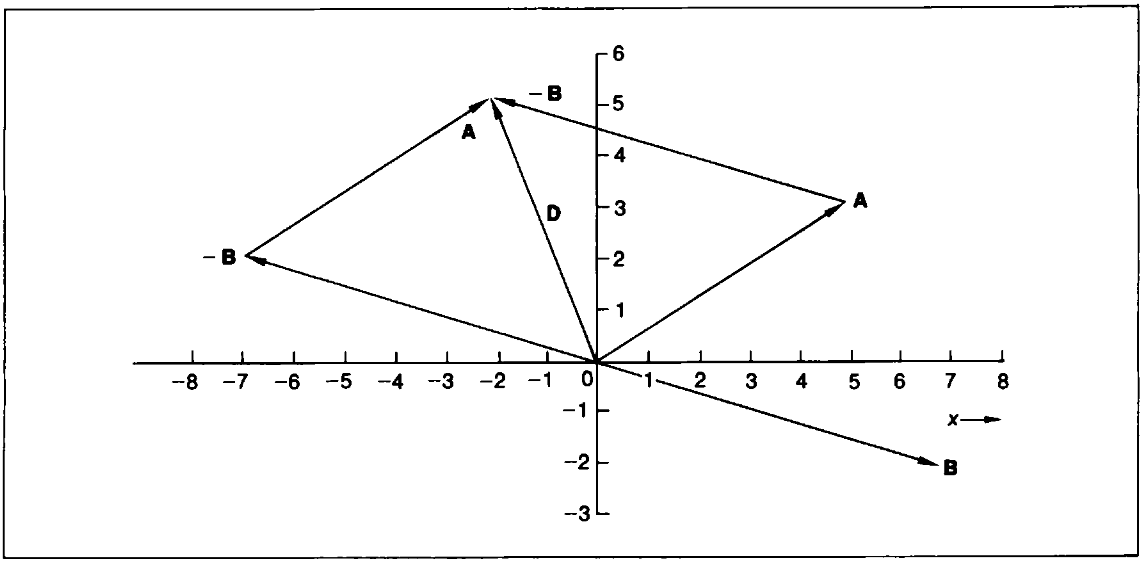


Figure 13-4 Graphical representation of **D**=**A**−**B**.

of this chapter, the information provided above is sufficient, and we turn now to the subject of polarization.

13-4 POLARIZATION AND POLARIZED LIGHT

The concept of polarization is used to identify the possible vibrational configurations that can be associated with transverse waves. From the discussions of the previous sections, we can consider a transverse wave in terms of a vibrational displacement vector that lies in a plane perpendicular to the direction of propagation and whose components along two perpendicular axes are sinusoidal functions of time and position along the wave, with some phase angle difference between the two components. The various configurations of Table 13-1 are thus identified as different states of polarization. For example, the first entry in the table represents a wave *linearly polarized* along the y-axis, while the last entry represents an elliptically polarized wave. From the analysis of Section 13-2, we see that transverse waves will have either elliptical, circular, or linear polarization. Furthermore, elliptical or circular polarization was shown to be the result of the vector addition of two linearly polarized waves.

A transverse wave for which the vibrational displacement vector points in a direction that varies completely randomly with time is called an unpolarized wave since there is no preferred vibrational direction. To produce a linearly polarized wave from an unpolarized wave, one needs a device known as a polarizer, which will allow only one polarization component to pass through it. For example, a transverse wave on a string can be polarized by passing the string through the slot between the boards of a picket fence. Any vibrational displacement component perpendicular to the slot will not pass through, while the parallel component will be passed. A second slot aligned perpendicular to the first will stop any of the vibrational displacement component transmitted by the first slot. If the second slot is instead aligned parallel to the first, there will be no reduction of the vibrational displacement. Since the second slot can thus be used to analyze the polarization of the vibrational displacements passed by the polarizer, the second slot is called an analyzer. For example, if a polarizer and an analyzer are used with their axes of transmission at right angles and a non-zero vibrational amplitude is observed, then the wave must have a longitudinal component of vibration. (Why?)

Let us now consider the case of light waves. From the discussion above, if optical devices analogous to the picket fence slots (polarizer and analyzer) for waves on a string can be devised, then the vibrational character of light waves can be determined. There are in fact several ways of accomplishing this. The simplest is to make use of materials which bear the general name polaroid (their inventor is Edwin Land). A polaroid material transmits nearly all the light with one linear polarization but almost none having a linear polarization at right angles to the first. Thus, the axis of polarization of the polaroid is analogous to the slots in the picket fence of the string example. A polarizer and an analyzer each made of polaroid with crossed axes (polarization axes at 90° to each other) will produce complete extinction for any light beam incident on the combination. This is consistent with our earlier assertion that light can be understood as a purely transverse wave phenomenon. Note, however, that using more than two pieces of polaroid at various angles gives results that have no analogue with the string-picket fence case. See Problems 13-10 through 13-12.

An unpolarized light beam passing through a polarizer will exhibit no variation in brightness as the polarization axis of the polarizer is rotated in a plane perpendicular to the propagation direction of the light. This is because an unpolarized beam has no preferred or prominent polarization features. The random nature of such a beam requires that the magnitude of the vibration component in one direction be no more (and no less) than that in any other direction (transverse to the propagation direction). On the other hand, a linearly polarized light wave is easily identified by this test since the brightness will decrease as the axis of the polarizer goes from an alignment parallel to the polarization axis of the wave to a perpendicular alignment.

Consider the following situation. The polarization axis of an analyzer makes an angle θ with the polarization axis of the polarizer. From our previous discussion, it follows that the amplitude of an initially unpolarized light wave that is transmitted by the polarizer can be described by a vector **A** aligned parallel to the polarizer axis. Therefore, since the analyzer axis makes an angle θ with the polarizer axis, the amplitude of the light wave transmitted by the analyzer will be reduced to a magnitude $A' = A \cos\theta$ and will be linearly polarized parallel to the analyzer axis, as shown in Figure 13-5.

Quantitative measurements of the reduction of transmitted light in a polarizer-analyzer system are made by a variety of devices (photocells, for example) that determine the *intensity* of light incident on the measuring device. The intensity of a light beam is given by the amount of energy that is incident on a given cross-sectional area of the detector per unit time. In MKS units, the intensity is measured in joules per meter squared per second. When such measurements are made, it is found that I (the intensity transmitted by the analyzer) is related to I_0 (the intensity transmitted by the polarizer) by the relation

$$I = I_0 \cos^2\theta, \tag{13-12}$$

where θ is the angle between the polarization axes of polarizer and analyzer. This relation is known as Malus' law, after its discoverer. We conclude that the intensity of a light wave is proportional to the square of the amplitude of the wave. That is,

$$\frac{I}{I_0} = \cos^2\theta = \left(\frac{A'}{A}\right)^2, \tag{13-13}$$

since $A' = A\cos\theta$ from above. It should be noted that for an unpolarized wave I_0 is only one-half the intensity of the wave incident on the polarizer,

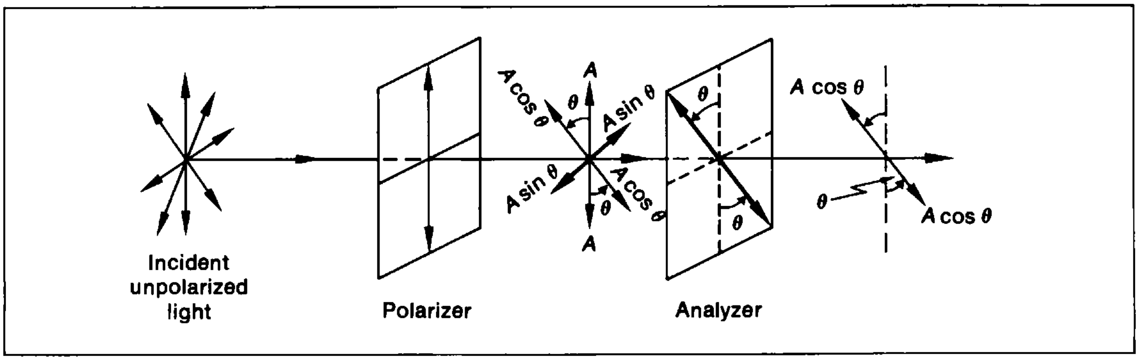


Figure 13-5 Polarizer-analyzer effects on a normally incident unpolarized light wave.

since the directions parallel and perpendicular to the polarizer axis should be equally effective in carrying the energy associated with the unpolarized wave.

13-5 MATTER AND POLARIZED LIGHT

In addition to the use of polaroid materials, other methods for producing polarized light can be used. One of the simplest techniques is based upon the experimental fact that when light is incident upon a plane surface the reflected part of the beam will be linearly polarized to an extent that varies with the angle of incidence. That is, for initially unpolarized light one can regard the vibrations as being a combination of transverse vibrations that are perpendicular to each other and to the direction of propagation. It is conventional to identify the plane defined by incident ray and the normal to the reflecting surface as the plane of incidence (as in Chapter 6). As Figure 13-6 indicates, the component of the polarization parallel to the plane of incidence is largely refracted rather than reflected. When the angle of incidence has been adjusted so that a 90° angle results between the reflected ray and the refracted ray, complete linear polarization perpendicular to the plane of incidence results. When this is true, the angle of reflection $\theta_i = 90 - \theta_r$, where θ_r is the angle of refraction. Using Snell's law and letting the index of the reflecting medium be n, we obtain for a beam incident in air ($n = 1$):

$$\sin\theta_i = n \sin\theta_r = n \sin(90 - \theta_i)$$

or

$$\sin\theta_i = n\cos\theta_i.$$

Therefore, when

$$\tan\theta_i = n, \tag{13-14}$$

total linearly polarized light in the reflected beam results. This expression is known as *Brewster's law*, and the angle satisfying Eq. (13-14) is called the *Brewster angle.*

Example 2. What angle of incidence is required to produce linearly polarized light by reflection using a flint glass plate ($n = 1.65$)?

SOLUTION

Using Eq. (13-14),

$$\tan\theta_i = 1.65,$$

so that

$$\theta_i \approx 58.8°.$$

That the reflected light for the Brewster angle is indeed linearly polarized can be demonstrated by causing it to impinge upon a second reflecting surface of the same material so arranged that the angle of incidence is the same but with the plane of incidence at right angles to that of the first surface. When this is done, the light beam is extinguished,

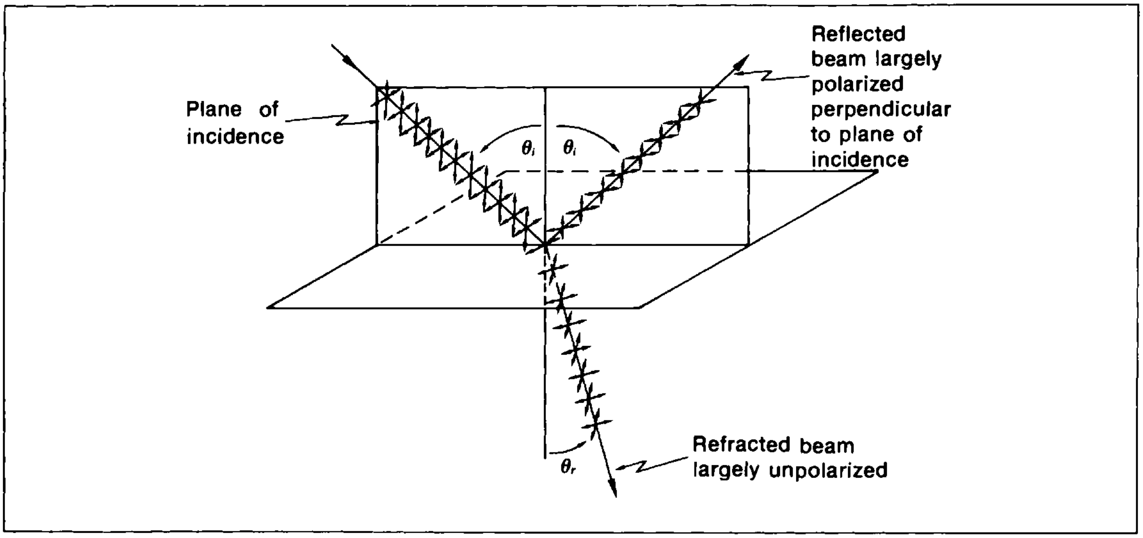


Figure 13-6 Polarization of a light wave by reflection.

showing that the polarized light is parallel to the now-rotated plane of incidence. At the Brewster angle, total absorption occurs and no reflection is observed. (One could of course use a polaroid analyzer to show the same thing, but it is not necessary to do so.)

Another polarization phenomenon was discovered by Bartholinus in the seventeenth century. Materials such as calcite, tourmaline, and quartz, for example, were found to be capable of separating a light beam into two components in the following sense. If a flat plate of such a crystal were placed over a spot of light, two spots could be observed at the upper surface. If the plate were rotated about an axis parallel to the normal, one spot remained fixed in position, but the other rotated about the first. It is conventional to refer to materials exhibiting this property as doubly refracting or *birefringent*. The ray giving rise to the non-rotated spot is called the *ordinary ray*, and the ray related to the rotating spot is called the *extraordinary ray*.

Experiments with these materials have shown that there does exist one direction in such crystalline materials for which the ordinary and extraordinary rays are not separated. This unique direction is called the *optic axis*. Huygens explained double refraction by assuming that in the direction of the optic axis the wavelets associated with the ordinary or O ray spread out spherically, while those of the extraordinary or E ray spread out in ellipsoidal form, with the optic axis as an axis of revolution for the ellipsoids. For directions not parallel to the optic axis, the wavelets of the E- and O-rays travel in the same direction but at different speeds. Thus, there will be two different indices of refraction for a given wavelength. It is common to quote values of the indices of refraction for a direction at right angles to the optic axis in which the E-ray wavelets have a maximum (or a minimum) speed. As Table 13-2 indicates, the speed associated with the E-ray wavelets may be larger or smaller than that of the O-rays.

Although we will not pursue the matter, it is worth noting that double refraction has its origin in the anisotropy, or non-symmetric, arrangement of the atoms or molecules making up the crystalline material. Since a light wave propagates in a transmitting material by excitation of the constituent electrons, any variation in environment encountered in various directions might be expected to influence the propagation character of incident light.

Table 13-2 Indices of Refraction.†

	n_O	n_E
Calcite	1.6583	1.4864
Quartz	1.5442	1.5533

†For sodium light, $\lambda = 5893$ Å.

The value of doubly refracting materials in the case of polarization rests on the fact that the ordinary and extraordinary waves are found to be linearly polarized at right angles to one another. There are two methods for obtaining linearly polarized light that make use of this fact. In the one method, a material such as calcite is cut into a particular shape, separated into two pieces, and cemented together again using a transparent material (such as Canada balsam) with an index of refraction that can cause total internal reflection of the O-ray while transmitting the E-ray, thus producing a linearly polarized beam. Such a polarizer is known as a Nicol prism, and the interested reader will find additional details in almost any advanced optics text.

The second method relies on the phenomenon of dichroism—the ability of some doubly refracting materials to strongly absorb one component of the light passing through them, either the E-ray or the O-ray. The inventor of polaroid, E. H. Land, discovered that needle-like crystals of herapathite (quinone iodosulfate), arranged to lie with their optic axes parallel and embedded in thin sheets of a cellulose material, are extremely dichroic. This discovery and subsequent improvements have made possible the development of large surface-area polarizers which have had widespread applications in science and industry.

PROBLEMS

1. Two transverse waves of equal amplitude, frequency, and wavelength propagate simultaneously along the same string. They are described by the relations

$$x_1 = A \sin\left[\frac{2\pi}{\lambda}(z - vt) + 90°\right]$$

and

$$x_2 = A \sin\left[\frac{2\pi}{\lambda}(z - vt) + 30°\right].$$

   (a) Using the principle of superposition, find the resulting wave disturbance on the string.
   (b) Then repeat the process for the case $A_1 = 2A_2 = 2A$.

2. Starting with Eqs. (13-3) and (13-4), carry out the sequence of operations indicated to verify Eq. (13-5).

3. Two sinusoidal transverse waves propagating in the direction of the z-axis have vibrations in the x- and y-directions, respectively. The frequency of vibration in the x-direction is one-half the frequency of vibration in the y-direction ($f_x = \frac{1}{2}f_y$). Using the graphical technique illustrated in Figure 13-1, construct the Lissajous figures which result when $\Delta\phi$ = (a) 0, (b) 30°, (c) 45°, (d) 60°, and (e) 90°. These figures are what is observed with an oscilloscope for various $\Delta\phi$.

4. Repeat Problem 3 for the case $2f_x = 3f_y$ and $\Delta\phi = 90°$. Deduce from the result that the frequency ratio $f_x/f_y = n_v/n_H$, where n_v is the number of times the figure touches a vertical edge of the oscilloscope screen and n_H is the number of times the figure touches a horizontal edge of the screen. (Warning: As illustrated in Problem 3, there are values of $\Delta\phi$ for which this interpretation would give incorrect results. See, for example, the figure in Problem 3 for $\Delta\phi = 90°$.)

5. A vector **A** has components A_x, A_y, and A_z in the x-, y- and z-directions, respectively. If α, β, γ are the angles that **A** makes with the x-, y- and z-axes, respectively,
   (a) show that $\cos\alpha = A_x/A$, $\cos\beta = A_y/A$, and $\cos\gamma = A_z/A$, where $A = [A_x^2 + A_y^2 + A_z^2]^{1/2}$, and
   (b) show that $\cos^2\alpha + \cos^2\beta + \cos^2\gamma = 1$. This result shows that if two directional angles of the vector **A** are known, the third follows from this identity.

6. Suppose $\mathbf{A} + \mathbf{B} + \mathbf{C} = 0$ and also that $A = B = C$. Find the angle between **A** and **B**.

7. Vector **A** is oriented at an angle θ with respect to vector **B**. For $\mathbf{A} = A_x\mathbf{i} + A_y\mathbf{j}$ and $B = B_x\mathbf{i}$,
   (a) show that $|\mathbf{A} + \mathbf{B}| = [(A_x + B_x)^2 + A_y^2]^{1/2}$, and that
   (b) $A_x = A\cos\theta$, $A_y = A\sin\theta$.
   (c) Therefore, show also that

$$|\mathbf{A} + \mathbf{B}| = [A^2 + B^2 + 2AB\cos\theta]^{1/2}.$$

   This relation is a slightly rearranged form of the law of cosines for trigonometry. In texts on vector analysis, it is shown that much of plane and solid geometry and trigonometry follows readily from vector operations.

8. Suppose vector **A** lies in the x-y plane and makes an angle of 45° with the x-axis. Give the magnitude and orientation of a vector **B** such that $|\mathbf{A} + \mathbf{B}| = |\mathbf{A} - \mathbf{B}|$.

9. Given $\mathbf{A} = 5\mathbf{i} - 10\mathbf{j} + 8\mathbf{k}$, and $\mathbf{B} = 1\mathbf{i} + 20\mathbf{j} + 12\mathbf{k}$.
   (a) Find the magnitude of $\mathbf{A} + \mathbf{B}$, $\mathbf{A} - \mathbf{B}$, and $2\mathbf{A} + \mathbf{B}$.
   (b) Find the angles **A** makes with the x-, y- and z-axes, respectively. (See Problem 5.)

10. Two polaroids with crossed axes of polarization are placed in a beam of light. A third polaroid is placed between the other two with its axis of polarization at an angle θ with the axis of the first polaroid. If the incident light is unpolarized and has an intensity I_0, show that the intensity of the light transmitted by the third polaroid is given by

$$I = \frac{1}{8} I_0 \sin^2 2\theta.$$

11. (a) For the situation described in the previous problem, determine the angle θ which provides the maximum transmitted light intensity.
    (b) Determine the maximum transmitted light intensity.
12. A stack of 7 polaroids is assembled with the axis of polarization of each polaroid making an angle of 15° with the one preceding it. Thus, the first and last polaroids are crossed. If unpolarized light of intensity I_0 is incident on the stack, what is the intensity of the transmitted light?
13. In Problems 10 to 12, it was assumed that the polaroids are ideal—that is, that light is transmitted only if its polarization axis is parallel to the polarization axis of the polaroid. In practice, it is more nearly true that a polaroid transmits a fraction p^2 of the incident light if the polaroid axis and the polarization axis are parallel and a fraction s^2 is transmitted when the two axes are perpendicular. (For an ideal polaroid, $p^2 = 1$ and $s^2 = 0$.) Unpolarized light of intensity I_0 is incident on a pair of real polaroid sheets that have an angle θ between their polarization axes. Show that the transmitted intensity is given by

$$I = \tfrac{1}{2}(p^4 + s^4)\cos^2\theta + p^2s^2\sin^2\theta.$$

    Show also that this reduces to the ideal expression discussed previously.
14. A beam of light in water ($n = 1.33$) is incident on flint glass ($n = 1.65$). For what angle of incidence will the reflected light be linearly polarized?
15. (a) Sunlight is reflected from the smooth surface of a pond ($n = 1.33$). At what angle relative to the vertical should one view the surface through a polaroid if surface glare is to be eliminated most effectively?
    (b) Should the polarization axis of the polaroid be vertical or horizontal?
16. Light from a sodium lamp ($\lambda = 5893$ Å) is incident normally on a rectangular plate of calcite which is cut so that the optic axis is parallel to the faces of the plate. Determine the wavelengths λ_O, λ_E of the ordinary and extraordinary rays in the plate.
17. When the plate of Problem 16 is cut to the proper thickness, the number of wavelengths of the ordinary ray in the plate will differ from the number of wavelengths of the extraordinary ray in the plate by one-quarter of a wavelength. Such a plate is called a quarter-wave plate. Determine the thickness of a quarter-wave plate of calcite for sodium light.
18. Show that the thickness of a half-wave plate of a birefringent material for light of wavelength λ (in vacuum) is given by

$$t = \frac{\lambda}{2(n_O - n_E)} \quad \text{or} \quad \frac{\lambda}{2(n_E - n_O)},$$

    where the denominator is to be positive. This, of course, depends on whether $n_E > n_O$ (like quartz) or $n_E < n_O$ (like calcite). In a half-wave plate, there is one-half wave more of one ray in the plate than there is of the other ray. Thus, the phase difference $\Delta\phi$ between the ordinary and extraordinary rays is π when the light has traversed the plate between two polaroids.
19. A half-wave plate of calcite (for sodium light) between two polaroids is placed in an unpolarized beam of sodium light. The plate is oriented so that the optic axis is at 45° to the polarization axis of the first polaroid. If the polarization axis of the second polaroid is parallel to the axis of the first no light will be transmitted by the system.
    (a) Explain why. Hint: See Problem 17.
    (b) What will be the situation when the two polaroids are crossed (but the optic axis of the half-wave plate remains at 45° to the axis of the first polarizer)?
20. Repeat Problems 16 and 17 for a quartz quarter-wave plate.

14 Particle Motion

14-1 INTRODUCTION

In this chapter, we develop a vocabulary for discussing motion in a quantitative manner. Motion is a concept with an imprecise meaning for us in terms of everyday experience. If we are to gain an understanding of the connection between motion and its physical causes, it is first necessary that we develop a clearer picture of motion in terms of geometric relationships and time. Such a study of motion (excluding its causes) is known as kinematics, while the more inclusive study (which correlates motion and its causes) is called dynamics.

The study of kinematics is made simpler initially by discussing the motion of a fictional or ideal object, the point-particle. Such a particle is completely described kinematically if the location of the point in space is known as a function of time. By contrast, a real, or physical, particle occupies a volume in space, the dimensions of which as well as the location of which may vary with time in a complicated manner. The concept of a point-particle is a simplification that is not only useful but is also a quite reasonable approximation for real situations in which the object in question has dimensions that are of negligible extent by comparison with other objects in its environment. For example, one may regard the motion of a Tiros satellite about the Earth as that of a point-particle because of the extreme disparity in dimensions.

Furthermore, the point-particle concept can be useful for discussing a rigid body of non-negligible dimensions. In this case, the object is regarded as an aggregation of a large number of point-particles that move in a correlated way because of influences whose nature will be discussed at a later point in our study of physics. To see that this point of view is plausible, one need only recall that any physical object is composed of one or more extremely small entities known as molecules, each of which can be said to approximate the ideal point-particle. A more familiar example of a collection of point-particles moving together is provided by a football that has been kicked off and is in motion. To the receiver watching it move toward him, it is irrelevant to observe that the football is in reality many many molecules of "pigskin" bound together in such a way that they entrap and thus compel a great number of air molecules to accompany the "pigskin" in its motion downfield. To the receiver, it is only necessary that he control his motion so that his hands arrive at a given limited region of space at the same instant that the collection of "pigskin" and air molecules arrive at that same location.

With these considerations as a motivation, we adopt the point-particle concept as a means of acquiring a clear understanding of motion, deferring until later a consideration of the complications that can accompany the motion of a large aggregate of correlated (essentially) point-particles when the aggregate is of non-negligible dimensions.

14-2 DISPLACEMENT

In Chapter 13, we indicated that a point in space is uniquely located if its three spatial coordinates are given with respect to a particular origin of coordinates. The vector specifying this point, called the position vector, in rectangular coordinate form is written as

$$\mathbf{r} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k} \tag{14-1}$$

where **i**, **j**, and **k** are the vectors of unit length in the x-, y-, and z-directions respectively as in Chapter 13, and x, y, and z represent the multiples of the unit lengths in the three directions, respectively.

Let us consider a particle (henceforth in this chapter particle will mean point-particle) whose position is initially given by

$$\mathbf{r}_1 = x_1\mathbf{i} + y_1\mathbf{j} + z_1\mathbf{k},$$

and which is then shifted to a new position given by

$$\mathbf{r}_2 = x_2\mathbf{i} + y_2\mathbf{j} + z_2\mathbf{k}.$$

Such a particle is said to have experienced a displacement, defined by the vector equation

$$\mathbf{d} \equiv \mathbf{r}_2 - \mathbf{r}_1 = (x_2 - x_1)\mathbf{i} + (y_2 - y_1)\mathbf{j} + (z_2 - z_1)\mathbf{k}. \tag{14-2}$$

From the laws of vector algebra (Chapter 13), the magnitude of **d** is given by

$$|\mathbf{d}| = |(\mathbf{d}\cdot\mathbf{d})^{1/2}| = |[(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2]^{1/2}|, \tag{14-3}$$

and the direction cosines of the displacement with respect to the coordinate axis are obtained by forming the dot or scalar product of **d** with the respective coordinate axes. Thus, for example, the angle between **d** and the x-axis is given by the relation

$$\mathbf{d}\cdot\mathbf{i} = |\mathbf{d}||\mathbf{i}| \cos(\mathbf{d}, \mathbf{i}) = (x_2 - x_1)$$

or

$$\cos(\mathbf{d}, \mathbf{i}) = \frac{(x_2 - x_1)}{d}, \tag{14-4}$$

with similar expressions for the other two coordinate axes.

It should be stressed here that the displacement is not the same as the distance traveled by the particle in getting from $\mathbf{r}_1$ to $\mathbf{r}_2$. The distance traveled is a scalar quantity that is vitally dependent upon the path taken in moving from $\mathbf{r}_1$ to $\mathbf{r}_2$, while the displacement is a vector quantity giving the magnitude and direction of the separation distance between $\mathbf{r}_1$ and $\mathbf{r}_2$. To see the distinction, consider a particle that was originally located in New York City, that is moved to Chicago via turnpikes, and returned to its starting point by some alternative route. For this situation, the displacement is zero since the initial and final positions are identical. The distance traveled, however, is non-zero. Furthermore, the actual distance traveled cannot be given without identifying the complete route taken.

Example 1. A particle moves along a circular path of radius R from point 1 to point 2 and on to point 3 as in Figure 14-1. Find the displacement and the distance traveled in moving from point 1 to (a) point 2, and (b) point 3.

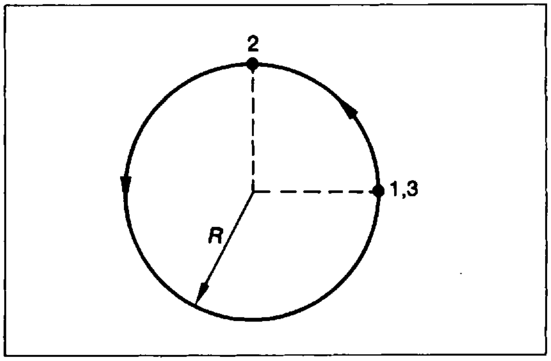


Figure 14-1 System diagram for Example 1.

SOLUTION

$\mathbf{r}_1 = R\mathbf{i} = \mathbf{r}_3$, $\mathbf{r}_2 = R\mathbf{j}$. Therefore, for case (a), $\mathbf{d} = (\mathbf{r}_2 - \mathbf{r}_1) = R(\mathbf{j} - \mathbf{i})$ so that $d = R((1)^2 + (-1)^2)^{1/2} = \sqrt{2}R$, and the distance traveled is given as $s = \pi R/2$ since it represents one-fourth of the circumference of the circle.

For case (b), $\mathbf{d} = \mathbf{r}_3 - \mathbf{r}_1 = 0$, and the distance traveled is $s = 2\pi R$, the circumference of the circle.

14-3 VELOCITY AND ACCELERATION

Although changes in location are adequately described by the displacement, one frequently is more interested in the manner in which the displacement varies with time or the extent to which the change in displacement with time is itself a function of time. Accordingly, we define the velocity of a particle as the time rate of displacement. It is a vector quantity, since displacement is a vector quantity while time is a scalar quantity.

If the particle displacement $\mathbf{d} = \mathbf{r}_2 - \mathbf{r}_1 = \Delta\mathbf{r}$ occurs in the time interval $\Delta t = t_2 - t_1$, where Δt is finite, then the average velocity is defined to be

$$\bar{\mathbf{v}} = \frac{\Delta\mathbf{r}}{\Delta t} = \frac{(x_2 - x_1)\mathbf{i} + (y_2 - y_1)\mathbf{j} + (z_2 - z_1)\mathbf{k}}{t_2 - t_1}, \quad (14\text{-}5)$$

so that

$$\bar{\mathbf{v}} = \bar{v}_x\mathbf{i} + \bar{v}_y\mathbf{j} + \bar{v}_z\mathbf{k}, \quad (14\text{-}6)$$

where

$$\bar{v}_x = \frac{x_2 - x_1}{t_2 - t_1}, \text{ etc.}$$

The magnitude of $\bar{\mathbf{v}}$ is found by application of the scalar product to be

$$|\bar{\mathbf{v}}| = (\bar{v}_x^2 + \bar{v}_y^2 + \bar{v}_z^2)^{1/2}. \quad (14\text{-}7)$$

Example 2. Suppose that the displacement in Example 1(a) takes place in 10 seconds. Then the average velocity will be

$$\mathbf{v} = \frac{\mathbf{r}_2 - \mathbf{r}_1}{t_2 - t_1} = \frac{R}{10}(\mathbf{j} - \mathbf{i}),$$

and the magnitude of the average velocity is

$$v = \sqrt{\left(\frac{R}{10}\right)^2 + \left(\frac{-R}{10}\right)^2} = \frac{\sqrt{2}R}{10}.$$

If the time interval during which the displacement occurs becomes very short, it is usually found that the magnitude of $\Delta\mathbf{r}/\Delta t$ approaches a constant value as a result of the physical fact that as the time interval decreases so also does the displacement. Thus, if we reduce the time interval until subsequent reductions do not change the value obtained for $\Delta\mathbf{r}/\Delta t$, we have effectively determined the velocity at a particular instant of time within the interval $t_2 - t_1$. This limiting process is, of course, more rigorously presented in the differential calculus in defining the process of differentiation. Thus, one writes that the instantaneous velocity v is defined by the relation

$$\mathbf{v} = \lim_{\Delta t \to 0} \frac{\Delta\mathbf{r}}{\Delta t} = \frac{d\mathbf{r}}{dt} \quad (14\text{-}8)$$

Graphically, the distinction between $\mathbf{v}$ and $\bar{\mathbf{v}}$ is clear. Figure 14-2 shows a schematic graph of $\mathbf{r}$ versus t for one-dimensional motion (along the x-axis, for example). From our discussion above, $\bar{\mathbf{v}}$ is represented by the ratio of the chord length between points 1 and 2 and the time interval $\Delta t = t_2 - t_1$. On the other hand, $\mathbf{v}$ is the value of the slope of the curve at a particular instant (or at least in a very small time interval in the neighborhood of that instant), for example, at point 3. At any interval about point 3 for which the chord has the same slope as the curve at point 3, the average velocity and the instantaneous velocity at point 3 will be the same. For example, if $\mathbf{r}$ is a linear function of t ($\mathbf{r} = \mathbf{r}_0 + \mathbf{c}t$), then

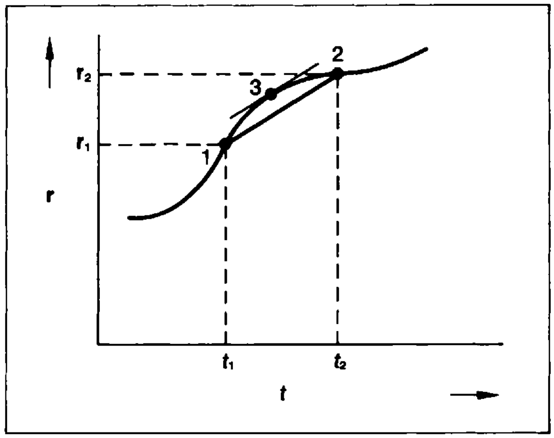


Figure 14-2 Schematic graph of one-dimensional displacement as a function of time.

$$\bar{\mathbf{v}} = \frac{\Delta\mathbf{r}}{\Delta t} = \frac{\mathbf{r}_2 - \mathbf{r}_1}{t_2 - t_1} = \frac{(\mathbf{r}_0 + \mathbf{c}t_2) - (\mathbf{r}_0 + \mathbf{c}t_1)}{t_2 - t_1} = \mathbf{c}.$$

But

$$\mathbf{v} = \frac{d\mathbf{r}}{dt} = \frac{d(\mathbf{r}_0 + \mathbf{c}t)}{dt} = \mathbf{c}.$$

Therefore, the average velocity and the instantaneous velocity are identical when the position of a particle varies linearly with time. On the other hand, if $\mathbf{r}$ is not a linear function of time but is known either as an analytic function of time or as a table of data relating $\mathbf{r}$ and t values, then $\bar{\mathbf{v}}$ and $\mathbf{v}$ can be found either by analytic or numerical (or graphical) techniques in a straightforward manner.

It also should be noted that the speed of a particle in general is not equal to the magnitude of the velocity, since the speed is defined to be the distance traveled divided by the time, a scalar quantity which is path dependent, while the velocity is a path independent vector quantity.

There are abundant examples in our daily lives of motion for which the velocity is non-constant. Thus, the scene available on any busy city street will show several "particles" whose velocities are changing in various ways (to "beat the light" or avoid a dented fender, for example). To discuss such motions, we define the acceleration of a particle in a manner analogous to that used in defining the velocity. Thus, the acceleration is defined as a vector quantity that gives the change of velocity with respect to time. The average and instantaneous acceleration, respectively, are defined as

$$\bar{\mathbf{a}} = \frac{\Delta \mathbf{v}}{\Delta t} = \frac{\mathbf{v}_2 - \mathbf{v}_1}{t_2 - t_1} = \frac{\Delta}{\Delta t}\left(\frac{\Delta \mathbf{r}}{\Delta t}\right) \qquad (14\text{-}9)$$

and

$$\mathbf{a} = \lim_{\Delta t \to 0} \frac{\Delta \mathbf{v}}{\Delta t} = \frac{d\mathbf{v}}{dt} = \frac{d^2\mathbf{r}}{dt^2}. \qquad (14\text{-}10)$$

Example 3. A particle obeys a displacement-time relationship given by $\mathbf{r} = (5 + 2t + 3t^2)\mathbf{i} + (4t - 6t^2 + 3t^3)\mathbf{j}$, where $\mathbf{r}$ is in meters and t is in seconds.

(a) Find the velocity and acceleration at $t = 0$ seconds.

(b) Find the velocity and acceleration at $t = 2$ seconds.

SOLUTION

(a) $$\mathbf{v}(t) = \frac{d\mathbf{r}}{dt} = [(2 + 6t)\mathbf{i} + (4 - 12t + 9t^2)\mathbf{j}] \text{ m/sec},$$

$$\mathbf{a}(t) = \frac{d\mathbf{v}}{dt} = [6\mathbf{i} + (-12 + 18t)\mathbf{j}] \text{ m/sec}^2;$$

therefore,

$$\mathbf{v}(0) = (2\mathbf{i} + 4\mathbf{j}) \text{ m/sec}, |\mathbf{v}(0)| = 2\sqrt{5} \text{ m/sec},$$

$$\mathbf{a}(0) = (6\mathbf{i} - 12\mathbf{j}) \text{ m/sec}^2, |\mathbf{a}(0)| = 6\sqrt{5} \text{ m/sec}^2.$$

Similarly, for

(b) $$\mathbf{v}(2) = (14\mathbf{i} + 16\mathbf{j}) \text{ m/sec}, |\mathbf{v}(2)| = 2\sqrt{113} \text{ m/sec},$$

$$\mathbf{a}(2) = (6\mathbf{i} + 24\mathbf{j}) \text{ m/sec}^2, |\mathbf{a}(2)| = 6\sqrt{17} \text{ m/sec}^2.$$

The average velocity and acceleration between $t = 0$ and $t = 2$ seconds are found to be

$$\bar{\mathbf{v}} = \frac{\mathbf{r}(2) - \mathbf{r}(0)}{2 - 0} = \frac{1}{2}[21\mathbf{i} + 8\mathbf{j}) - 5\mathbf{i}]$$

$$= (8\mathbf{i} + 8\mathbf{j}) \text{ m/sec}, \quad |\bar{\mathbf{v}}| = 8\sqrt{2} \text{ m/sec},$$

and

$$\bar{\mathbf{a}} = \frac{\mathbf{v}(2) - \mathbf{v}(0)}{2 - 0} = \frac{1}{2}[14\mathbf{i} + 16\mathbf{j} - (2\mathbf{i} + 4\mathbf{j})]$$

$$= (6\mathbf{i} + 6\mathbf{j}) \text{ m/sec}^2, \quad |\bar{\mathbf{a}}| = 6\sqrt{2} \text{ m/sec}^2.$$

The reader can determine the appropriate direction of the various vector quantities by means of the scalar product as in Eq. (14-4). It should be noted that the motion in the x- and y- directions had different characteristics. The acceleration in the x-direction was constant so that the velocity in the x-direction was a linear function of time, while the acceleration in the y-direction varied linearly with time so that the y-velocity was a quadratic function of time. This is a result that is quite general in that the type of motion in one direction can usually be discussed independently of the motion in either of the two remaining directions in space, with the total motion being obtained by vector addition of the motion in the three space directions. Thus, for example,

$$\mathbf{a} = a_x\mathbf{i} + a_y\mathbf{j} + a_z\mathbf{k}$$

$$= \frac{dv_x}{dt}\mathbf{i} + \frac{dv_y}{dt}\mathbf{j} + \frac{dv_z}{dt}\mathbf{k}$$

$$= \frac{d^2x}{dt^2}\mathbf{i} + \frac{d^2y}{dt^2}\mathbf{j} + \frac{d^2z}{dt^2}\mathbf{k}.$$

14-4 UNIFORMLY ACCELERATED MOTION

Let us now consider the kinematical description appropriate for a particle for which the acceleration is constant. Then, since

$$\mathbf{a} = \frac{d\mathbf{v}}{dt} = \textbf{constant},$$

it follows that

$$\mathbf{a} = \bar{\mathbf{a}} = \frac{\mathbf{v} - \mathbf{v}_0}{t - 0}.$$

Therefore,

$$\mathbf{v} = \mathbf{v}_0 + \mathbf{a}t. \qquad (14\text{-}11)$$

To determine the displacement in a time t, we first find the average velocity in the interval $(t - 0)$ to be the average

$$\bar{\mathbf{v}} = \frac{\mathbf{v} + \mathbf{v}_0}{2} = \frac{\mathbf{v}_0 + \mathbf{a}t + \mathbf{v}_0}{2} = \mathbf{v}_0 + \frac{\mathbf{a}t}{2}. \qquad (14\text{-}12)$$

(Equation (14-12) simply states that the average velocity for a particle undergoing constant acceleration for a time t is given by the initial velocity plus the acceleration times one-half of the total time interval $(t - 0)$.) Next, the displacement during the time interval $(t - 0)$ is found from the relation

$$\mathbf{d} = \bar{\mathbf{v}}(t - 0) = \mathbf{v}_0 t + \frac{1}{2}\mathbf{a}t^2. \qquad (14\text{-}13)$$

Since

$$\mathbf{d} = \mathbf{r} - \mathbf{r}_0,$$

we finally have

$$\mathbf{r} = \mathbf{r}_0 + \mathbf{v}_0 t + \frac{1}{2}\mathbf{a}t^2. \qquad (14\text{-}14)$$

Alternatively, we can relate the initial and final velocities and the displacement by eliminating t in Eqs. (14-11) and (14-13). This is most readily seen in component form. Thus, for the x-components, we have

$$v_x = v_{0x} + a_x t$$

and

$$d_x = x - x_0 = v_{0x}t + \frac{1}{2}a_x t^2,$$

from which

$$t = \frac{v_x - v_{0x}}{a_x}$$

and

$$d_x = v_{0x}\frac{(v_x - v_{0x})}{a_x} + \frac{1}{2}a_x\left(\frac{v_x - v_{0x}}{a_x}\right)^2$$

$$= \frac{v_x^2 - v_{0x}^2}{2a_x}$$

or

$$v_x^2 = v_{0x}^2 + 2a_x d_x. \qquad (14\text{-}15)$$

Similar relations in the y- and z-directions combine vectorially to yield

$$v^2 = v_0^2 + 2\mathbf{a}\cdot\mathbf{d}. \qquad (14\text{-}16)$$

Returning to Eq. (14-13), we see that the displacement is the sum of two terms. The first, $\mathbf{v}_0 t$, represents the displacement due to a constant velocity $\mathbf{v}_0$. The second term, $(\frac{1}{2}\mathbf{a}t)t$, represents the displacement due to a linear change in the velocity from $\mathbf{v}_0$ to $\mathbf{v}_0 + \mathbf{a}t$ in the time interval $(t - 0)$. A graph of $\mathbf{v}$ versus t for one-dimensional motion with constant acceleration is shown in Figure 14-3. The two terms above correspond to the areas labeled 1 and 2 under the linear curve, $\mathbf{v}_0 t$ and $\frac{1}{2}t(\mathbf{v}_0 + \mathbf{a}t - \mathbf{v}_0)$, respectively. Hence, $\mathbf{d} = \mathbf{v}_0 t + \frac{1}{2}\mathbf{a}t^2$, as we have already found.

This graphical interpretation is more generally applicable than just to the case of uniformly accelerated motion. To see that this is true, consider the situation illustrated in Figure 14-4(a). Here, $\mathbf{a}$ is not constant since $\mathbf{v}$ is a non-linear function of t. However, one can regard the curve as a sum of curve segments for time intervals small enough to regard $\mathbf{v}$ as a linear function of time in that interval as shown in Figure 14-4(b). Then the total displacement is the sum of the displacements for each small interval during which $\mathbf{a}$ assumes a constant value. Thus,

$$\mathbf{d} \approx \sum_{i=0}^{N} \mathbf{d}_i = \sum_{i=0}^{N} [\mathbf{v}(t_i + \Delta t) + \mathbf{v}(t_i)]\frac{\Delta t}{2}. \qquad (14\text{-}17)$$

It is at least reasonable that greater accuracy will result if Δt is made increasingly smaller. In the limit as Δt becomes vanishingly small, so that the number of terms N in the sum becomes infinite, the expression

$$\mathbf{d} = \lim_{\substack{t \to 0 \\ N \to \infty}} \sum_{i=0}^{N} [\mathbf{v}(t_i + \Delta t) + \mathbf{v}(t_i)]\frac{\Delta t}{2}$$

becomes identical to the definition of the definite integral,

$$\mathbf{d} = \int_0^t \mathbf{v}(t)dt. \qquad (14\text{-}18)$$

Therefore, if the velocity is known as an analytical or graphical function of time, the displacement

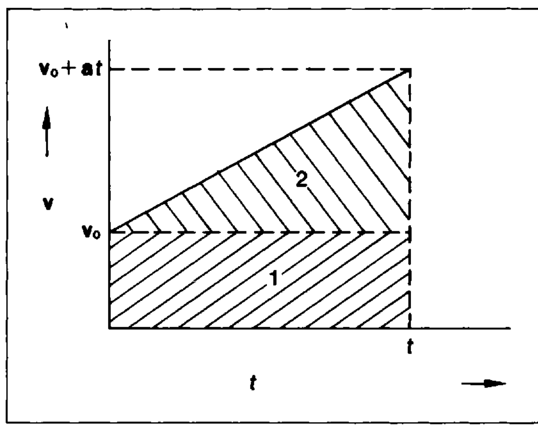


Figure 14-3 Velocity versus time for uniformly accelerated motion.

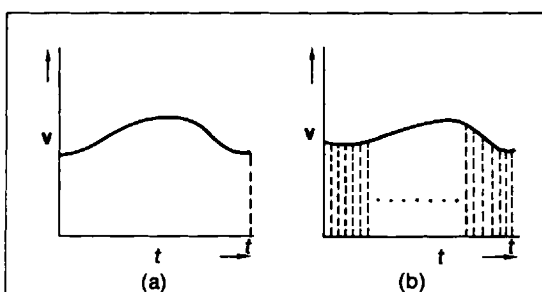


Figure 14-4 Velocity versus time for non-uniformly accelerated motion.

can be obtained as the integral of the velocity over the appropriate time interval (or, equivalently, by determining the area under the velocity-versus-time graph in the units of the graph). The reader might wish to show by an analogous argument that for an arbitrary **a**-versus-t relationship, $\mathbf{v}(t)$ can be found by the relation

$$\mathbf{v}(t)=\int_0^t \mathbf{a}(t)dt+\mathbf{v}_0 \tag{14-19}$$

or, equivalently, by determining the area under the **a** versus t curve in the units of the graph.

14-5 FRAMES OF REFERENCE

In Sections 14-3 and 14-4, we defined the quantities displacement, velocity, and acceleration with respect to a given system of coordinates. Thus, when we write $\mathbf{r}=x\mathbf{i}+y\mathbf{j}+z\mathbf{k}$, it is understood that x, y, and z are the spatial coordinates of a particle with respect to a given origin (for example, the center of the Earth). It is clear that the values for x, y, and z of the same point in space will be different if the point is referred to a new origin of coordinates (such as a particular point on the Earth's surface). Should observations be made of the particle by observers at rest at two different locations, they can correlate their information only if they are able to translate (both literally and figuratively!) their measurements from one origin to the other or vice versa. Such a translation corresponds to a transformation from one frame of reference to another. To be specific, suppose an observer is at rest at the origin of a system of coordinates (a reference frame) labeled S_1, and another observer is at rest at the origin of another reference frame labeled S_2. Suppose, further, that the origins of the two reference frames are at rest and separated by a displacement $\mathbf{r}_{21}$ (the subscripts are read as: the location of the origin of S_2 with respect to the origin of S_1). Thus, it follows that $\mathbf{r}_{21}=-\mathbf{r}_{12}$. Now a particle at point P in space would be described in S_1 as having a position vector $\mathbf{r}_{P1}$, and in S_2 as having a position vector $\mathbf{r}_{P2}$. As illustrated in Figure 14-5, the translation equation relating these position vectors is correctly given as

$$\mathbf{r}_{P1}=\mathbf{r}_{P2}+\mathbf{r}_{21}. \tag{14-20}$$

Example 4. Two students at opposite ends of a laboratory table are at rest with respect to the table when they observe a flash of light across the room in a direction parallel to the length of the table. One student reports a distance of 20 m to the point where the flash occurred, while the other observed a distance of 17 m. What is the length of the laboratory table?

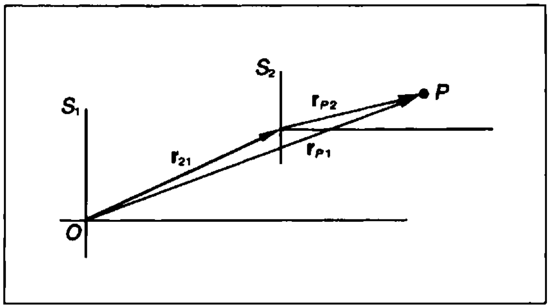


Figure 14-5 Translation of a position vector from one reference frame to another.

SOLUTION

From the information given, one can write $\mathbf{r}_{P1}=20\mathbf{i}$ m, $\mathbf{r}_{P2}=17\mathbf{i}$ m. Therefore, from Eq. (14-20),

$$20\mathbf{i}\text{ m}=17\mathbf{i}\text{ m}+\mathbf{r}_{21}\quad\text{or}\quad\mathbf{r}_{21}=3\mathbf{i}\text{ m},$$

so that the laboratory table is 3 m in length. Notice that we could have alternatively written $\mathbf{r}_{P1}=17\mathbf{i}$ m, $\mathbf{r}_{P2}=20\mathbf{i}$ m, which would lead to $\mathbf{r}_{21}=-3\mathbf{i}$ m. The physical interpretation here is still that the laboratory table is 3 m in length, but in this case the observer in S_1 is nearer to the flash of light than the observer in S_2, and the observer in S_2 is therefore 3 m behind him. (One would *not* conclude that the laboratory table was 3 m less than zero length!) In similar situations, one should always consider the physical situation involved to avoid any apparent mathematical paradoxes.

14-6 RELATIVE VELOCITY AND THE GALILEAN TRANSFORMATION

Now we consider the problem of transforming observations made by an observer in S_1 if the observer in S_2 is in motion relative to him. To be specific, suppose that at some instant of time the origins of S_1 and S_2 coincide, and that the two observers synchronize their watches. The reference frame S_2 and the observer at the origin of S_2 are moving with a constant velocity $\mathbf{v}$ relative to another observer at the origin of S_1. A somewhat frivolous example of such a situation could be approximated by two students in a laboratory corridor. One of the students is at rest while the other is moving down the corridor on a skateboard at a constant velocity v relative to the first student. At the instant the two are side by side, they synchronize their watches to read zero. We now assume that the motion does not alter the synchronization of the watches. (We shall see in the next chapter that this assumption breaks down if v is comparable in magnitude to the velocity of light.) It then follows that the separation of the two origins is given by

$$\mathbf{r}_{21} = \mathbf{v}t$$

or

$$\mathbf{r}_{12} = -\mathbf{v}t. \qquad (14\text{-}21)$$

To continue the example, suppose that at a time t the two students simultaneously observe a stationary rabbit down the corridor. The situation is depicted in Figure 14-6. It is clear that in this case

$$\mathbf{r}_{P1} = \mathbf{r}_{P2} + \mathbf{r}_{21} = \mathbf{r}_{P2} + \mathbf{v}t \qquad (14\text{-}22)$$

or, conversely

$$\mathbf{r}_{P2} = \mathbf{r}_{P1} - \mathbf{v}t. \qquad (14\text{-}23)$$

In the diagram, we imply that the motion is one-dimensional (along the corridor), but we could consider a more complex situation (for example, a hummingbird hovering at a definite point in space along the corridor while the student on the skateboard moves at constant speed diagonally along the corridor). In the more complicated case, we would simply decompose the vector motion into its spatial component equations and consider them separately. Thus, Eq. (14-22) would become the three equations

$$x_{P1} = x_{P2} + v_x t,$$

$$y_{P1} = y_{P2} + v_y t,$$

$$z_{P1} = z_{P2} + v_z t.$$

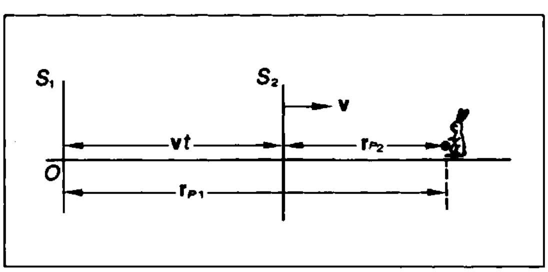


Figure 14-6 Diagram for the system described by Eq. 14-22.

We see, therefore, that there is no essential increase in conceptual difficulty if the relative motion is not one-dimensional.

To see how the relative motion of the two observers will affect a kinematical description, let us now assume that the rabbit moves from a point observed in S_1 given by $\mathbf{r}_{P1}(t_1)$ at time t_1 to a point $\mathbf{r}_{P1}(t_2)$ at time t_2. The corresponding locations as observed in S_2 will be found by means of Eq. (14-23). Now we ask: What relationship is there between the average velocities of the rabbit as measured by each observer? From Eq. (14-5), the average velocity $\bar{\mathbf{u}}_1$ as measured in S_1 is given by

$$\bar{\mathbf{u}}_{P1} = \frac{\Delta\mathbf{r}_{P1}}{\Delta t} = \frac{\Delta\mathbf{r}_{P1}(t_2) - \Delta\mathbf{r}_{P1}(t_1)}{t_2 - t_1}. \qquad (14\text{-}24)$$

From Eq. (14-22), this is equivalent to

$$\bar{\mathbf{u}}_{P1} = \frac{(\mathbf{r}_{P2}(t_2) + \mathbf{v}t_2) - (\mathbf{r}_{P2}(t_1) - \mathbf{v}t_1)}{(t_2 - t_1)}$$

$$= \frac{\mathbf{r}_{P2}(t_2) - \mathbf{r}_{P2}(t_1)}{(t_2 - t_1)} + \mathbf{v} = \frac{\Delta\mathbf{r}_{P2}}{\Delta t} + \mathbf{v}$$

$$= \bar{\mathbf{u}}_{P2} + \mathbf{v} \qquad (14\text{-}25)$$

or, conversely,

$$\bar{\mathbf{u}}_{P2} = \bar{\mathbf{u}}_{P1} - \mathbf{v}. \qquad (14\text{-}26)$$

Once again in the limit as Δt becomes vanishingly small, the average velocities in these equations may be replaced by the corresponding instantaneous velocities. Thus, we find that the observers obtain results that differ by a velocity representing their relative motion. Do they also obtain differing results for average and instantaneous accelerations observed for the rabbit? The answer is no and follows from the use of Eq. (14-9):

$$\bar{\mathbf{a}}_{P1} = \frac{\mathbf{u}_{P1}(t_2) - \mathbf{u}_{P1}(t_1)}{t_2 - t_1} = \frac{\Delta\mathbf{u}_{P1}}{\Delta t}$$

$$= \frac{(\mathbf{u}_{P2}(t_2) + \mathbf{v}) - (\mathbf{u}_{P2}(t_1) + \mathbf{v})}{t_2 - t_1}$$

$$= \frac{\Delta\mathbf{u}_{P2}}{\Delta t} = \bar{\mathbf{a}}_{P2}. \qquad (14\text{-}27)$$

Similarly, in the limit as $\Delta t \to 0$, we obtain the result that

$$\mathbf{a}_{P1} = \mathbf{a}_{P2}. \qquad (14\text{-}28)$$

These results may be summarized by the statement that for reference frames that differ by a constant relative velocity the acceleration remains invariant to (is independent of) a transformation from one reference frame to the other, while the position and velocity transform according to Eqs. (14-22) and (14-25) or, conversely, Eqs. (14-23) and (14-26). These transformation equations are known as the Galilean transformation equations. They are satisfactory only if the magnitudes of the velocities involved are small compared to the speed of light. In the next chapter, we derive the transformation equations known as the Lorentz equations required for high velocity situations. Since there is ample experimental evidence that the Galilean transformation equations are adequate for low velocities, it is reasonable to anticipate that the Lorentz transformation equations must possess a form that reduces to the Galilean equations whenever the velocities are small in magnitude compared to the speed of light. This expectation, in fact, will assist us in finding the form of the Lorentz equations.

We conclude this chapter with a final example of relative motion, which also proves to be useful in the next chapter in describing the Michelson–Morley experiment.

Example 5. Consider a stream of width w, which has a uniform velocity (relative to the shore) of magnitude $\mathbf{v}$. Two swimmers A and B who swim at a velocity $\mathbf{u}$ relative to the water are to engage in a contest in which A swims directly across the stream and back, while B swims downstream a distance w and back again (see Figure 14-7). Who wins the contest, and what is the difference in times required by the two in completing their round trips?

SOLUTION

Swimmer B's velocity downstream relative to the shore is given by

$$\mathbf{v}_{BS} = \mathbf{v}_{BW} + \mathbf{v}_{WS} = \mathbf{u} + \mathbf{v},$$

and his upstream velocity is

$$\mathbf{v}_{BS} = \mathbf{u} - \mathbf{v}.$$

Therefore, the time required by B is given by

$$t_{\text{total } B} = \frac{w}{u+v} + \frac{w}{u-v} = \frac{2\dfrac{w}{u}}{1-\left(\dfrac{v}{u}\right)^2}.$$

For swimmer A, it is not sufficient to swim straight across the stream. If he did so, the movement of the water would carry him downstream, resulting in his covering a distance greater than $2w$. He must, therefore, head upstream at an angle θ (see Figure 14-8) such that the component of his velocity (relative to the water) parallel to the shore is equal in magnitude but opposite in direction to the stream velocity relative to the shore. If he does so, his speed perpendicular to the shore is reduced to

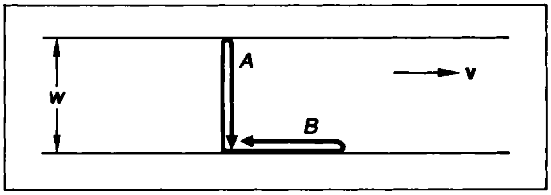


Figure 14-7 System diagram for Example 5.

$$u_{AW_\perp} = u\cos\theta.$$

Since $u\sin\theta = v$, it follows that

$$u_{AW_\perp} = u(1-\sin^2\theta)^{1/2} = u\left(1-\left(\frac{v}{u}\right)^2\right)^{1/2}.$$

As a result, his time of crossing one way will be

$$t_1 = \frac{w}{u_{AW_\perp}} = \frac{\dfrac{w}{u}}{\left[1-\left(\dfrac{v}{u}\right)^2\right]^{1/2}}.$$

On the return trip, it will again be necessary to head upstream by a similar amount for the same reason, so the return time would be identical to t_1. Therefore, the total time for A is

$$t_{\text{total } A} = \frac{2\dfrac{w}{u}}{\left[1-\left(\dfrac{v}{u}\right)^2\right]^{1/2}}.$$

Now, the expression

$$1-\left(\frac{v}{u}\right)^2 < \left[1-\left(\frac{v}{u}\right)^2\right]^{1/2}.$$

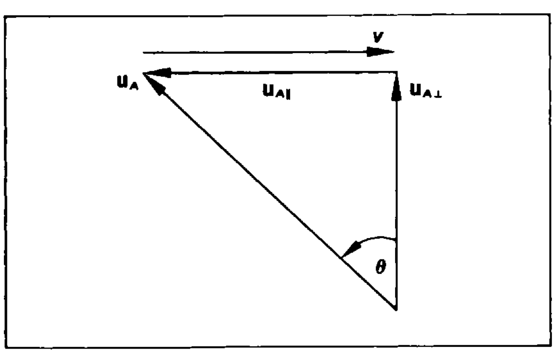


Figure 14-8 Velocity vector diagram for swimmer A.

Therefore, we conclude that swimmer A will take less time than B. The difference in elapsed times will be

$$\Delta t = t_B - t_A = \frac{2\frac{w}{u}}{\left[1-\left(\frac{v}{u}\right)^2\right]^{1/2}}\left[\frac{1}{\left[1-\left(\frac{v}{u}\right)^2\right]^{1/2}} - 1\right]. \tag{14-29}$$

Finally, consider the case for which $(v/u)^2 \ll 1$. Recall that the binomial expansion is written

$$(1 \pm x)^n = 1 \pm nx + \frac{n(n-1)x^2}{2!} + \frac{n(n-1)(n-2)x^3}{3!} + \cdots.$$

For $x \ll 1$,

$$(1 \pm x)^n \approx 1 \pm nx.$$

In our case, the term $[1-(v/u)^2]^{-1/2}$ becomes approximately

$$1 + \frac{1}{2}\left(\frac{v}{u}\right)^2,$$

so that

$$\Delta t \approx 2\left(\frac{w}{u}\right)\left[\frac{1}{2}\left(\frac{v}{u}\right)^2\right] \approx \frac{wv^2}{u^3}. \tag{14-30}$$

While this result is not realistic for swimmers in a river, we will make use of Eq. (14-30) in our discussion of the Michelson–Morley experiment.

PROBLEMS

1. At time $t = 0$, a body is traveling due east with a linear speed of 32.5 ft/sec. At time $t = 4$ seconds, the body is traveling due north with a linear speed of 32.5 ft/sec. What is the magnitude of the average acceleration experienced by the body?
2. At the instant a motor patrolman waiting at an intersection starts ahead with a green light, an automobile passes him with a constant speed of 50 ft/sec. The patrolman accelerates at 10 ft/sec^2. How far from the intersection will the patrolman overtake the automobile?
3. A stone is dropped from the top of a cliff 256 ft in height and, at the same instant, another stone is projected upward from the bottom level with a speed of 96 ft/sec. At what height will the two stones meet, or pass each other?
4. A dense ball is projected vertically upward at 30 m/sec. It rises higher than an observation point that is 15 m above its starting point.
   (a) Compute the magnitude of its velocity as it passes the observation point on the way down.
   (b) Compute the time from the start until the ball passes the observation point on the way down.
5. An automobile is initially traveling with a velocity of 40 ft/sec north. Two minutes later the automobile has a velocity of 120 ft/sec south. Determine the average acceleration experienced by the automobile.
6. A ball shot vertically upward from the edge of a vertical cliff travels at 50 ft/sec when it reached a point 25 ft below its starting point.
   (a) What was its initial velocity?
   (b) How high did it rise above the cliff?
7. A ball is projected upward from the ground with a velocity of 60 ft/sec.
   (a) How high will it rise?
   (b) Find its velocity and height 3 seconds after it left the ground.
8. A ball released to roll down an inclined track has constant acceleration of 30 cm/sec^2. When it passes one mark on the track, the ball has a velocity of 40 cm/sec, and when it passes a second mark, it is going 130 cm/sec. Compute:
   (a) the average velocity of the ball between the two marks,
   (b) the distance between the two marks, and
   (c) the time required for the ball to go from the first mark to the second mark.

9. It takes 30 seconds for a car to travel half way around a circular track from the eastern-most point to the western-most point of the track. The radius of the track is 1500 ft.
   (a) Determine the car's displacement during this 30 second interval.
   (b) Determine the car's average speed during this interval.
10. An automobile accelerates uniformly to 15 ft/sec in 10 seconds. Compute:
    (a) the average speed,
    (b) the average acceleration, and
    (c) the distance traveled in the 10 seconds.
11. A car having an initial velocity of 8 m/sec south experiences, by the application of the brakes, a constant deceleration of 2 m/sec^2. Determine:
    (a) the velocity after 3 seconds and
    (b) the distance traveled and the average velocity during this time.
12. A raindrop falls at a rate of 50 ft/sec, and is at the same time blown to the east at 10 ft/sec by the wind. If a car driving due east at 80 ft/sec has the drop graze its side window, find the angle its trace will make with the vertical.
13. A ball is thrown vertically upward from the ground, and a student gazing out of the window sees it moving upward past him at 16 ft/sec. The window is 32 ft above the ground.
    (a) How high does the ball go above the ground?
    (b) How long does it take to go from a height of 32 ft to its highest point?
    (c) What was the initial velocity given to the ball?
    (d) Find its velocity and acceleration one-half second after it left the ground.

15 Special Relativity

15-1 INTRODUCTION

In this chapter, we briefly discuss the modifications required when transformation equations are needed for situations where the magnitudes of the velocities involved are comparable to the speed of light. As was asserted in the previous chapter, the Galilean transformation equations are then no longer adequate. The need for modifications was recognized as a result of the Michelson–Morley experiment, which is described in the next section. In order to "explain" the results of this experiment, Hendrik Lorentz and George Fitzgerald independently proposed an *ad hoc* solution by suggesting that moving objects suffer a contraction in their length in the direction of motion. Such a contraction would predict results in agreement with the Michelson–Morley experiment, but such a bizarre postulate was scarcely satisfactory philosophically.

A more fundamental approach was provided by the special theory of relativity developed in 1905 by Albert Einstein. As a result of this theory, the desired transformation equations, known as the Lorentz transformation equations, are readily derived as in Section 15-3. The remaining sections of the chapter are devoted to an examination of the kinematical consequences of the Lorentz transformation and the extent to which they are at variance with our usual experience.

15-2 THE MICHELSON–MORLEY EXPERIMENT

In the 1880s, the wave theory of light (and, in fact, of electromagnetic radiation in general) was firmly established due to the experimental work of Michael Faraday (and later of Heinrich Hertz) and the theoretical work of Clerk Maxwell. A fundamental part of the wave theory was the assumption that an extremely rigid, yet transparent and highly permeable medium existed known as the luminiferous ether (for brevity, we refer to it henceforth as the ether), which served as the means of transfer of solar and stellar radiation. Since sound waves do not propagate without air or other molecules to transmit the disturbance, it was argued by analogy that electromagnetic waves also required a medium for their transmission.

Therefore, the ether was assumed to occupy all of space, and the Sun was assumed to be at rest in it while the planets moved freely through it. As a result, the ether (if its existence could be demonstrated) could be regarded as a unique or absolute frame of reference to which all others might be referred. Michelson first and then Michelson and Edward Morley sought to demonstrate by interference techniques that the ether did exist and possess the characteristics ascribed to it. The failure of their experiments to indicate its existence led first to great consternation, followed by the efforts of Lorentz, Fitzgerald, and Henri Poincaré to theoretically explain the null results.

Figure 15-1 is a schematic representation of the Michelson–Morley experiment. A Michelson interferometer (discussed in Chapter 11, Figure 11-5) is arranged so that one of the light paths will be at right angles to the ether through which the Earth is moving, while the other light path is parallel to the movement through the ether. (From the point of view of the interferometer, the ether is streaming through the interferometer with a speed equal in magnitude to the orbital velocity of the Earth about the Sun.) This physical situation was regarded as analogous to the contest of the two swimmers (Example 5 of Chapter 14) in that a difference in time for the light to traverse the two different paths would exist because of the movement relative to the ether.

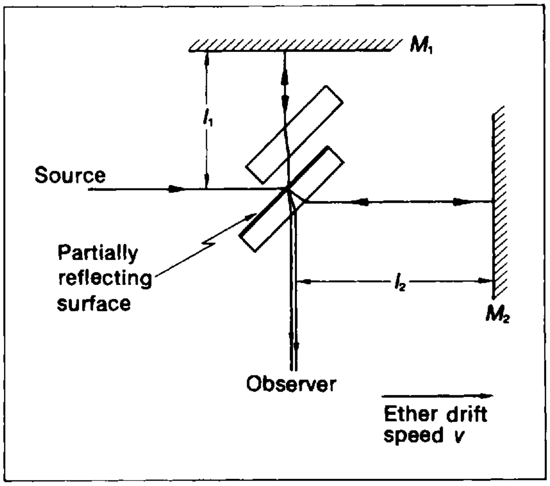


Figure 15-1 The Michelson interferometer as used in the Michelson–Morley experiment.

This difference in times was to be observed directly in terms of a shift of the fringe pattern when the interferometer was rotated through 90° to reverse the paths with respect to the ether. The interferometer fringe pattern arises because the two mirrors M_1 and M_2 are not exactly perpendicular. As the interferometer is rotated, a measurable shift of the fringe pattern was expected to occur. To estimate the magnitude of the shift, we make use of the special case of Example 5, Section 6, Chapter 14. When v is the ether drift velocity $\approx 3 \times 10^4$ m/sec and $u = c = 3 \times 10^8$ m/sec, the difference in time for the two paths becomes

$$\Delta t \approx \frac{wv^2}{c^3}. \qquad (15\text{-}1)$$

For the interferometer, a time difference Δt is equivalent to an optical path difference Δl through the relation

$$\Delta l = c\,\Delta t = n\lambda, \qquad (15\text{-}2)$$

where Δl represents n fringes of light of wavelength λ. Thus, the expected number of fringes will be

$$n \approx \frac{c}{\lambda}\frac{wv^2}{c^3} = \frac{w}{\lambda}\left(\frac{v}{c}\right)^2. \qquad (15\text{-}3)$$

Since each orientation will exhibit this fringe pattern, the shift should be

$$n_{\text{total}} = 2n \approx \frac{2w}{\lambda}\left(\frac{v}{c}\right)^2. \qquad (15\text{-}4)$$

By means of multiple reflections, Michelson and Morley were able to have $w \approx 10$ m and $\lambda \approx 5000$ Å. As a result, they anticipated $n_{\text{total}} \approx 0.4$, a value substantially within the experimental limits of precision.

As indicated earlier, however, their result was no observable shift, a result which has since been verified repeatedly by other experimenters at various locations and at various times of the year. Later experiments by Roy Kennedy and Edward Thorndike, utilizing an interferometer having interferometer arms of differing lengths, obtained the same null result. All these results demonstrate that the ether, if it exists, does not possess any experimentally observable properties. As a consequence, its value in any theory must be severely limited.

Furthermore, the experiments, including those of Kennedy and Thorndike, would strongly suggest that the most likely way of explaining the null result would be to assume that the speed of light in free space was not affected by any motion of the source or the observer. If this were true, then there would be no time difference for the two paths, which is in agreement with experiment. In this connection, it is interesting to note that when Einstein published his special theory of relativity in 1905 he was apparently unaware of the Michelson–Morley experiment. Instead, his paper dealt directly with the description of events as viewed in reference frames moving at high constant velocity relative to one another.

The results of his analysis completely resolved the puzzle of the Michelson–Morley experiment without directly attempting to do so.

15-3 EINSTEIN'S SPECIAL THEORY OF RELATIVITY

The conclusions Einstein reached in considering reference frames in constant relative motion are summarized in the two postulates that constitute the special theory of relativity. We state them as follows:

1. The speed of light in free space will have the same value for all observers, regardless of their relative motion.
2. The laws of physics must retain the same equation form for all frames of reference moving at constant velocity relative to one another.

In this section, we derive the transformation equations obtained by Einstein using these postulates. They are identical to those obtained by Lorentz a year earlier in an *ad hoc* fashion. We begin by considering two reference frames S_1 and S_2. S_2 is in motion at a constant velocity $\mathbf{v} = v\mathbf{i}$ with respect to S_1. (The restriction of relative motion along the x-axis is for convenience only. It does not represent a fundamental restriction.) As noted in the previous chapter, the transformation equations obtained here must reduce to the Galilean transformation equations in the limit of $v/c \ll 1$. We therefore assume (because of the simplicity obtained) that the transformation equations connecting S_1 to S_2 have the form

$$\begin{aligned} x_1 &= k(x_2 + vt_2), \\ y_1 &= y_2, \\ z_1 &= z_2. \end{aligned} \tag{15-5}$$

We note that for $v/c \ll 1$, k must go to 1 to give the Galilean transformation. By Postulate 2, we must have an identical form of transformation equations connecting S_2 to S_1, except that now v is replaced by $-v$ because we are considering S_1 from the point of view of S_2 instead of the former case. Therefore, we must require

$$\begin{aligned} x_2 &= k(x_1 - vt_1), \\ y_2 &= y_1, \\ z_2 &= z_1. \end{aligned} \tag{15-6}$$

Now, however, we note an unexpected result. That is, the times measured in the two frames are not identical as they were taken to be in the Galilean transformations. To see this, substitute Eq. (15-5) in Eq. (15-6), to obtain

$$x_2 = k\{k(x_2 + vt_2)\} - vt_1$$

or

$$(k^2 - 1)x_2 + kvt_2 = kvt_1,$$

so that

$$t_1 = t_2 + \frac{(k^2 - 1)}{kv}x_2. \tag{15-7}$$

Notice once again that if $v/c \ll 1$ so that $k \to 1$, then Eq. (15-7) reduces to $t_1 = t_2$—the Galilean transformation relation.

Next, we consider the following situation. Assume that at $t_1 = 0$, the origins of S_1 and S_2 coincide, so that $t_2 = 0$ also [from Eq. (15-7)]. At this instant, a light flash is generated at the common origin and subsequently propagates as a spherical wave. Each of the observers will describe the spreading wave front as a spherical wave by Postulate 2. Since the relative motion is along the x-direction only, the y and z relationships are trivial ($y_1 = y_2$, $z_1 = z_2$).

In S_1,

$$x_1 = ct_1, \tag{15-8}$$

and in S_2,

$$x_2 = ct_2, \tag{15-9}$$

by Postulate 1.

From Eqs. (15-6) and (15-7), Eq. (15-9) becomes

$$k(x_1 - vt_1) = ckt_1 + \frac{(1 - k^2)}{kv}x_1$$

or

$$x_1 = \frac{vk^2(v + c)}{k^2(v + c) - c}t_1. \tag{15-10}$$

But Eq. (15-8) requires that

$$x_1 = ct_1.$$

We see, therefore, that

$$c = \frac{vk^2(v + c)}{k^2(v + c) - c},$$

from which

$$k^2 = \frac{c^2}{c^2 - v^2} = \frac{1}{1 - (v/c)^2}$$

and

$$k = \left[1 - \left(\frac{v}{c}\right)^2\right]^{-1/2}, \tag{15-11}$$

where we have chosen the positive root to satisfy the requirement that $k \to 1$ when $v/c \ll 1$.

Thus, the Lorentz transformation equations are seen to be

$$x_2 = \frac{(x_1 - vt_1)}{\left[1 - \left(\frac{v}{c}\right)^2\right]^{1/2}},$$

$$y_2 = y_1,$$

$$z_2 = z_1,$$

$$t_2 = \frac{t_1 - \frac{vx_1}{c^2}}{\left[1 - \left(\frac{v}{c}\right)^2\right]^{1/2}}. \tag{15-12}$$

The inverse transformation from S_2 to S_1 can be shown by similar arguments to be given by the equations

$$x_1 = \frac{(x_2 + vt_2)}{\left[1 - \left(\frac{v}{c}\right)^2\right]^{1/2}},$$

$$y_1 = y_2,$$

$$z_1 = z_2,$$

$$t_1 = \frac{t_2 + \frac{vx_2}{c^2}}{\left[1 - \left(\frac{v}{c}\right)^2\right]^{1/2}}. \tag{15-13}$$

We see that our assumed transformation equation form, in addition to being simple, reduces properly to the Galilean transformation equation as required. What we had not anticipated was the prediction that not only position measurements but also time measurements are affected by the motion of the observer. In the sections that follow, we consider some of the consequences of these transformation equations on measurements required for a kinematic description of physical events as seen by observers in constant relative motion.

15-4 THE LORENTZ CONTRACTION

Consider the following experiment. An observer at rest in S_1 locates the ends of a rod also at rest in S_1. He finds that its length is given by $L_0 = x_{1R} - x_{1L}$, where the subscripts R and L refer to the right and left ends, respectively. A second observer in a frame S_2 moving in the x-direction with speed v relative to S_1 measures the length of the same rod to be $L = x_{2R} - x_{2L}$. How do L_0 and L compare? From Eqs. (15-13) we obtain

$$x_{1R} = \frac{(x_{2R} + vt_2)}{\left[1 - \left(\frac{v}{c}\right)^2\right]^{1/2}},$$

$$x_{1L} = \frac{(x_{2L} + vt_2)}{\left[1 - \left(\frac{v}{c}\right)^2\right]^{1/2}}.$$

Therefore,

$$L_0 = x_{1R} - x_{1L} = \frac{(x_{2R} - x_{2L})}{\left[1 - \left(\frac{v}{c}\right)^2\right]^{1/2}}$$

or

$$L = \left[1 - \left(\frac{v}{c}\right)^2\right]^{1/2} L_0. \tag{15-14}$$

We see that an object measured by an observer at rest in the same reference frame will have a maximum length. An observer in motion relative to the object will regard the object as moving relative to him, and will obtain a smaller or contracted dimension along the direction of constant relative motion. Notice that since v appears only as a squared quantity in Eq. (15-14), it is irrelevant which reference frame is regarded as the stationary one. Thus, an observer in S_1 will regard an object at rest in S_2 as contracted, while an observer in S_2 will similarly regard an object at rest in S_1 as contracted. The primary question is simply: What is the state of motion of the object relative to the observer? If they are in the same reference frame, their relative velocity is zero and no motional effects are noted. It is the relative motion that provides results at variance with our Galilean expectations.

Example 1. A moving meter stick appears to be shrunk in length to $\frac{1}{4}$ m as seen by a stationary observer. What is its speed relative to the observer?

SOLUTION

By Eq. (15-14), $L/L_0 = \frac{1}{4} = \sqrt{1-(v/c)^2}$, so that

$$v = \sqrt{\frac{15}{16}}c \approx 0.97c.$$

15-5 TIME DILATION

Let us slightly modify the experiment of Section 15-4. In this instance, a stationary clock in S_1 records a time interval $\Delta t_1 = t_{1f} - t_{10}$. What is the time interval as measured by a clock at rest in S_2, which is moving at speed v relative to S_1 along the x-direction?

Again, from Eq. (15-13), we obtain

$$\Delta t_1 = t_{1f} - t_{10} = \frac{t_{2f} - \frac{vx_2}{c^2}}{\left[1 - \left(\frac{v}{c}\right)^2\right]^{1/2}} - \frac{t_{20} - \frac{vx_2}{c^2}}{\left[1 - \left(\frac{v}{c}\right)^2\right]^{1/2}}$$

$$= \frac{t_{2f} - t_{20}}{\left[1 - \left(\frac{v}{c}\right)^2\right]^{1/2}} = \frac{\Delta t_2}{\left[1 - \left(\frac{v}{c}\right)^2\right]^{1/2}}. \qquad (15\text{-}15)$$

We conclude, therefore, that the moving clock runs slow; that is, the elapsed time registered by it will be less than that registered by the stationary clock. This seemingly bizarre prediction actually finds confirmation in data related to μ mesons (radioactive particles created in the upper atmosphere by cosmic rays). μ mesons created in the laboratory by high energy accelerators have been found to have a mean lifetime before decay of about 2 microseconds. Those created by cosmic rays are observed (by an earthbound observer) to have speeds of about 0.998 c. We might therefore incorrectly conclude that during their mean lifetime after creation they travel a distance given by $y = 0.998\,c \times 2 \times 10^{-6} \approx 600$ m, whereas they are actually created at altitudes an order of magnitude higher and are still detected at the Earth's surface. We can resolve this paradox in either of two ways.

First, in the rest frame of the μ mesons (S_2), the distance traveled is given by $y_2 = 0.998\,c \times 2 \times 10^{-6} \approx 600$ m, since the mean lifetime was measured in the laboratory for which the μ meson velocity was negligible. On the other hand, an earthbound observer would translate the distance traveled by the moving μ meson to be

$$y_1 = \frac{y_2}{\sqrt{1 - \left(\frac{v}{c}\right)^2}} \approx \frac{600}{\sqrt{1 - (0.998)^2}} \approx 9000 \text{ m},$$

in agreement with observations.

Second, the earthbound observer could measure the lifetime of the μ meson to be not 2×10^{-6} sec, but the longer time (since the μ meson's "clock" is in motion) of

$$\Delta t_1 = \frac{2 \times 10^{-6}}{\left[1 - \left(\frac{v}{c}\right)^2\right]^{1/2}} = \frac{2 \times 10^{-6}}{[1 - (0.998)^2]^{1/2}} \approx 3 \times 10^{-5} \text{ sec},$$

from which $y_1 = v\Delta t_1 \approx 9000$ m as before.

In either case, we find that the proper use of the Lorentz transformation equations leads to results in accord with experiment. At this point, we pause to recognize a common objection expressed by students of physics encountering time dilation or length contraction for the first time. That is the feeling that these results are in violent contrast to "common sense." However, one must realize that the term "common sense" necessarily refers to our everyday experience, which is (for nearly all of us) restricted to kinematic situations for which $v \ll c$. As a result, these seemingly absurd effects are not observable. Therefore, while the predictions are at variance with "common sense" as identified here, they are in no way rendered impossible. It would be more satisfactory in fact to redefine "common sense" to mean "in agreement with experimental results." In this way, the apparently unsatisfactory predictions should be as acceptable as the predictions of the Galilean transformation equations.

15-6 RELATIVISTIC ADDITION OF VELOCITIES

In this section, we consider the effect of the Lorentz equations in the addition of velocities as in Eq. (14-22). From Eqs. (15-12), we can write

$$x_{2R} - x_{2L} = \frac{(x_{1R} - x_{1L}) - v(t_{1f} - t_{10})}{\left[1 - \left(\frac{v}{c}\right)^2\right]^{1/2}}$$

and

$$t_{2f} - t_{20} = \frac{(t_{1f} - t_{10}) - \frac{v(x_{1R} - x_{1L})}{c^2}}{\left[1 - \left(\frac{v}{c}\right)^2\right]^{1/2}}.$$

Since

$$u_{2x} = \frac{x_{2R} - x_{2L}}{t_{2f} - t_{20}} \quad \text{and} \quad u_{1x} = \frac{x_{1R} - x_{1L}}{t_{1f} - t_{10}}, \qquad (15\text{-}16)$$

we obtain from Eqs. (15-16) the result

$$u_{2x} = \frac{u_{1x} - v}{1 - \frac{vu_{1x}}{c^2}}. \qquad (15\text{-}17)$$

Similarly, we find that

$$u_{2y} = \frac{y_{2R} - y_{2L}}{t_{2f} - t_{20}}$$

$$= \frac{y_{1R} - y_{1L}}{\dfrac{1}{\left[1-\left(\dfrac{v}{c}\right)^2\right]^{1/2}}\left(t_{1f} - t_{10} - \dfrac{v(x_{1R} - x_{1L})}{c^2}\right)}$$

$$= \frac{\left[1-\left(\dfrac{v}{c}\right)^2\right]^{1/2} u_{1y}}{1 - \dfrac{vu_{1x}}{c^2}} \qquad (15\text{-}18)$$

and

$$u_{2z} = \frac{\left[1-\left(\dfrac{v}{c}\right)^2\right]^{1/2} u_{1z}}{1 - \dfrac{vu_{1x}}{c^2}}. \qquad (15\text{-}19)$$

As before, the converse equations are obtained by the interchange of quantities from S_2 with those from S_1, and by replacing v by $-v$. In addition, we see again that for $v/c \ll 1$, Eqs. (15-17) through (15-19) reduce to Eq. (14-26) as they should.

On the other hand, the results are strikingly different from the Galilean predictions if the velocities involved are not small. For example, suppose S_2 moves at speed $v = c/2$ along the x-axis relative to S_1, and that the event observed is the transmission of a light signal from a source at rest in S_1, so that $u_1 = c$. Then, Eq. (15-17) predicts that

$$u_{2x} = \frac{c - \dfrac{c}{2}}{1 - \dfrac{c^2}{2c^2}} = c,$$

which is, of course, the result required by the first postulate of special relativity. The reader can in fact show that for any sequence of observers, each moving at a constant velocity relative to the preceding one, successive application of Eqs. (15-17) through (15-19) will provide the correct result.

Example 2. An observer at rest sees two high speed objects approaching him from opposite directions, each with a speed of $0.6\,c$. What is the relative speed of the two rockets?

SOLUTION

From the point of view of the rocket at the left (S_1), the observer has a speed of $0.6\,c$ to the left. The speed of the rocket at the right is denoted by u_{1x} relative to the first rocket. Then, we have

$$u_{1x} = \frac{u_{2x} + v}{1 + \dfrac{u_{2x}v}{c^2}} = \frac{2(0.6)c}{1 + 0.36} = \frac{0.6}{0.68}c = \frac{1}{1.133}c \approx 0.89\,c.$$

Thus, the relative speed of the two rockets is $\approx 0.89\,c$ and not $1.2\,c$ as predicted by Galilean relativity.

Finally, we remark that this brief discussion by no means exhausts the new predictions that are a consequence of the special theory of relativity. It was our main purpose here to justify the assertion of the previous chapter that at high velocities the effect of motion on measurements is to produce differences in the results obtained, depending upon the state of relative motion. There is abundant literature at all levels (from the elementary to the advanced) on this topic, which is quite properly a subject for additional study in physics. In the following chapters, we shall only occasionally need to take further direct notice of special relativity theory, since our principal interest is in classical physics fundamentals. For this reason, we do not consider at all the general theory of relativity, which takes into account the motional effects that arise when one reference frame is in accelerated motion relative to another.

PROBLEMS

1. Verify the result $n_{\text{total}} \approx 0.4$ fringes for the Michelson–Morley experiment. Use the data given in Section 15-2.
2. Show that applying the Lorentz contraction to the interferometer arm parallel to the ether drift velocity leads to a null result in the Michelson–Morley experiment.
3. Verify Eqs. (15-13). Note that these relations follow from Eqs. (15-12) if one substitutes $-v$ for v and interchanges subscripts 1 and 2 throughout. What is the physical significance of this procedure?
4. Determine the value of the ratio v/c for $L/L_0 = 0.95$, 0.90, 0.75, 0.50, 0.10, and 0.05.
5. What would be the value of the ratio L/L_0 for a speed v of magnitude 600 mi/hr?

6. A high speed runner runs at a constant speed over a distance he judges to be 432 m. His wristwatch indicates the time required was 2.88×10^{-6} sec.
(a) What was his speed? (This will be the same as the speed of the ground relative to him.)
(b) What distance do stationary judges see him travel?
(c) How long does the run take him according to the stationary judges?

7. Two events occurring at the same time in S_1 are separated by a distance of 1 m along the x-axis. In a frame S_2 moving at constant speed along the x-axis, the events are separated by a distance of 2 m. What is the time interval between the two events in S_2?

8. A cube with sides of length L moves with S_2 at speed v relative to S_1.
(a) What is the volume of the cube as measured in S_1?
(b) What is the value of v/c if the volume measured by S_1 is half that measured in S_2?

9. A uranium nucleus traveling at 10^8 m/sec away from S_1 ejects an alpha particle at 1.5×10^8 m/sec relative to itself toward S_1. What is the velocity of the alpha particle as viewed in S_1?

10. A stick of length L_1 is inclined at an angle θ_1 relative to the x-axis in S_1. Find its length L_2 and its angle of inclination θ_2 relative to the x-axis as viewed by an observer in S_2 moving at a speed v relative to S_1 along the x-axis.

11. Starting with the relations

$$u_{2x} = \frac{u_{1x} - v}{1 - \frac{u_{1x}v}{c^2}} \quad \text{and} \quad t_2 = \frac{\left(t_1 - \frac{vx_1}{c^2}\right)}{\left[1 - \left(\frac{v}{c}\right)^2\right]^{1/2}},$$

use the relations

$$a_{2x} = \frac{du_{2x}}{dt_2}, \quad a_{1x} = \frac{du_{1x}}{dt_1}, \quad \frac{dx_1}{dt_1} = u_{1x},$$

and the fact that v is constant, to show that

$$a_{2x} = \frac{\left[1 - \left(\frac{v}{c}\right)^2\right]^{3/2} a_{1x}}{\left[1 - \frac{vu_{1x}}{c^2}\right]^3}.$$

This transformation relation for the x-component of the moving object viewed from two reference frames with a constant relative velocity will be valid only as long as the velocities u_{1x} and u_{2x} remain less than c.

12. A cosmic ray forms a particle having a lifetime of 10^{-7} sec when measured at rest. How far will it travel before decaying if its speed is $0.95\,c$ when it is created?

16 Principles of Dynamics I

16-1 INERTIA, MOTION, AND FORCES

In this chapter, we shall be concerned with dynamics, the relationship between the motion of physical objects, and the causes producing the motion. Everyone has an intuitive understanding of the relationship, for we are aware that muscular effort in the form of pushes or pulls can cause motion. Furthermore, we distinguish between objects in terms of their relative response to a given push or pull. For example, a man who kicks with equal vigor first a beach ball and then a stone of similar dimensions will have no trouble identifying which one experiences a greater change in its state of motion. To describe this response, we introduce the property of inertia and say that, of two objects subject to the same push or pull, the motion of the one with the greater inertia will be affected least. Alternatively, we can say that an object whose state of motion is observed to remain unchanged is not subject to any net push or pull. That is, because of the inertia property possessed by physical objects, the natural state of motion is one of complete rest or constant velocity and changes from such a state of motion can only occur by the application of some outside effort. Such outside efforts need not be in the form of muscular pushes or pulls. In physics, the term force is applied to any form of effort that can produce a change in the state of motion of an object. At the present time, physicists recognize four fundamental kinds of force: gravitational, electromagnetic, strong nuclear, and weak nuclear forces. Gravitational forces are relatively very weak and will be discussed in Section 16-5. Electromagnetic forces include both electric and magnetic forces on charged particles and are much stronger than gravitational forces (see Chapters 29 and 32). Strong nuclear forces refer to the forces protons and neutrons exert on one another within a given nucleus, while weak nuclear forces are related to the phenomenon of beta decay for some nuclei. These nuclear forces are stronger than electromagnetic forces. However, they are important only for very short distances ($\approx 10^{-15}$ m) and hence for nuclear particles. Since it is necessary to apply the principles of quantum mechanics in this domain, we are not prepared to discuss these forces in the remainder of the text.

16-2 THE CONCEPT OF MASS

Let us return to the kicked beach ball and stone for the purpose of assigning quantitative significance to the inertia property. Since we assumed similar dimensions, it is clear that the volume of an object is not a suitable measure of inertia. It seems intuitively clear that the quantity of matter involved in the case of the stone is greater than that of the beach ball, which would explain a greater inertia. Instead of pursuing the idea of quantity of matter to the microscopic or atomic level, we introduce the concept of the mass of an object to explain the inertia property. In particular, if we regard mass as a fundamental quantity (as we have already done with length and time), then we can define a unit mass and subject it to some reproducible force in order to determine the acceleration (change in motion) produced by the force. Any other mass can be measured in terms of the unit mass by determining the acceleration produced by the same force when applied to the unknown object. We have implied earlier that for a given force the change in motion is less for greater inertia. This suggests that the force should be proportional to the product of mass and acceleration or, since the force F is the same in each case,

$$\mathbf{F} = m\mathbf{a} = (1)\mathbf{a}_1, \qquad (16\text{-}1)$$

where

$\mathbf{a}_1$ = the acceleration of the unit mass

and

$\mathbf{a}$ = the acceleration of mass m.

Thus, the mass m will be given by

$$m = \frac{|\mathbf{a}_1|}{|\mathbf{a}|}. \qquad (16\text{-}2)$$

For any two arbitrary masses subject to the same force, the relation becomes

$$\frac{m_1}{m_2} = \frac{|\mathbf{a}_2|}{|\mathbf{a}_1|}. \qquad (16\text{-}3)$$

In the MKS system of units, the unit mass is the kilogram (kg) and is, by definition, the mass of the platinum-iridium cylinder known as the standard kilogram, which is preserved in France. In addition, from Eq. (16-1), we can derive the unit of force in terms of the three fundamental quantities of mass, length, and time. In the MKS system, the force unit is the newton (nt). A force of one newton will produce an acceleration of one meter per second squared when applied to a mass of one kilogram, or

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

We have previously noted that the second is defined in terms of the atomic vibrations of cesium.†

†See the article "Standards of Measurement," *Scientific American*, June 1968, p. 50.

Specifically, 1 second = 9,192,631,770 Cs vibrations. Since non-atomic standards are perishable, it would be desirable to replace the standard kilogram by an atomic standard (for example, the number of atoms of the ordinary isotope of hydrogen in a particular state of excitation having a mass equal to the standard kilogram). For the immediate future, however, experimental difficulties make it impossible to obtain a precision comparable to that available with the standard kilogram (1 kg masses can be compared to 1 part in 10^8 at the US National Bureau of Standards). Therefore, standards laboratories in the United States, Great Britain, Germany, etc. have carefully prepared replicas of the standard kilogram for use as secondary standards of mass in their respective countries.

Although, for the most part, we shall employ MKS units, it is proper to present a brief discussion of the British engineering system of units because of its widespread use in many English-speaking nations. In this system, the second retains the definition of the MKS system. The unit of length, the foot (ft), is also defined in terms of Kr^{86} orange-red wavelengths;

$$1 \text{ ft} = 12 \text{ in} = 12 \times 41{,}929.399 \text{ wavelengths of } Kr^{86} \text{ orange-red light.}$$

In this system of units, force is chosen to be the third fundamental quantity, and the unit of mass is then defined in terms of force, length, and time. The force unit chosen is the pound (lb), and the mass unit, the slug, is defined as the mass which will experience an acceleration of 1 ft/sec^2 when a force of 1 lb is applied to it, or

$$1 \text{ slug} = \frac{1 \text{ lb}}{1 \text{ ft/sec}^2}.$$

By defining the pound in terms of the pull of the Earth on a certain standard body at a certain place on the Earth, and comparing the standard body with the standard kilogram, it can be demonstrated that

$$1 \text{ lb} = (0.45359237 \text{ kg})(9.8066 \text{ m/sec}^2) \approx 4.448 \text{ nt},$$

and that

$$1 \text{ slug} = 0.45359237 \times 32.1740 \text{ kg} \approx 14.59 \text{ kg}.$$

Example 1. Two masses m_1 and m_2 are placed on a frictionless, horizontal surface. Each is subjected to the same constant accelerating force. Both are initially at rest. After the time t in each case it is found that the distance traveled by m_1 is three times the distance traveled by m_2. What is the ratio m_1/m_2?

SOLUTION

From Eq. (16-1),

$$F = m_1a_1 = m_2a_2 = \text{constant}.$$

Since the masses are constant, it follows that the accelerations are also constant. As a result, by Eq. (14-13) we can write

$$d_1 = \frac{1}{2}a_1t^2,$$

$$d_2 = \frac{1}{2}a_2t^2$$

or

$$\frac{d_1}{d_2} = \frac{a_1}{a_2}.$$

Since

$$d_1 = 3d_2,$$

$$\frac{a_1}{a_2} = 3.$$

Therefore,

$$\frac{m_1}{m_2} = \frac{a_2}{a_1} = \frac{1}{3}.$$

16-3 WEIGHT

When an object is placed near the Earth's surface, the gravitational force (Section 16-5) acts on the object. If it is released, it will fall to the surface with an acceleration designated by **g**—the acceleration due to gravity. This gravitational force is called the weight (**W**) of the object. Both force and acceleration are vector quantities, while mass is a scalar quantity. Therefore, we write

$$\mathbf{W} = m\mathbf{g}, \qquad (16\text{-}4)$$

which indicates that the gravitational force **W**, and the acceleration **g** it would produce on a mass m if it were released (that is, if **W** were the only force acting on m), both lie in the same direction. The magnitude of **g** varies with location on the Earth's surface, ranging from a value of about 9.78 m/sec^2 at the Equator to about 9.83 m/sec^2 at the North Pole. At a given location, however, **g** is essentially uniform over a limited region of the Earth's surface. Local variations in the value of **g** indicate non-uniform mass concentrations (mascons). Geological surveys for oil and minerals make use of this situation.

The equal arm balance utilizes this fact by letting the force due to gravity act upon two objects—one of which is the unknown, while the other is made up of known masses. When the known mass total on one side of the balance has the same mass as the unknown on the other side, the arm of the balance (see Figure 16-1) becomes horizontal, and the unknown mass is thus determined. This process, unfortunately, is called weighing in practice; the known masses used are called weights, which leads to some confusion. The reason the masses are stamped on the so-called "weights" is that mass is a measure independent of location, while the weight is a measure of the force of gravity on the object, and it varies with location. For example, a mass of 1 kg will weigh about 9.83 nt at the North Pole but only 9.80 nt in New York City. The mass will remain 1 kg anywhere, barring nuclear reactions (fission, fusion) or similar circumstances which we do not consider here.

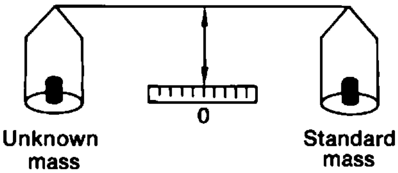


Figure 16-1 The balance (schematic).

16-4 INERTIAL AND GRAVITATIONAL MASS

In Section 16-2, we indicated a procedure for determining the mass of an object in terms of the acceleration produced by the application of a given force. The mass value thus obtained is said to be an *inertial* mass. On the other hand, if we can determine both the weight of the same object (this is discussed in the next section) and the acceleration it experiences when released to fall near the Earth's surface, then we can calculate a *gravitational* mass by means of Eq. (16-4). It is reasonable, but by no means obvious, that the two values should be the same, and many physicists (including Newton) devoted much effort toward providing experimental verification for this assumption. That they are identical to within 3 parts in 10^{11} was demonstrated in 1964 by R. H. Dicke and collaborators at Princeton University.†

This assumed equality was used by Einstein in 1911 as the basis for the principle of equivalence—an important part of his theory of general relativity. Briefly stated, the principle of equivalence says that it is not possible to distinguish between an inertial (non-accelerating) frame of reference in which uniform gravitational forces act, and a non-inertial (accelerating) reference frame where there are no gravitational forces but the reference frame experiences an acceleration $-\mathbf{g}$ relative to an inertial frame.

As an example of this principle, consider two elevator cars, each of which is sealed with an identical observer inside who cannot see outside. The first car is at rest in a location where the acceleration due to gravity is uniform and equal to $\mathbf{g}$. The other is removed to empty space beyond the galaxy where there is no acceleration due to gravity. If this second car is given an acceleration $-\mathbf{g}$ by the application of the necessary force, the observer in this car will experience the same force as the observer in the first car. Furthermore, within their cars there is no experiment that either can perform enabling them to discover which car is which.

One of the important consequences of the principle of equivalence that has been observed experimentally is the predicted bending of a beam of light by a strong gravitational force (for example, a light beam from a distant star bent by the Sun as it approaches the Earth). Unfortunately, it is not possible to pursue this fascinating subject here any further.‡

†See Dicke's article, "The Eötvös Experiment," *Scientific American*, December 1961, p. 84.

‡The interested reader is referred to Chapter 3 in R. Skinner, *Relativity*, Blaisdell, 1969.

16-5 THE LAW OF UNIVERSAL GRAVITATION

In Section 16-3, we referred to the weight of an object at the Earth's surface as a gravitational force. This is because the object and the Earth are observed to attract (gravitate toward) each other unless the object is subject to a restraining force whose magnitude equals the weight but whose direction is opposite to that of the weight.

Example 2. An object rests on a bathroom scale which reads 10 lb. This reading indicates that the scale is exerting an upward force of 10 lb on the object. Since the object is at rest, the weight of the object is 10 lb and has a direction downward, so that no net force is acting on the object.

The distance-dependence of the attraction of one mass for another was first discovered by Newton around 1670, but not published until 1686 (in part because of previous irritating controversies his discoveries had aroused, and also in part because of mathematical questions which were not fully resolved until he had invented the methods of calculus). Actually Newton's discovery, the law of universal gravitation, was of extreme importance for it related not just two masses but *all* masses in the universe. In words, the law states that:

> Every particle of mass in the universe attracts every other particle with a force which is directly proportional to the product of the masses, inversely proportional to the square of the distance between them, and which acts along the line joining them.

In equation form, the law of universal gravitation is expressed as

$$\mathbf{F}_g = \frac{Gm_1m_2}{r_{12}^2}\hat{r}, \qquad (16\text{-}5)$$

where

$\mathbf{F}_g$ = magnitude of the force exerted by particle 1 on particle 2 *and* vice versa;
m_1 = mass of particle 1;
m_2 = mass of particle 2;
r_{12} = separation distance of the two particles;
G = universal gravitational constant;
$\hat{r}$ = unit vector along the line joining the particles.

There are some philosophically unsatisfying features about gravitational forces that the statement of the law of gravitation and Eq. (16-5) fail to clarify. Perhaps the biggest puzzle is, why do masses attract one another especially if they are some distance apart? Experiment will demonstrate that the attraction cannot be electrical in nature because gravitational forces are far weaker than such forces. (For that matter, why do electric charges attract or repel each other as observed?) The reader must understand that, in stating the law of universal gravitation, Newton made no attempt to explain why masses attract one another. What he did was assert a relationship for determining such forces, and then use it to explain a number of observed phenomena (see Section 18-5). Since Eq. (16-5) can be used to successfully predict experimental outcomes ranging from satellite motions to the discovery of hitherto unknown planets, it seems reasonable to accept the reality of gravitational forces, even though their exact nature remains unknown. Certainly it is more fruitful to adopt without a fundamental derivation a law enabling us to obtain a more quantitative understanding of our universe than to reject this understanding because we cannot prove the "truth" of the law. (This attitude, in fact, was another of the many great contributions that Newton made to the development of physics.)

Another feature that is apparently a source of difficulty is the assertion that every particle of mass in the universe is attracting every other particle of mass at every instant. If this is so, how can one possibly ever hope to isolate two masses and perform an experimental test of Eq. (16-5)? Furthermore, we have already defined units of mass, length, and time [so that the unit of force is also determined by Eq. (16-1)]. How then are we to find the value of G, the constant which makes an equation out of the proportionality asserted in the statement of the law of gravitation?

The first question can be answered in terms of an example. Consider a point mass m resting on the surface of the Earth. One can easily show that the force between any two particles is very small by hanging them from slightly separated long thin threads. Because the force each exerts on the other is very small, the strings will both hang vertically; after all, the mass of the Earth is so much greater than any one small particle. Since the attractive forces decrease as the square of the separation distance, it begins to be clear that except for very careful measurements the attractive force due to the rest of the universe on a particle on or near the Earth's surface will be primarily due to the mass of the Earth alone acting on the particle. (Similarly, an astronaut on the surface of the Moon will experience a gravitational force that is almost exclusively due to the mass of the Moon.)

As a result, to determine the gravitational force exerted on the particle, one has only to sum up (vectorially) the forces of attraction between the particle and each element of the Earth. Newton was able to show that if the mass of the Earth is a function only of the radial distance from the center, then the gravitational force exerted on a particle on or above the surface can be calculated by assuming the mass of the sphere is concentrated at its center, giving a single separation distance in Eq. (16-5). In this connection, recall that we have already called this gravitational force on the particle its weight $\mathbf{W}$ in Eq. (16-4). Therefore, if we combine Eqs. (16-4) and (16-5), we obtain

$$\mathbf{W} = m\mathbf{g} = m\frac{Gm_E}{R_E^2}\hat{r}, \qquad (16\text{-}6)$$

where m_E and R_E are the mass and radius, respectively, of the Earth. Since g and R_E are easily found, it follows that a knowledge of G is essentially equivalent to a knowledge of the mass of the Earth. For this reason, when Sir Henry Cavendish determined a value for G in 1798, he was said to have "weighed" the Earth in the sense of Section 16-3.

Let us briefly consider the Cavendish experiment.† Figure 16-2 shows a light rigid arm bearing

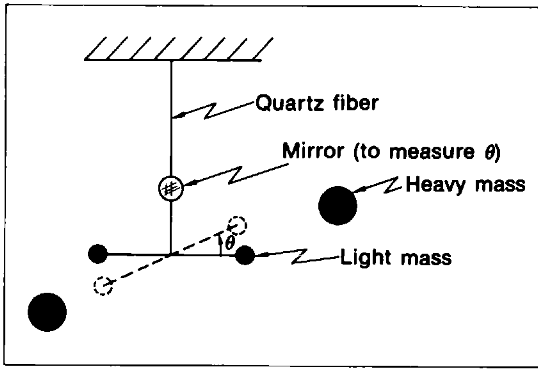


Figure 16-2 Schematic diagram of the Cavendish experiment.

†An annotated version of Cavendish's description of his experiment is given in M. H. Shamos, (ed.), *Great Experiments in Physics*, Holt, Rinehart, and Winston, New York, 1959.

equal spherical masses at its ends, supported at its center by a fine quartz fiber attached to a fixed position. The large dashed circles represent two larger spheres that are brought close (symmetrically) to the smaller masses on the arm. The gravitational force between the pairs of masses causes the quartz fiber to twist through an angle such that the twisting or torsional force due to the internal structure of the fiber just balances the gravitational force due to the mass pairs that are now at a measurable (fixed) separation distance. By measuring the angle of twist and separately measuring the force required to cause that angle of twist, one can calculate G (since the masses and r are also known). The currently accepted value in MKS units is

$$G = 6.670 \times 10^{-11} \text{ nt m}^2/\text{kg}^2. \qquad (16\text{-}7)$$

The reader should show that the units given are correct.

PROBLEMS

1. Starting from rest, a 2 kg mass moves 9 m in 6 seconds due to the application of a constant force. What is the magnitude of the force?
2. A 1 kg mass is observed to experience an acceleration of 3 m/sec^2 when subject to a given force. What will be the acceleration of a 5 kg mass subject to the same force?
3. It was suggested in Section 16-2 that the kilogram be defined (for example) as the number of atoms of ordinary hydrogen required to equal the mass of the standard kilogram. List objections that can be raised to such a procedure.
4. By Eq. (16-4), the weight of a 1 slug mass at a point where g has a magnitude of 32.2 ft/sec^2 is 32.2 lb. Calculate your mass in slugs and in kg. Also find your weight in newtons.
5. For homogeneous materials, the mass m is directly proportional to the volume V of the material. The relation can be written in terms of the density ρ:

$$m = \rho V,$$

where ρ is a constant characteristic of the material. The density of water is 1×10^3 kg/m^3, and the mass of a water molecule is about 30×10^{-27} kg. If the water molecules are in contact with one another, what is their approximate size?
6. A particle of mass m is placed along a line joining two particles (masses M and $3M$) that are separated by a distance of 2 m. At what point will the forces due to M and $3M$ just cancel each other?
7. At what point along the line from the Earth to the Moon is the gravitational pull of the Moon equal to that of the Earth? Give your answer as a fraction of R_{EM}, the distance from the center of the Earth to the center of the Moon. The mass of the Moon is about $\frac{1}{81}$ the mass of the earth. This is the point at which a moonship is said to "leave" the Earth's influence and enter the Moon's.
8. The planet Mars has a radius about 0.52 that of the Earth, and its mass is 0.11 that of the Earth.
   (a) Find the ratio $g_{\text{Mars}}/g_{\text{Earth}}$ (at the surface).
   (b) If an astronaut in full equipment can just leap 2 ft upward on the Earth, how high will he be able to leap on Mars?
9. Two homogeneous spheres of equal radii and made of a material of density ρ are in contact. Show that the force of gravitational attraction between them is given by

$$F_g = \frac{16\pi^2}{9} G r^4 \rho^2.$$

10. Given the value of G from Eq. (16-7) and the value of g (take 9.80 m/sec^2), use Eq. (16-6) to determine the mass of the Earth. The radius of the Earth is about 6.4×10^6 m. Then calculate the average density of the Earth.

11. Find an expression for the variation of the acceleration due to gravity as a function of height h above the Earth's surface in terms of g_0, the acceleration due to gravity at the Earth's surface, and R_E, the radius of the Earth. Also show that at large height ($h \gg R_E$),

$$g \approx g_0 \left(\frac{R_E}{h}\right)^2.$$

12. For an astronaut *on the surface of the Moon*, compute the ratio of the pull of gravity due to the Moon to the pull of gravity on him due to the Earth. The radius of the Moon's orbit about the Earth is about 60 times the Earth's radius. Assume the astronaut is on the side of the Moon facing the Earth.

13. A space explorer becomes separated by 6 m from the surface of a spherical asteroid he has come to call home. His mass including equipment is 90 kg, while the mass of his asteroid is 45,000 kg. If the radius of the asteroid is 30 m, estimate the greatest amount of time required for him to drift back due to gravitational attraction. Assume he starts from rest.

14. To appreciate the weakness of the gravitational force, determine the magnitude of the force between a 1 kg mass and a 10 kg mass separated by a distance of 0.10 m.

15. For the system shown in Figure 16-3:

$m_1 = 1$ kg, $m_2 = 2$ kg, and $m_3 = 3$ kg.

(a) Find the force on m_2 due to m_1.
(b) Find the force on m_2 due to m_3.
(c) Find the vertical and horizontal acceleration components of m_2 due to m_1 and m_2.
(d) Find the direction and magnitude of the resultant acceleration.

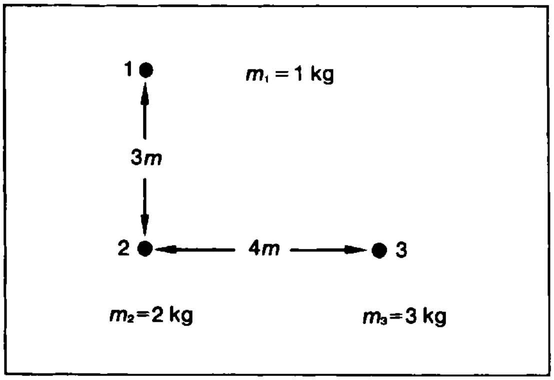


Figure 16-3

17 Principles of Dynamics II

17-1 STATIC AND DYNAMIC EQUILIBRIUM

In the previous chapter, the mass of a particle was related to the applied force and the resulting acceleration by the relation in Eq. (16-1):

$$\mathbf{F} = m\mathbf{a}.$$

If the applied force is zero (equivalently, if the vector sum of all forces acting on the particle is zero), the resulting acceleration is also zero and the particle is said to be in a state of equilibrium. When the velocity is also zero, the particle is at rest or in a state of static equilibrium. When the particle is in uniform motion (constant velocity), the situation is one of dynamic equilibrium. The feature common to both types is a balance of forces resulting in no acceleration.

Example 1. A particle of mass m is supported in a state of rest by two forces $\mathbf{F}_1$ and $\mathbf{F}_2$, which are oriented as shown in Figure 17-1. In addition to these forces, $w = mg$, the weight of the particle is vertically directed downward. Find $\mathbf{F}_1$ and $\mathbf{F}_2$ in terms of the weight w.

SOLUTION

Since the particle is at rest, the horizontal and vertical components of the forces must separately sum to zero. (There is no way a vertical force can balance a horizontal force.) For the horizontal force components,

$$F_1 \cos 37° - F_2 \cos 53° = 0,$$

$$0.80F_1 = 0.60F_2,$$

or

$$F_1 = \frac{3}{4}F_2.$$

For the vertical force components,

$$F_1 \sin 37° + F_2 \sin 53° - w = 0,$$

$$0.60F_1 + 0.80F_2 = w.$$

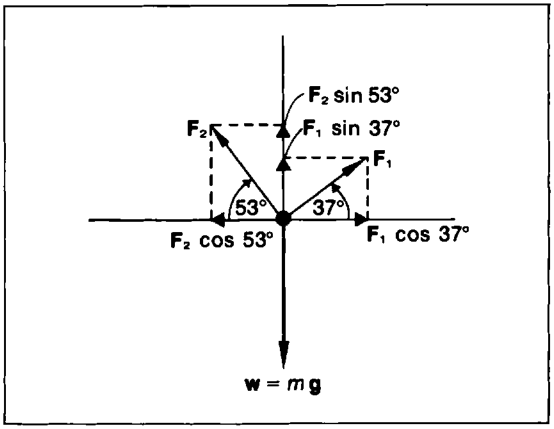


Figure 17-1 Force diagram for Example 1.

Combining these relations, we obtain

$$0.60\left(\frac{3}{4}F_2\right)+0.80F_2 = w,$$

so that

$$1.25F_2 = w$$

or

$$F_2 = 0.80w$$

and

$$F_1 = \frac{3}{4}F_2 = 0.60\,w.$$

Example 2. Figure 17-2 represents a particle of mass m moving to the right with a constant velocity v_x. It is acted upon by a pulling force equal in magnitude to the weight ($F_P = w$), a drag force F_D, an upward force F_N due to the surface and the downward gravitational force w. Find F_N and F_D in terms of w.

SOLUTION

Again, the horizontal force components and the vertical force components must separately sum to zero. Therefore, referring to Figure 17-2, we have

$$F_P \cos\theta - F_D = 0$$

and

$$F_N + F_P \sin\theta - w = 0,$$

so that

$$F_D = F_P \cos\theta = w \cos\theta,$$

$$F_N = w - F_P \sin\theta.$$

Substituting the values given, we obtain:

$$F_D = w \times \frac{1}{2} = \frac{w}{2}\ \text{lb}$$

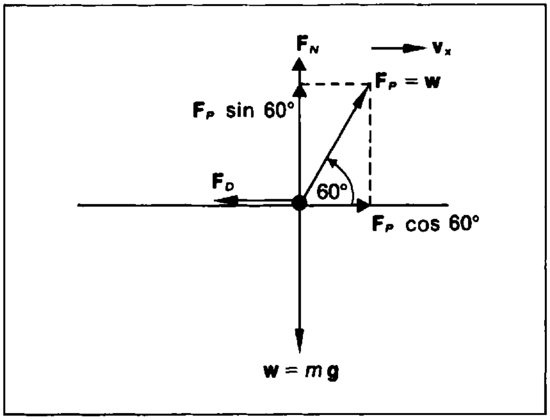


Figure 17-2 Force diagram for Example 2.

and

$$F_N = w - \frac{w\sqrt{3}}{2} = \left(1 - \frac{\sqrt{3}}{2}\right)w$$

$$\approx 0.13w.$$

As these examples illustrate, the analysis of a system in equilibrium can be described as a bookkeeping process in that there must be a complete balance of forces. In cases involving many forces, the process is made much more systematic if the force components in the x-, y-, and z-directions are entered in separate columns of a tabular form. Summing the entries in each column must produce a result of zero. The advantages of the tabular form are that it minimizes the chance of omitting one or more components and it facilitates checking the arithmetic work to eliminate errors.

17-2 NON-EQUILIBRIUM MOTION

From the discussion of the previous section, the reader may already have correctly concluded that non-equilibrium motion occurs when the vector sum of all forces acting on the particle does *not* equal zero. In this situation, Eq. (16-1) is used to determine the acceleration, from which the velocity and displacement as a function of time can be found as outlined in Sections 14-3 and 14-4. Alternatively, one can use a knowledge of the acceleration to determine further details of the system of forces acting on the particle.

Example 3. A particle of mass m is released from rest to slide down a smooth incline that is elevated at an angle θ above the horizontal. It is subject to a downward gravitational force mg and a normal force F_N directed perpendicularly upward from the surface of the incline. Determine

(a) the magnitude of F_N and

(b) the acceleration in terms of mg and θ.

SOLUTION

Since the motion occurs along the surface of the incline, the normal force F_N must be equal in magnitude but oppositely directed to the component of the weight acting normal to the incline surface. Thus (see Figure 17-3),

$$F_N - w \cos\theta = 0$$

or

$$F_N = mg \cos\theta.$$

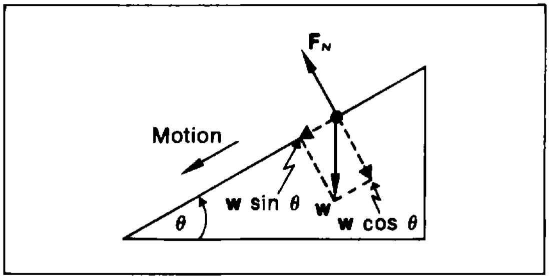


Figure 17-3 Force diagram for Example 3.

The component of the weight acting in a direction parallel to the incline surface is the only other force. Therefore, the acceleration must be directed along the incline in the same direction. It is given by

$$mg \sin \theta = ma$$

or

$$a = g \sin \theta.$$

This example provides an opportunity to illustrate one method of checking the correctness of our solution. Thus, consider the extreme situations (i) $\theta = 0°$, and (ii) $\theta = 90°$. In the first case, the incline would become horizontal, F_N would equal mg, and a would become zero (static equilibrium). In the second case, the incline becomes vertical, F_N would become zero, and a would become equal to g. The reader can check that the equations for F_N and a yield these results, and can therefore be expected to apply as well to intermediate values of θ. Thus, by considering special cases for which we know the answer, we check the general solution.

In obtaining our solution to this problem, the physical requirement that the motion had to take place along the incline was used to simplify the analysis. In effect, a new set of axes parallel and perpendicular to the incline were used in place of the usual x- and y-axes. Let us now analyze the problem in terms of x- and y-axes to demonstrate that the results are the same. As Figure 17-4 illustrates, we can write the component parts of Eq. (16-1) as follows:

$$F_N \sin \theta = ma_x = ma \cos \theta, \qquad \text{(i)}$$

$$mg - F_N \cos \theta = ma_y = ma \sin \theta. \qquad \text{(ii)}$$

From (i), we obtain

$$F_N \tan \theta = ma. \qquad \text{(iii)}$$

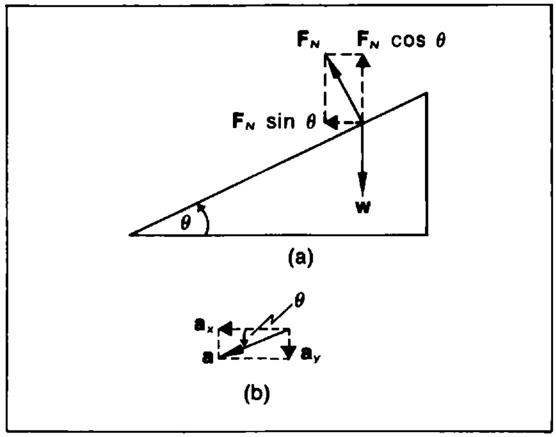


Figure 17-4 Force diagram for Example 3 in x-y coordinates.

Combining (ii) and (iii),

$$mg - F_N \cos \theta = F_N \tan \theta \sin \theta,$$

$$mg \cos \theta = F_N(\cos^2 \theta + \sin^2 \theta)$$

or

$$F_N = mg \cos \theta,$$

so that from (iii)

$$a = \frac{1}{m}(mg \cos \theta)\tan \theta = g \sin \theta,$$

as before.

Example 4. A particle of mass $m = 1$ kg is observed to move up a smooth incline with an acceleration of 4.9 m/sec^2 as a result of a pulling force F_P applied at an angle of 37° above the plane of the incline (which is oriented at an angle of 30° above the horizontal).

(a) Find F_P.
(b) Find F_N.

SOLUTION

As in Example 3, it is easier to consider motion parallel and perpendicular to the plane of the incline. Therefore, we resolve $\mathbf{F}_P$ and $\mathbf{w}$ into components in those directions. Equation (16-1) then gives:

parallel components

$$F_P \cos 37° - w \sin 30° = ma = \frac{w}{g}a;$$

perpendicular components

$$F_N + F_P \sin 37° - w \cos 30° = 0.$$

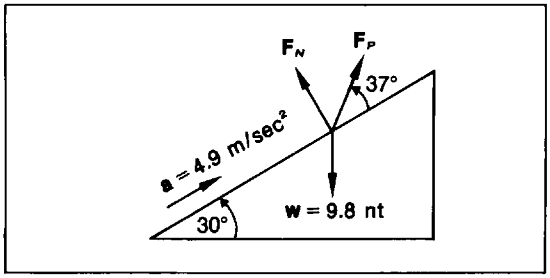


Figure 17-5 Force diagram for Example 4.

Therefore,

$$F_P = \frac{w\left(\sin 30° + \frac{a}{g}\right)}{\cos 37°}$$

and

$$F_N = w\left[\cos 30° - \left(\sin 30° + \frac{a}{g}\right)\tan 37°\right].$$

Using the numerical values given above, we find that

$$F_P = 12.2 \text{ nt}$$

and

$$F_N = 1.14 \text{ nt}.$$

The reader should verify that the parallel and perpendicular force components were correctly determined.

17-3 NEWTON'S LAWS OF MOTION

In analyzing the examples in the last two sections, we have made use of a limited number of basic principles. First, Eq. (16-1) proved to be adequate as a cause and effect relationship for both equilibrium ($a = 0$) and non-equilibrium ($a \neq 0$) situations. A less obvious relationship was employed when we identified the normal force (F_N) exerted by a constraining surface in the various examples. A normal force exists when a particle of mass m is in contact with the surface. As a result of the contact, the particle exerts a net force on the surface due to other forces (gravitational, pulling, etc.) acting upon it. If the particle is to remain on the surface (rather than sinking into it or rising above it), a force directed outward from the surface of the correct magnitude must exist. In Section 16-5, for example, the object resting on the scales exerted a force of 10 lb downward on the scales (because of its weight). Since the object remained at rest on the scales, the normal force exerted on it by the scales is an upward force of 10 lb.

In the language of physics, the normal force exerted by the surface on the object is said to be in reaction to the net force normal to the surface exerted on it by the object. If the surface were to vanish, so also would the normal force and the constrained motion associated with it. In the next section, we will discuss at greater length the concept of reaction forces (and the action forces that oppose them). The remainder of this section is devoted to a brief historical discussion of the above relationships.

Galileo (1564–1642) found flaws in the mechanics of Aristotle. His experiments involving moving bodies led to the idea that a state of motion can exist without an applied force, while a change in the state of motion can only occur as a result of an applied force. It remained for Newton (1642–1727) to formulate a complete basis for analyzing problems of classical mechanics, the law of universal gravitation, as well as his contributions to optics and mathematics.

In 1687, Newton published his *Principia*, or *Mathematical Principles of Natural Philosophy*. The first two parts of the book deal with the motions of bodies by means of propositions, theorems, and corollaries. In addition, these parts contain his development of the principle of universal gravitation. The third part of the book is devoted to the solar system as an example of the applicability of the principles presented earlier. His description of mechanics is summarized in three relations which have acquired the status of "laws" of motion. In formulating these laws, it was necessary for him to clarify the meaning of mass and force, which we have discussed briefly in Chapter 16.

Newton's three laws of motion can be stated (in modern terminology) as follows:

I Every object continues in a state of rest or of uniform motion in a straight line unless acted upon by a net applied force.

II The product of the acceleration and the mass of an object is proportional to the net force applied to the object.

III When object A exerts a force on object B, object B exerts on object A a reaction force which is equal in magnitude but opposite in direction.

Some comments are in order. First, Eq. (16-1) expresses in equation form the statements of law I ($a = 0$) and law II ($a \neq 0$). In addition, however, one must recognize that a frame of reference (see Section 14-5) must be specified if accelerations are to be unambiguously identified. A frame of reference with respect to which a particle is not accelerated is called an inertial reference frame, and law I is also known as the law of inertia. Finally, it should be noted that Newton's original statement of law II could also be expressed symbolically in the form

$$F = \frac{d(mv)}{dt}. \tag{17-1}$$

If the mass m is not a function of time, Eq. (17-1) is identical to Eq. (16-1). On the other hand, if m does depend upon time, the two equations will not give the same result. We will return to this question in Chapter 19, where it will be seen that Eq. (17-1) is the most generally satisfactory form of law II.

17-4 ACTION AND REACTION

Law III can be interpreted as a statement that a single isolated force cannot exist. Whenever one object experiences a force, it does so because of the existence of another object which must necessarily be subject to a force in reaction to its action on the first. Furthermore, one should understand that the two objects need not be in contact, as we assumed for the object on the scales. For example, the law of universal gravitation [Eq. (16-5)] applies to any two particles at any separation. It states that each exerts on the other a force of the same magnitude but opposite direction. (Thus a self-centered individual could describe his weight as a force equal in magnitude but opposite in direction to the force acting on the Earth due to his mass.)

It is important to stress that the action-reaction pair of forces act on separate objects and not on the same one. As a result when one applies law II to a given object, only one of the action-reaction pairs will be included in the vector sum of forces applied to the object. If this were not true, one would never have a non-equilibrium situation. A common paradox used to emphasize this fact is the following: when a pitched baseball is hit by a bat swung by a batter, it is observed to move away from the bat. However, law III says the bat exerts a force *on* the ball equal in magnitude and opposite in direction to the force it experiences due to the ball. Therefore, since the vector sum of these two forces vanishes, a change of motion cannot occur. The paradox is easily resolved by recognizing that whether or not the motion of the ball changes depends only upon the forces acting upon it and *not* upon any reaction forces it exerts on other objects.

As a final example emphasizing this point, consider once more the 10 lb object at rest on the scales. This system involves four distinct forces—the pull of the Earth on the object, the pull of the object on the Earth, the force of the object on the scales, and the force of the scales on the object. All of these forces have the same magnitude but the direction of the first and third forces are toward the Earth, while the second and fourth are in the opposite direction. In applying law I to the object (since $a = 0$), only the first and fourth forces are involved since the other two forces are exerted *on* objects other than the 10 lb object in question.

In analyzing situations involving more than one particle, it is useful to schematically subdivide the problem into separate parts (called free-body diagrams) and indicate all forces, whether action-reaction or not. In this way, it becomes possible to determine missing details of the total system.

Example 5. Figure 17-6 shows a system of two particles whose masses are m_1 and m_2, respectively. They are in contact on a smooth horizontal surface, and have an acceleration **a** together to the right due to an applied force **F**. Find the magnitude of the force $\mathbf{F}_{21}$ that the first mass exerts on the second.

SOLUTION

Figure 17-7 shows the free-body diagrams for the system, with all the forces indicated. Since the motion is along the horizontal surface $F_{N_1} = m_1 g$, $F_{N_2} = m_2 g$. Law II for the horizontal motion of 1 and 2 yields the following expressions:

$$F - F_{12} = m_1 a \tag{i}$$

$$F_{21} = m_2 a \tag{ii}$$

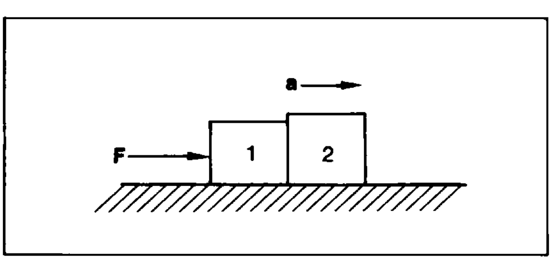


Figure 17-6 System diagram for Example 5.

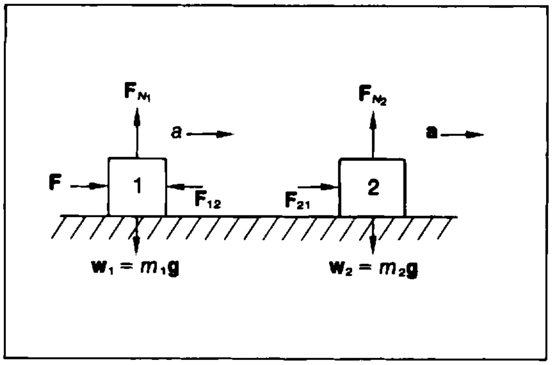


Figure 17-7 Free-body diagrams for Example 5.

Law III requires that $F_{12} = -F_{21}$ (equal magnitudes but opposite directions, as shown). Therefore, adding Eq. (i) and Eq. (ii) gives

$$F = (m_1 + m_2)a$$

or

$$a = \frac{F}{m_1 + m_2}. \qquad \text{(iii)}$$

Substituting Eq. (iii) in Eq. (ii) gives the result

$$F_{12} = F_{21} = \frac{m_2 F}{m_1 + m_2}. \qquad \text{(iv)}$$

The reader can show by an identical argument that if the two masses are interchanged (see Figure 17-8), then

$$F_{12} = \frac{m_1 F}{m_1 + m_2}.$$

Is this last result consistent with Eq. (iv)?

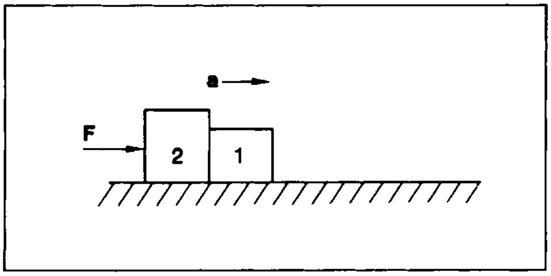


Figure 17-8 System diagram of Example 5 with masses reversed.

PROBLEMS

1. Figure 17-9 shows a body of weight W supported at the midpoint of a wire of length l. The forces at point B exerted along and by the wire segments AB and BC (such forces are called tension forces) are opposed by the weight W.
   (a) Show that the tension forces $\mathbf{T}_1$ and $\mathbf{T}_2$ in the wire segments are equal in magnitude.
   (b) Find the magnitude of tension force T in terms of the weight W and the angle ϕ.
   (c) For what angle ϕ is the magnitude of the tension force T equal to the weight W?

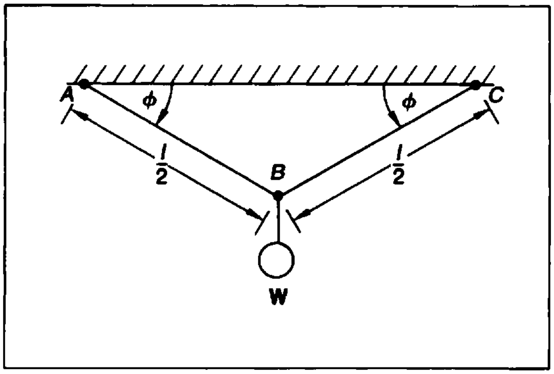


Figure 17-9

2. In Figure 17-10, $\phi_1 = 60°$, $\phi_2 = 30°$, and $W = 200$ lb. Find the magnitudes of the tension forces T_1 and T_2. In this case, the wire lengths AB and BC are unequal.

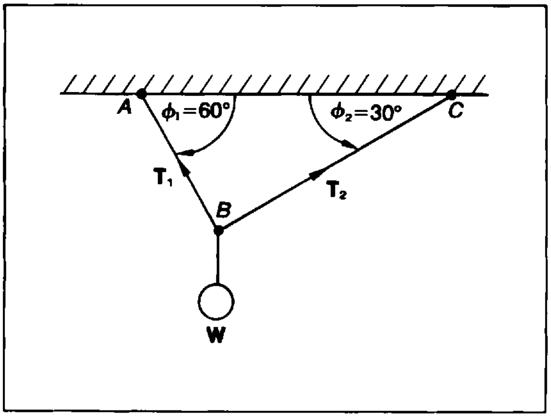


Figure 17-10

3. A 3 kg block resting on a smooth horizontal surface is subjected to a force of 8 nt.
   (a) What is the acceleration of the block?
   (b) If the horizontal force on the block is supplied by hanging a weight of 8 nt over a pulley and releasing the system, what will be the acceleration of the block and the tension T in the connecting cord? See Figure 17-11.

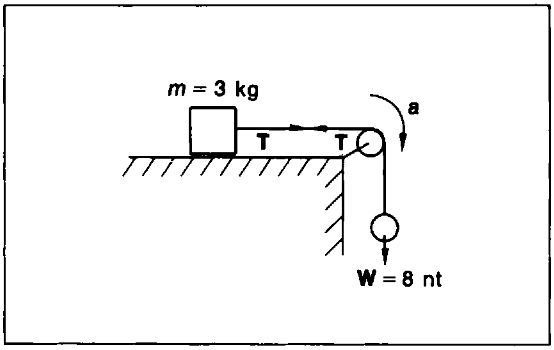


Figure 17-11

4. What average force is required to stop a 224 lb fullback in a distance of 3 ft if his initial speed is 24 ft/sec?

5. An automobile weighing 4000 lb and traveling 60 mi/hr is to be brought to rest in 300 ft. What average force must be exerted on the car?

6. A rope fastened to a 60 lb block that is on a smooth plane inclined at 30° with the horizontal extends upward parallel to the plane and over a pulley at the top of the plane. A 70 lb ball is hung on the end of the rope vertically below the pulley. Find:
   (a) the acceleration of the block and
   (b) the tension in the rope.

7. A 1920 gm block rests on a frictionless table. A cord attached to the block passes over a pulley at the edge of the table. A 40 gm block is attached to the hanging end of the cord. Calculate:
   (a) the acceleration of the blocks and
   (b) the tension in the cord.

8. The blocks and the ball have the weights shown in Figure 17-12. The plane is smooth. Find:
   (a) the acceleration of the ball and
   (b) the magnitudes of the tension forces T_1 and T_2 in the connecting ropes.

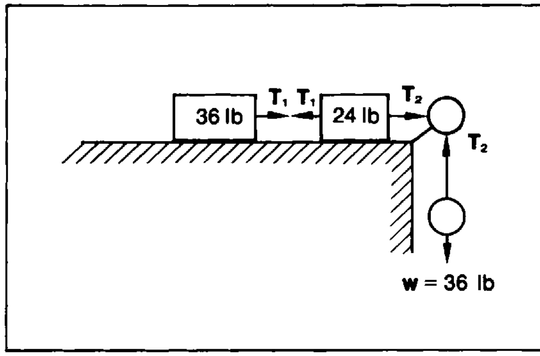


Figure 17-12

9. Atwood's machine consists of two particles of masses m_1 and m_2 connected by a light string, which is passed over a light, smooth pulley as shown in Figure 17-13. The magnitude of the tension T in the string is the same on either side of the pulley. By applying Newton's laws of motion, show that the acceleration of the masses is given by

$$a = (m_1 - m_2)g/(m_1 + m_2).$$

   Provide a physical interpretation of this result for the cases
   (a) $m_1 > m_2$;
   (b) $m_1 < m_2$;
   (c) $m_1 = m_2$.

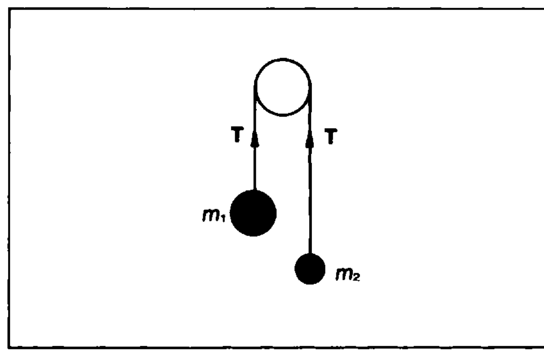


Figure 17-13

10. Show that the tension in the string of an Atwood's machine has a magnitude given by the expression

$$T = \left(\frac{2m_1m_2}{m_1 + m_2}\right)g.$$

    Provide a physical interpretation of this result for the special case $m_1 = m_2$.
11. In an Atwood's machine experiment, the acceleration of the system is measured to be 0.49 m/sec^2. The lighter particle has a mass of 1.9 kg.
    (a) What is the magnitude of the tension T?
    (b) What is the mass of the heavier particle?
12. A 0.5 kg block is allowed to slide down a rough inclined plane set at 30° to the horizontal. It starts from rest and accelerates uniformly, traveling 2 m in 4 seconds. Find the magnitude of the constant retarding force exerted on the block due to the rough surface of the incline. This force acts in a direction parallel to the incline surface.
13. A man weighing 160 lb enters an elevator, stands on a bathroom scale, and is accelerated upward at 2 ft/sec^2.
    (a) What weight is indicated by the scale in this upward accelerated motion?
    (b) What would be the reading of the scale if the acceleration were 2 ft/sec^2 downward?
    (c) Suppose the elevator cable snaps while the car is in motion. What would the scale reading then be?

18 Applications of the Laws of Motion

18-1 PROBLEM SOLVING—A PLAN OF ATTACK

If the motion of a particle (or system of particles) is to be analyzed in terms of Newton's laws of motion, there are several aspects of the problem to be dealt with. First, one must be sure to include all the forces that can have a measurable effect on the particle. Furthermore, the nature of each force must be understood at least well enough to assign approximate values to the various parameters upon which the force depends. For example, the effect of a gravitational force (due to one particle) that is acting on another particle cannot be fully evaluated unless the masses of the two particles and their separation distance are known. It is one of the aesthetically pleasing aspects of physics that most of the important force laws are simple in form however powerful their consequences.

Second, an inertial reference frame must be chosen that is suitable for the analysis of the particle motion. In a non-inertial frame of reference, the laws of motion will not give satisfactory results unless one or more "fictitious" forces are introduced in the analysis. These forces have the effect of "transforming" the problem from the non-inertial reference frame to an inertial frame. For example, one can generally regard a frame of reference attached to a stationary laboratory table as an inertial reference frame for particle motion on the table top. However, for a laboratory on the surface of the Earth, the table is in fact both rotating about the axis of rotation of the Earth and traveling in a nearly circular orbit about the Sun (see Section 18-5). Precise measurements would show that the resultant acceleration of the table leads to small discrepancies between the observed particle motion and the predictions obtained from the laws of motion assuming an inertial frame of reference. In Section 18-4, we will consider the effect of the Earth's rotation upon the weight of an object on the surface of the Earth. Except for this example, we will not consider the complications of motion relative to non-inertial frames of reference.

The remaining facets of the analysis required in applying the laws of motion have to do with what can be called procedural details, in contrast to the conceptual or physical aspects discussed thus far. As noted in Section 17-1, when the analysis has been systematized to resemble a problem in bookkeeping, one can have confidence that few errors will arise. Those which do occur will not be difficult to locate if the desired systemization has been accomplished. The following suggestions should be helpful in systematizing the analysis of any problem.

1. Determine all the non-negligible details of the problem. In reading a "textbook" problem, it is useful to sketch in a diagram the essential details of the problem as they are encountered in reading. Show a symbol for every quantity involved, and, if numerical values are given, indicate them as well. Be sure that all similar quantities have common units—do not attempt to add pounds of force to newtons of force.
2. Determine what physical laws (gravitation, for example) are necessary in addition to the laws of motion. Substitute in these expressions the symbols of the relevant quantities from the sketch of step 1.
3. Separate the vector equations where necessary into algebraic component equations. Solve these equations systematically for the desired quantities *in symbol form.*
4. Check to see that the resulting solution equations are dimensionally correct *before* performing any numerical work.
5. After the calculations are complete, try to determine whether or not the results are reasonable. This is a value judgment that becomes easier to make with experience.

18-2 FRICTION

It is common experience that the external force required to cause (or maintain) the motion of an object kept in contact with a solid surface depends upon surface characteristics such as degree of cleanliness and smoothness. Thus, a rough surface will offer a relatively great resistance to a change in the state of motion of an object resting upon it. The same surface if smoothed or lubricated will become much easier to negotiate. One attributes this decreased resistance to motion to a lessening of what are called *frictional forces* between two bodies in contact. It is also observed that a less massive body on a given surface will require a smaller force to overcome friction forces than will a more massive body on the same surface. Once a body has been set in motion on a surface, it is usually observed that the force required to maintain motion at constant speed is markedly less than the force required to set it in motion initially. Within rather wide limits, this force is independent of the speed of the object. Finally, it is found that the force required to overcome friction forces is relatively insensitive to the magnitude of the area of contact between the object and the surface.

To attempt a detailed analysis of friction forces is out of the question at this point, for it necessarily involves the interaction of atomic or at least microscopic portions of the objects involved. Such interactions are extremely sensitive to many factors such as crystal structure, impurities or imperfections in the structure, and surface contaminations to name a few. While the effects of these factors could in principle be calculated with modern quantum mechanical techniques, it is experimentally not possible to obtain sufficiently accurate values of the necessary physical parameters, although much progress has been made.†

It is possible, however, to satisfactorily account for the effect of frictional forces by introducing empirical relationships based on the observations cited above. In doing so, we stress the fact that these relationships must *not* be regarded as fundamental physical laws like Newton's laws of motion. They are valid *only* for limited ranges of the physical parameters involved, and their use is justified only by the realization that without them it would be difficult to obtain even approximate solutions for many common particle motion problems. One need not apologize for the use of empirical relations, provided care is exercised to avoid their use outside their range of validity.

We proceed by identifying the important parameters involved in a particularly simple situation, illustrated in Figure 18-1. A block of mass m is at

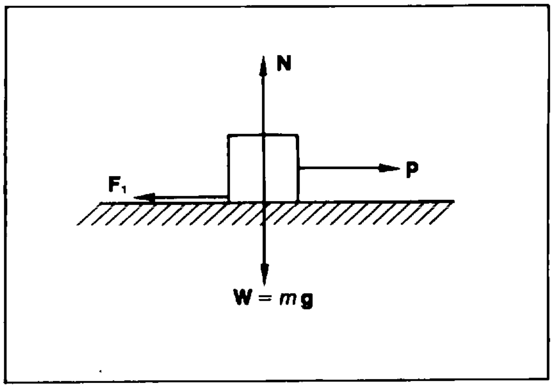


Figure 18-1 Force diagram for a system with friction.

†See, for example, the article, "Friction," by R. Palmer in the February 1951 issue of *Scientific American.*

rest upon a horizontal surface, subject to four different forces. The weight of the block ($\mathbf{W} = m\mathbf{g}$) acts downward, pressing the block into contact with the surface. Since the block is at rest on the surface, it follows that there must be a force exerted upward on the block by the surface which has the same magnitude. Since this force is perpendicular (or normal) to the surfaces that are in contact, it is common to identify this as the normal force (**N**). The force designated (**P**) represents a pulling force, and the frictional force due to the contact of the block with the surface is labeled ($\mathbf{F}_f$). Since the block is at rest, the magnitude of (**P**) must equal that of ($\mathbf{F}_f$). Experimentally, one finds that **P** can be increased in magnitude from zero up to a maximum value. Exceeding that maximum value will produce motion of the block.

This indicates that, for a static situation (no motion occurs), the friction force ($\mathbf{F}_f$) can assume values ranging from zero to a maximum value. This maximum value is related empirically to the magnitude of the normal force (**N**) by the relation

$$|\mathbf{F}_{f\max}| = \mu_s |\mathbf{N}|, \qquad (18\text{-}1)$$

where μ_s is a dimensionless constant called the *static coefficient of friction*. This relation just expresses the observed fact that as the force of contact (**N**) between the surface and the block is increased, the friction force will increase in direct proportion. This relationship holds true almost independently of the area of contact. It does not remain valid for a heavy object whose area of contact is made to approach zero—for example, by grinding the bottom to a (balanced) point surface. The reason it fails in such a case is that when the entire weight of the object is applied to such a small region of the surface, a local deformation of the surface can occur. Motion in such a case would then be similar to pulling a plow through soil rather than along the surface of the soil. An example of this type of deformation is the unfortunate creation of dents in tile (and even hardwood or aluminum airplane floors) caused by women wearing spike-heeled shoes.

To summarize the results of this discussion involving Figure 18-1, the block will remain at rest so long as the pulling force (**P**) has a magnitude given by the relation

$$|\mathbf{P}| \leq \mu_s |\mathbf{N}| = \mu_s m |\mathbf{g}|. \qquad (18\text{-}2)$$

Equation (18-2) is valid because $|\mathbf{P}| = |\mathbf{F}_f|$ and $|\mathbf{N}| = m|\mathbf{g}|$. When $|\mathbf{P}| = \mu_s |\mathbf{N}|$, one is to infer that motion impends, that is, we are to regard $|\mathbf{P}|$ as being infinitesimally close to, but less than $\mu_s |\mathbf{N}|$. Once motion has begun, it is usually observed that the pulling force required to maintain motion of the block *at constant speed* along the surface can be accounted for empirically by the relation

$$|\mathbf{P}| = \mu_K |\mathbf{N}|, \qquad (18\text{-}3)$$

where μ_K is the *kinetic coefficient of friction*. In general, $\mu_K < \mu_s$, which can be understood qualitatively by referring to Figure 18-2, which is a representation of an enlarged portion of two surfaces in contact. In general, an apparently smooth surface is microscopically very irregular so that two such surfaces are not in contact everywhere but only at irregularly located points such as *A*, *B*, and *C*. At such points, the interatomic forces of attraction are quite large and can be thought of as atomic size "spot welds." To break these "spot welds" so that relative motion can begin, requires relatively large forces. Once motion is begun, the protruding portions of the two surfaces move relative to each other, which tends to reduce the strength of the "spot welds." Thus, once motion begins, a smaller pulling force will be required to maintain relative motion of the surfaces at a constant speed. As indicated earlier, it is observed that within rather wide limits the force required to maintain motion is independent of the speed. Thus, Eq. (18-3) requires the additional relation

$$|\mathbf{P}| = |\mathbf{F}_f| = \mu_K |\mathbf{N}|. \qquad (18\text{-}4)$$

We conclude this section with an example intended to emphasize that, while **N** is perpendicular to the surfaces of contact and $\mathbf{F}_f$ acts tangent to the surfaces of contact and in a direction opposite to any actual or impending motion, it is *not* always true that $|\mathbf{N}| = m|\mathbf{g}|$ and $|\mathbf{F}_f| = |\mathbf{P}|$.

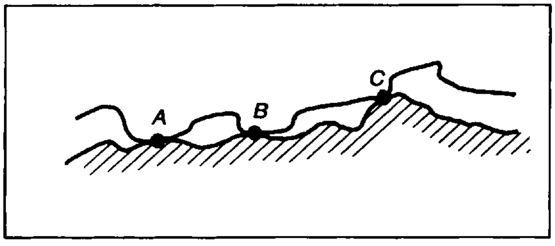


Figure 18-2 Schematic enlargement of two surfaces in contact.

Example 1. A block of mass m and weight mg is held against a vertical plane surface due to a horizontal force $\mathbf{P}$ as shown in Figure 18-3. The static and kinetic coefficients of friction are μ_s and μ_K, respectively.

(a) Determine the magnitude of $|\mathbf{P}|$ if downward motion of the block impends.

(b) Determine the magnitude of $|\mathbf{P}|$ if downward motion at constant speed is taking place.

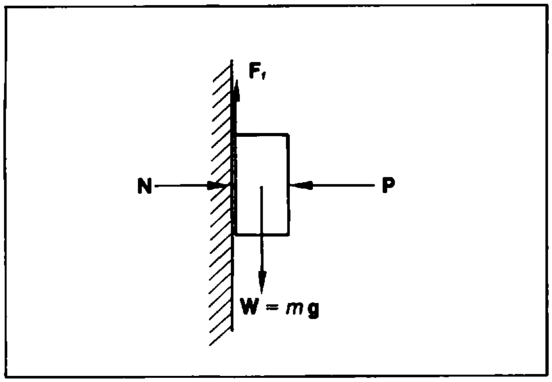


Figure 18-3 Force diagram for Example 1.

SOLUTION

(a) As indicated in the figure, in this case $\mathbf{N}+\mathbf{P}=0$ and $\mathbf{F}_f+m\mathbf{g}=0$, or $P=N$, $F_f=mg$. Since $F_f=\mu_s N$, it follows that $mg=\mu_s P$ or $P=mg/\mu_s$.

(b) In this case, the vertical and horizontal forces are still balanced (since the speed is constant, $\mathbf{a}=0$). Therefore, we obtain by the analysis used in (a)

$$P=\frac{mg}{\mu_K}.$$

In Table 18-1, we list some values for μ_s and μ_K for various surfaces. Since friction forces depend markedly on microscopic details of both surfaces (surface cleanliness, crystalline imperfection, etc.), these values should be considered as typical values only.

Table 18-1 Coefficients of Friction (dry surfaces).

Substances	μ_s	μ_K
Teflon on teflon	0.04	0.04
Copper on steel	0.50	0.35
Aluminum on steel	0.60	0.50
Lead on steel	0.95	0.95
Copper on cast iron	1.05	0.30

18-3 PROJECTILE MOTION

There are many familiar examples of the type of motion known as projectile motion. A projectile fired from some type of cannon, a football put in motion by a kicker, and a golf ball sent in flight by a golf club all can be described by the same physical characteristics with only minor variations in details. Figure 18-4 illustrates the principal features: at the origin, a particle of mass m has an initial velocity of magnitude v_0 directed at an angle θ_0 relative to the horizontal, brought about by the action of some force (explosive propellant, muscular effort of the kicker, or the force transmitted by the driven golf club for the examples cited). We now limit our discussion to low speeds and ignore the rotation of the Earth and air resistance. In this situation, the only force that acts upon the projectile after launch is the downward force of (or due to) gravity. Newton's second law for this situation, in vector component form is, from $\Sigma\mathbf{F}=m\mathbf{a}$,

$$-mg\mathbf{j}=m(a_x\mathbf{i}+a_y\mathbf{j}). \qquad (18\text{-}5)$$

Equating components on both sides of this equation gives

$$a_x=\frac{dv_x}{dt}=\frac{d^2x}{dt^2}=0, \qquad (18\text{-}6)$$

$$a_y=\frac{dv_y}{dt}=\frac{d^2y}{dt^2}=-g. \qquad (18\text{-}7)$$

From Eq. (18-6), it follows that $v_x=$ constant. Now from Figure 18-4, $v_{0x}=v_0\cos\theta_0$ is the initial x-component of the velocity. In the absence of any forces in the x-direction, we see that $v_x=v_{0x}=v_0\cos\theta_0$. In the y-direction, Eq. (18-7) requires that (refer to Section 14-4)

$$\int_{v_{0y}}^{v_y} dv_y=-g\int_0^t dt$$

or

$$v_y=v_{0y}-gt. \qquad (18\text{-}8)$$

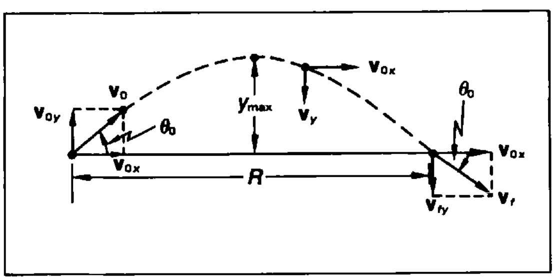


Figure 18-4 Schematic diagram of projectile motion.

Since

$$v_{0y} = v_0 \sin \theta_0,$$

we have

$$v_y = v_0 \sin \theta_0 - gt. \qquad (18\text{-}9)$$

Similarly, we can find the x- and y- components of the displacement by noting that

$$v_x = \frac{dx}{dt} = v_0 \cos \theta_0$$

or

$$\int_0^x dx = v_0 \cos \theta_0 \int_0^t dt,$$

$$x = v_0 \cos \theta_0 t; \qquad (18\text{-}10)$$

and that

$$v_y = \frac{dy}{dt} = v_0 \sin \theta_0 - gt,$$

$$\int_0^y dy = \int_0^t (v_0 \sin \theta_0 - gt) dt$$

or

$$y = v_0 \sin \theta_0 t - \frac{1}{2} g t^2. \qquad (18\text{-}11)$$

What is the physical content of our mathematical manipulations? First, the absence of forces in the x-direction yields a constant velocity component in the x-direction. As a result, the x-component of the displacement is a linear function of the time. In the y-direction, the acceleration is constant because the force (due to gravity) in the y-direction is constant. As a result, the y-velocity-component is a linear function of time; the y-component of the displacement, therefore, is a quadratic function of time. We see further that the application of Newton's second law for our projectile motion example leads to a separation of the x- and y-components of the motion. (The extension of the problem to include 3 components adds no further complications.)

Let us turn now to the questions:

1. What is the maximum height (y_{max}) attained by the projectile and how long does it take to attain it?
2. What is the range (R), the surface-to-surface horizontal distance traveled by the projectile, and the time of flight (t_{total})?
3. What is the shape of the trajectory (the path traversed by the projectile)?

It is clear, first of all, that v_y must be equal to zero when the particle is at the maximum height. (If it were not, either the particle would still be moving upward or would be in motion to a lower position. For either possibility, the y-displacement could not be y_{max}.) Therefore, Eq. (18-8) gives

$$0 = v_0 \sin \theta_0 - g t_{up},$$

where t_{up} is the time from launch ($t = 0$) to the attainment of y_{max}. Thus,

$$t_{up} = \frac{v_0 \sin \theta_0}{g}. \qquad (18\text{-}12)$$

Substitution of $t = t_{up}$ from Eq. (18-12) and $y = y_{max}$ into Eq. (18-11) yields

$$y_{max} = \frac{v_0^2 \sin^2 \theta_0}{2g}. \qquad (18\text{-}13)$$

From Eq. (18-13), we can deduce (as might seem obvious from experience) that a maximum height for given v_0 and g will occur for a projectile fired vertically ($\theta_0 = 90°$).

To determine the range R, we first note that when $t = t_{total}$, $y = 0$ (the projectile has returned to the launch elevation). Thus, Eq. (18-11) requires that

$$0 = v_0 \sin \theta_0 t - \frac{1}{2} g t^2$$

or

$$0 = \left(v_0 \sin \theta_0 - \frac{gt}{2} \right) t. \qquad (18\text{-}14)$$

Equation (18-14) has two solutions: (a) $t = 0$, the initial situation; and (b) $t = t_{total} = 2v_0 \sin \theta_0 / g$, the desired time of flight. Notice that t_{total} is exactly twice the time required to attain y_{max}. This tells us that the flight is symmetric; that is, it takes as long to go up as it does to come down when one considers only the vertical force due to gravity. Substituting the value for t_{total} in Eq. (18-10) for the x-motion, we get

$$R = \frac{2v_0^2 \sin \theta_0 \cos \theta_0}{g} \qquad (18\text{-}15)$$

or

$$R = \frac{v_0^2 \sin 2\theta_0}{g}. \qquad (18\text{-}16)$$

For fixed values of v_0 and g, the maximum value of R will be obtained when $\sin 2\theta_0$ attains its maximum value, which is unity. Since this occurs for $2\theta_0 = 90°$, we see that for maximum range the launch angle must be 45°. (The reader who is anxious to test the concept of "max-min" determinations by calculus techniques can arrive at this same result by solving the equation $dR/d\theta_0 = 0$.)

The shape of the trajectory is not difficult to determine. First of all, we already know the path is symmetric about the point $(R/2,\ y_{max})$ since $t_{total} = 2t_{up}$. To determine the analytical equation, we simply eliminate the time parameter from Eqs. (18-10) and (18-11). Thus, Eq. (18-10) gives

$$t = \frac{x}{v_0 \cos \theta_0}. \qquad (18\text{-}17)$$

Substituting this result in Eq. (18-11) yields

$$y = \frac{v_0 \sin \theta_0}{v_0 \cos \theta_0} x - \frac{1}{2} \frac{gx^2}{v_0^2 \cos^2 \theta_0}$$

$$= x \tan \theta_0 - \frac{gx^2}{2v_0^2 \cos^2 \theta_0}. \qquad (18\text{-}18)$$

Since y varies quadratically with x, Eq. (18-18) predicts that the path of the projectile will be parabolic. It is left for the reader to show that Eq. (18-18) predicts a parabola that is symmetric about the point $x = R/2$. (That is, show that when $x = R/2 + \Delta x$ the same value of y is obtained for positive or negative Δx.)

Finally, it should be noted that the analysis of projectile motion carried out here is also valid for such seemingly different situations as skiers leaving the horizontal end of a ski jump or golf balls driven from a tee to a green at a higher elevation.

Example 2. A skier leaves the horizontal end of a ski jump with a speed of 20 m/sec. If the skier lands 40 m out from the end of the ski jump, how far below the end of the ski jump is the landing point?

SOLUTION

Referring to Figure 18-5, it is clear that since $\theta_0 = 0°$, $v_{0x} = v_0$, $v_{0y} = 0$. Therefore, from Eq. (18-17), we find the time of flight to be

$$t_{total} = \frac{x}{v_0} = \frac{40}{20} = 2 \text{ sec.}$$

Using this result in Eq. (18-18) gives the result

$$y = -\frac{1}{2} g \left(\frac{x}{v_0}\right)^2 = -19.6 \text{ m.}$$

(The minus sign indicates a vertical displacement *below* the origin.) Notice that this result also follows from Eq. (18-11) for $\theta_0 = 0°$, as it should.

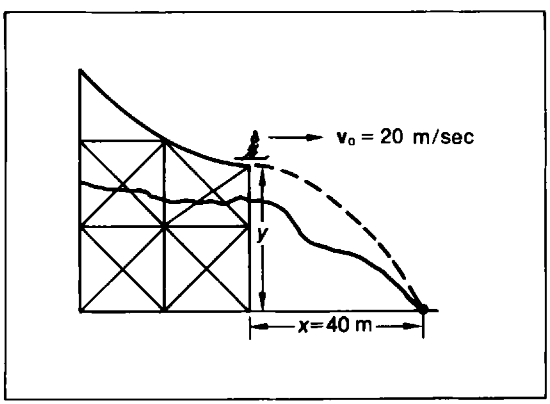


Figure 18-5 System diagram for Example 2.

18-4 UNIFORM CIRCULAR MOTION

A particle in uniform circular motion follows a circular path with a velocity that is constant in magnitude. Its acceleration can have no component tangent to the path since this would cause the magnitude of the velocity to change. The changing direction of the velocity requires an acceleration component perpendicular to the velocity vector (and the circular path)—that is, directed radially inward. Such an acceleration is called centripetal (center-seeking). The force associated with such an acceleration (via the second law) is known as a centripetal force, and it can arise in many ways. For planetary motion discussed in the next section, the force is gravitational in nature, while a whirling stone on a string experiences a centripetal force due to the tension in the string.

Before we consider examples of such systems, let us first determine the kinematical relationship between the centripetal acceleration and the velocity of a particle in uniform circular motion. Figure 18-6(a) shows the particle at two different points along its path corresponding to positions separated by a distance $R\Delta\theta$, which was traversed in a time Δt. The displacement Δs between the two points can be viewed as the base of an isosceles triangle with equal sides of length R. The two velocity vectors v and v' are also of equal length v, and each is perpendicular to a radius vector of length R. Figure 18-6(b) illustrates the vector relation $v + \Delta v = v'$ that the velocities must satisfy. Here again we have an isosceles triangle whose equal sides are each perpendicular to one of the equal sides of the triangle of Figure 18-6(a). As you may recall, the two triangles are said to be similar when such conditions exist—that is, with identical apex angles, which in this case are of magnitude $\Delta\theta$. Further-

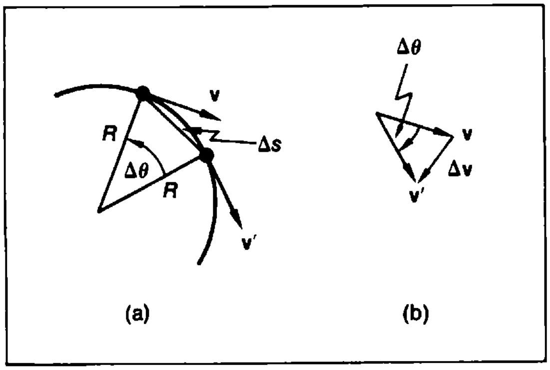


Figure 18-6 Displacement (a) and velocity (b) vector diagrams for uniform circular motion.

more, Δv will be perpendicular to Δs. In terms of the two triangles, therefore, we can write

$$\Delta\theta = \frac{\Delta s}{R} = \frac{\Delta v}{v} \qquad (18\text{-}19)$$

or

$$\Delta v = \frac{v}{R}\Delta s. \qquad (18\text{-}20)$$

Since the change in velocity Δv occurs in a time Δt, the average acceleration $\bar{a}$ is

$$\bar{a} = \frac{\Delta v}{\Delta t} = \frac{v}{R}\frac{\Delta s}{\Delta t}. \qquad (18\text{-}21)$$

In the limit as $\Delta t \to 0$, Δs will become tangent to the path so that

$$\lim_{\Delta t \to 0} \frac{\Delta s}{\Delta t} = \frac{ds}{dt} = v,$$

and $\lim_{\Delta t \to 0} \bar{a} = a$. When $\Delta t \to 0$, Δs becomes tangent to the path so that Δv is directed radially inward as required. We see, therefore, that the expression relating the centripetal acceleration and the velocity in uniform circular motion is

$$a = \frac{v^2}{R}. \qquad (18\text{-}22)$$

In situations involving non-uniform circular motion, it is rather common to find the centripetal acceleration denoted by a_R, while the component tangent to the path (which is responsible for the change in magnitude of v) is given by a_T—the tangential acceleration.

From the discussion above, we conclude that for a particle in uniform circular motion Newton's second law takes the form

$$\Sigma \mathbf{F} = m\mathbf{a}_R. \qquad (18\text{-}23)$$

The physical significance of Eq. (18-23) is that regardless of the nature or number of forces applied to the particle of mass m, if it is in uniform circular motion, the vector sum of these forces must be directed radially inward.

Example 3. A boy fastens a small stone of mass m to a light string of length l, raises it over his head, and sets the stone in motion so that it travels a horizontal circular path of radius r with a velocity having a constant magnitude. Determine the tension T in the string and the magnitude of v in terms of m, g, l, and r. (See Figure 18-7.)

SOLUTION

As indicated in Figure 18-7, we first resolve the tension into vertical and horizontal components. Thus,

$$T_x = T\sin\theta \qquad (18\text{-}24)$$

and

$$T_y = T\cos\theta. \qquad (18\text{-}25)$$

Applying the second law to the vertical forces, we have equilibrium (no vertical motion)

$$T_y - mg = 0. \qquad (18\text{-}26)$$

For the horizontal motion, we obtain

$$T_x = ma_r = \frac{mv^2}{r}. \qquad (18\text{-}27)$$

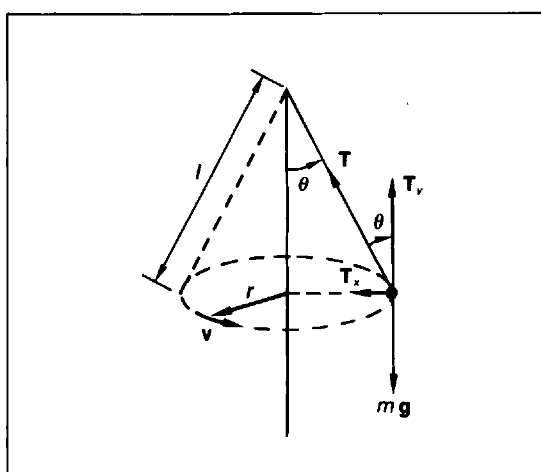


Figure 18-7 Force diagram for Example 3.

Combining these last four relations yields

$$\frac{T_x}{T_y} = \tan\theta = \frac{\frac{mv^2}{r}}{mg} = \frac{v^2}{gr}.$$

Since we can also write

$$\tan\theta = \frac{r}{\sqrt{l^2 - r^2}},$$

it follows that

$$v^2 = \frac{gr^2}{\sqrt{l^2 - r^2}}. \qquad (18\text{-}28)$$

By combining Eqs. (18-24) and (18-27), we also find

$$T = \frac{mgl}{\sqrt{l^2 - r^2}} = \frac{mg}{\sqrt{1 - \left(\frac{r}{l}\right)^2}}. \qquad (18\text{-}29)$$

Notice that our solution [Eq. (18-29)] indicates that $T \to \infty$ as $r \to l$. This means that it is not physically possible to swing the stone so that the string is horizontal. If it were possible, Eq. (18-26) would have to become

$$-mg = 0,$$

because there would now be no vertical component of the tension to counteract the force due to gravity on the stone. Alternatively, Eq. (18-28) would require that $v \to \infty$ as $r \to l$, which is also not possible.

As another example of uniform circular motion, we now consider an object supported by a spring scale at some point on the surface of the Earth. The problem is to determine W, the reading of the spring scale that will be the effective weight of the object. At first glance, it would appear reasonable to say that due to equilibrium the reading of the scales will be equal to the product of the mass of the object, and

$$g_E = \frac{GM_E}{R_E^2},$$

the acceleration due to gravity at the Earth's surface when rotation is neglected. However, because the Earth rotates about its axis, a point on the surface will be in uniform circular motion about the axis of rotation. As a result, the object experiences an acceleration toward the axis of rotation. In other words, a point on the surface of the Earth cannot be used as the origin of an inertial reference frame. Instead, we must describe the motion from an inertial frame (with an origin at 0) that is fixed in space and relative to which the Earth rotates about its axis.

Figure 18-8 shows a particle of mass m at rest at point P on the surface of the Earth at an angle θ (the latitude angle) above the Equator. At this location, the magnitude of uniform circular velocity of the particle due to the Earth's rotation will be the circumference of its path $2\pi r$ divided by the time T corresponding to the rotation of the Earth (24 hours or 8.64×10^4 sec). Since $r = R_E \cos\theta$, it follows that

$$v = \frac{2\pi R_E \cos\theta}{T}. \qquad (18\text{-}30)$$

This means that there will be a resultant acceleration directed inward along r toward the axis of rotation given by

$$a_r = \frac{v^2}{r} = \frac{4\pi^2 R_E \cos\theta}{T^2}. \qquad (18\text{-}31)$$

For the Earth, $R_E \approx 6.4 \times 10^6$ m, $T = 8.64 \times 10^4$ sec, so that $a_r \approx 3.34 \times 10^{-2} \cos\theta$ m/sec^2.

Now let us consider the forces acting on the particle. Figure 18-9 illustrates the situation with W' representing the force supplied by the spring. W' can be replaced by component forces $W'_{R_E} = W' \cos\alpha$ and $W'_T = W' \sin\alpha$, respectively. Similarly, the centripetal acceleration directed along r can also be replaced by components along OP and tangent to the surface at P. Thus, the component along OP has a magnitude

$$a_{R_E} = a_r \cos\theta \approx 3.34 \times 10^{-2} \cos^2\theta, \qquad (18\text{-}32)$$

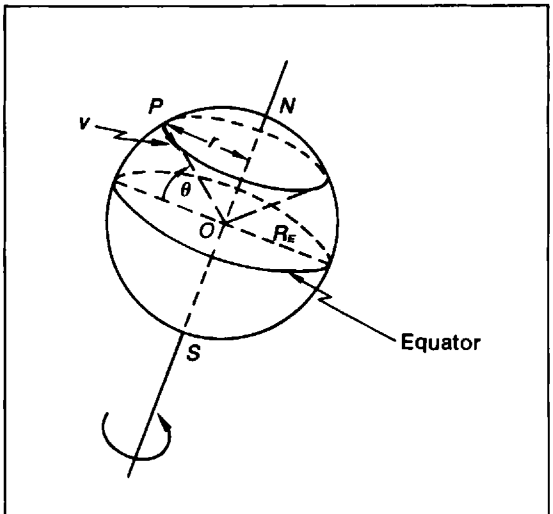


Figure 18-8 Schematic diagram of a particle at rest on a rotating earth.

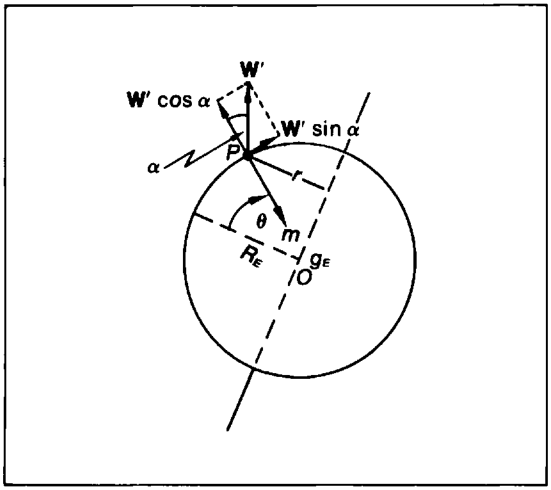


Figure 18-9 Force diagram for the system of Figure 18-8.

and the magnitude of the component tangent to the surface is

$$a_T = a_r \sin\theta \approx 3.34 \times 10^{-2} \cos\theta \sin\theta$$
$$\approx 1.67 \times 10^{-2} \sin 2\theta. \qquad (18\text{-}33)$$

Now we can apply Newton's second law in component form to the radial and tangential parts just evaluated. Thus,

$$W' \cos\alpha - mg_E = -ma_{R_E} = -ma_r \cos\theta \qquad (18\text{-}34)$$

and

$$W' \sin\alpha = ma_T = ma_r \sin\theta. \qquad (18\text{-}35)$$

By eliminating W' from Eqs. (18-35) and (18-36), we can deduce that α must be rather small. Eliminating W' yields

$$\tan\alpha = \frac{a_r \sin\theta}{g_E - a_r \cos\theta}$$

or

$$\tan\alpha = \frac{(a_r/g_E)\sin\theta}{1 - \dfrac{a_r}{g_E}\cos\theta}. \qquad (18\text{-}36)$$

Since

$$\frac{a_r}{g_E} \approx \frac{3.34 \times 10^{-2}}{9.832} \approx 3.3 \times 10^{-3}$$

and

$$0 \leq \theta \leq 90°,$$

it follows that $\tan\alpha \approx 3.3 \times 10^{-3} \sin\theta$. We see, therefore, that even when $\sin\theta$ is unity (P is at the North Pole), α will be substantially less than 1° so that $\cos\alpha \approx 1$. Returning now to Eq. (18-34), we have

$$W' - mg_E \approx -ma_r \cos\theta.$$

Using Eq. (18-31) and rearranging gives

$$W' \approx m(g_E - 3.34 \times 10^{-3} \cos^2\theta)$$

or

$$W' = mg_{\text{eff}} \approx m(9.832 - 3.34 \times 10^{-3} \cos^2\theta). \qquad (18\text{-}37)$$

This analysis shows that the effective value of g is reduced (except at the North Pole, where $\cos\theta$ and, hence, r vanish). Therefore, the use of Newton's laws assuming that the Earth is an inertial reference frame cannot give a correct result unless a fictitious "centrifugal" force F_c of magnitude given by $ma_r \cos\theta$ and directed outward (from O to P in Figure 18-9) is assumed to act. When this is done, Newton's first law (assuming an inertial reference frame with origin at P) yields

$$W' + F_c - mg_E = 0$$

or

$$W' = mg_E - ma_r \cos\theta,$$

which is just Eq. (18-37), as it should be. As noted in Section 18-1, we will not pursue further the complications due to motion relative to non-inertial frames of motion. As a final remark, it should be noted that this approximate analysis provides an expression [Eq. (18-37)] for the variation of g_{eff} with latitude, which is in good agreement with experimentally determined values, ranging from 9.780 m/sec^2 at the Equator to 9.832 m/sec^2 at the North Pole.

18-5 PLANETARY MOTION AND KEPLER'S LAWS

To begin this discussion, we consider a simplified version of planetary motion. Figure 18-10 shows a star of mass M, about which a single planet of mass m orbits in uniform circular motion. The radius of the orbit is R, and T is the period or time required to complete one orbit around the star. We assume for further simplicity that the planet does not rotate about its axis. For the Earth–Sun system, this would involve a time error of only ≈ 240 seconds in one year, so it is not a serious departure from reality.

If it is further assumed that the star does not move, then we can place the origin of an inertial reference frame at the center of the star and use this

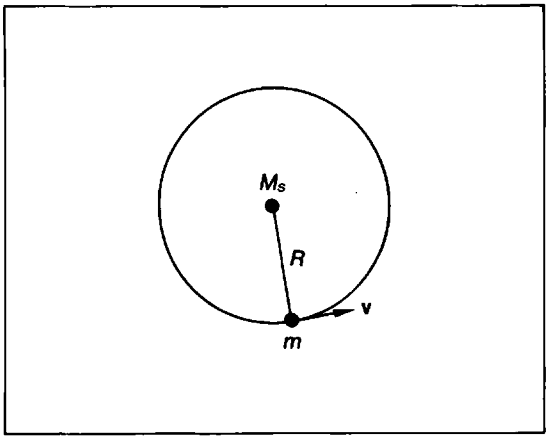


Figure 18-10 Schematic diagram of a planet orbiting a star in uniform circular motion.

Table 18-2 Solar System Data.

Planet	Mass (kg)	Period (sec)	Mean orbit radius (m)
Mercury	3.28×10^{23}	7.60×10^{6}	5.79×10^{10}
Venus	4.83×10^{24}	1.94×10^{7}	1.08×10^{11}
Earth	5.98×10^{24}	3.16×10^{7}	1.49×10^{11}
Mars	6.40×10^{23}	5.94×10^{7}	2.28×10^{11}
Jupiter	1.90×10^{27}	3.74×10^{8}	7.78×10^{11}
Saturn	5.68×10^{26}	9.30×10^{8}	1.43×10^{12}
Uranus	8.67×10^{25}	2.66×10^{9}	2.87×10^{12}
Neptune	1.05×10^{26}	5.20×10^{9}	4.50×10^{12}
Pluto	5.37×10^{24}†	7.82×10^{9}	5.91×10^{12}

†Uncertain

frame to apply Newton's laws to analyze the motion of the planet. The only force acting on m will be the gravitational force due to M_s, which will act along the line joining the centers of the two masses toward M_s. Since the orbital velocity is tangent to the path and constant in magnitude, the resultant acceleration is also radially inward. Therefore, combining Newton's second law and the law of gravitation gives

$$\frac{GM_s m}{R^2} = \frac{mv^2}{R}. \qquad (18\text{-}38)$$

Since

$$v = \frac{2\pi R}{T},$$

we can rearrange Eq. (18-38) to yield

$$T^2 = \frac{4\pi^2}{GM_s} R^3. \qquad (18\text{-}39)$$

Now, if we extend our model by assuming not one but several planets orbit the star without perturbing the motions of each other, then Eq. (18-39) shows that for any two planets regardless of their masses,

$$\frac{T_1^2}{R_1^3} = \frac{T_2^2}{R_2^3}. \qquad (18\text{-}40)$$

Data obtained for the solar system (see Table 18-2) show all the planets have nearly circular orbits (only those of Mercury and Pluto are appreciably non-circular). However, the large masses of the planets lead to some small but significant departures from simple circular orbits for the outermost planets. Historically, in fact, the barely visible distant planets Uranus, Neptune, and Pluto were each discovered only after the minor variations in the orbit of a neighboring planet were accounted for by assuming the existence of another planet further removed from the sun.

The reader may well be puzzled by the fact that planetary masses are given in Table 18-2, when it is clear from Eq. (18-39) that the planetary mass is not involved in the period-orbital radius relation. However, astronomers have measured periods and orbital radii for the satellites (moons) of the planets and they are observed to also move according to Eq. (18-39) with M_s now understood to be the mass of the planet. For example, the mean radius of the Moon's orbit about the Earth is 3.84×10^8 m, and its period is 2.36×10^6 sec, so that

$$M_E = \frac{4\pi^2(3.84 \times 10^8)^3}{6.67 \times 10^{-11}(2.36 \times 10^6)^2} = 5.98 \times 10^{24}\ \text{kg}.$$

(A similar calculation shows that the mass of the Sun is 1.98×10^{30} kg.)

Historically, Newton used the empirical relations governing planetary motion that Johannes Kepler had obtained (by careful analysis of astronomical data gathered by Tycho Brahe) to deduce the law of universal gravitation. We therefore state Kepler's "laws" and try to confirm that such a connection exists. There are three relations and they can be stated as:

I The planets describe elliptical orbits, with the Sun at one focus.

II The position vector of any planet relative to the Sun sweeps out equal areas of the orbital ellipse in equal times.

III The squares of the orbital periods are proportional to the cubes of the average distances of the planets from the Sun.

The reader may recall that a circle is a special case of an ellipse (where the two foci coincide). Then laws I and II are obeyed by our simple model, and we have already seen in Eq. (18-40) that law III is valid as well. It is not as easy to demonstrate the fact that the law of universal gravitation satisfactorily predicts Kepler's laws for a general elliptical orbit. However, since the difficulties are connected with the geometry of ellipses rather than the physical principles involved we shall not pursue the question here. There are numerous sources to which the interested reader may refer.†

Example 4. An Earth satellite moves in a circular orbit at a height of 6×10^5 m above the Earth's surface. Determine:

(a) the period of revolution and

(b) the magnitude of its velocity.

Note that the mean radius of the Earth is 6.40×10^6 m.

SOLUTION

(a) Since we were given the period and radius of the orbit of the Moon about the Earth in calculating the mass of the Earth, we can use them in applying Eq. (18-40):

$$T_s^2 = T_M^2 \left(\frac{R_s}{R_M}\right)^3$$

$$= (2.36 \times 10^6)^2 \left(\frac{7 \times 10^6}{3.84 \times 10^8}\right)^3,$$

$$T_s = 2.36 \times 10^6 \left(\frac{7}{3.84}\right)^{3/2} \times 10^{-3}$$

$$= 5.80 \times 10^3 \text{ sec}$$

$$\approx 97 \text{ min.}$$

(b) $$v = \frac{2\pi R_s}{T_s} = \frac{2\pi(6.40 + 0.60) \times 10^6}{5.80 \times 10^3}$$

$$= \frac{7\pi \times 10^3}{2.9} = 7.58 \times 10^3 \text{ m/sec}$$

$$\approx 17{,}000 \text{ mi/hr.}$$

†See, for example, M. Alonso and E. Finn, *Fundamental University Physics* (Vol I), Addison-Wesley, 1967, Chapter 13.

18-6 CENTRAL FORCE PROBLEMS

We have seen in the last section that the law of universal gravitation and Newton's laws of motion are sufficient for analyzing the motion of planetary systems. The analysis can be made far more general by noting that the form of the gravitational force is rather special in that the force acts along a line joining the two interacting bodies. Thus, the direction of the gravitational force on the planet was taken to be toward the center of the star, so that even though the planet moved in a circular orbit, the force always pointed toward the same central point. Although we have not proved it, when the orbital motion is elliptical rather than circular, the interaction force must still be centrally directed. A force that is always directed toward a fixed point is called a *central force.* It can be shown that motion analogous to that described by Kepler's laws will result regardless of the exact nature of the central force.

There are many situations in nature for which the forces are either central or very nearly central. As a result, understanding the characteristics of planetary motion can lead to the analysis of any new situation *if* the interaction force is basically central in form.

Although we will not discuss them here, two important cases of this sort from the field of atomic physics are: (1) the motion of an electron about a proton in a hydrogen atom is due to a central force that is attractive and electrostatic in nature, and (2) the scattering of alpha particles (helium nuclei) by gold atoms in a thin foil is due to an essentially central force that is repulsive and electrostatic in nature.

One should not feel that there is little merit in studying the motion of planets (or satellites) simply because such studies have been continuing for over 300 years. After all, Rutherford used the alpha particle scattering technique to deduce the nuclear structure of atoms in 1911, and the present space exploration activities of the USA and the USSR are based on the satellite motions that are in principle no different from those of the planets around the Sun.

PROBLEMS

1. A 4.8 lb trolley moves along a horizontal track for which the coefficient of friction is $\frac{1}{6}$. It is pulled by a string that passes over a frictionless pulley at the end of the track attached to a 1.6 lb piece of lead that hangs freely at the end of the string. Find:
   (a) the acceleration of the trolley and
   (b) the tension in the string.
2. A 39 lb block A is placed on an inclined surface and connected by a cord to a 25 lb hanging block B as shown in Figure 18-11. The coefficient of friction between the block and the surface is 0.10. Compute:
   (a) the acceleration of blocks A and B and
   (b) the tension in the cord.

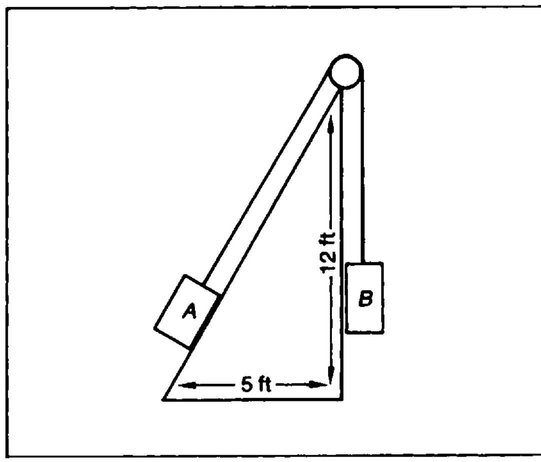


Figure 18-11

3. Find the acceleration that will be imparted to a 10 kg block lying on an inclined plane by a force of 100 nt parallel to the plane, if the plane is inclined at 30° to the horizontal and the coefficient of friction is 0.10 between the block and the plane.
4. A cord extends horizontally from an 8 kg block on a horizontal plane over a pulley and then vertically downward with a 5 kg ball at its lower end. The kinetic coefficient of friction between the block and the plane is 0.250. Compute:
   (a) the acceleration of the block and
   (b) the tension in the cord.
5. In Figure 18-12, m_1 is 1 kg and m_2 is 2 kg. The coefficient of static friction between m_1 and m_2 is 0.40. The coefficient of sliding friction between m_2 and the table is 0.15. The body m_3 has that mass which, when m_3 is released, gives the system the maximum acceleration possible without m_1 slipping relative to m_2.
   (a) What is the maximum acceleration?
   (b) What is the tension in the cord when the system has the maximum acceleration?
   (c) What is the mass of m_3?

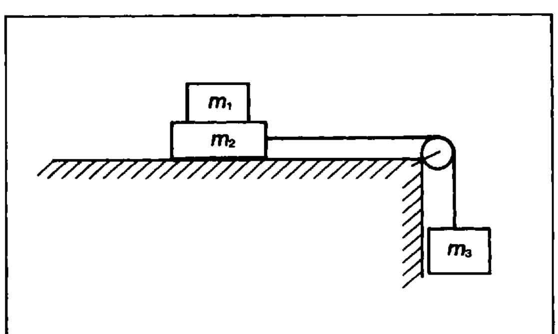


Figure 18-12

6. An 80 lb block on a long plane inclined 30° to the horizontal, starts from rest and slides down the plane. The coefficient of sliding friction between the block and the plane is 0.25. Compute:
   (a) the resultant force on the block,
   (b) the acceleration of the block, and
   (c) the time for the block to slide 20 ft along the inclined plane starting from rest.
7. A block weighing 20 lb is on a plane inclined at 40° with the horizontal. The coefficient of friction between the block and the plane is 0.50.
   (a) What is the minimum push on the block, applied parallel to the plane, that will keep the block from starting to slide down the plane?
   (b) What push applied parallel to the plane will start the block sliding up the plane?
   (c) What push applied parallel to the plane will give the block a constant acceleration of 8 ft/sec^2 up the plane?
8. A 20 lb block on a long plane inclined at 30° to the horizontal is attached to a hanging ball of x lb by means of a cord and pulley as shown in Figure 18-13. The coefficient of sliding friction between the block and the plane is 0.25.
   (a) When the block is sliding up the plane at constant speed, compute the weight of x.
   (b) Later, the cord is cut, allowing the block to slide down the plane. Find its acceleration.

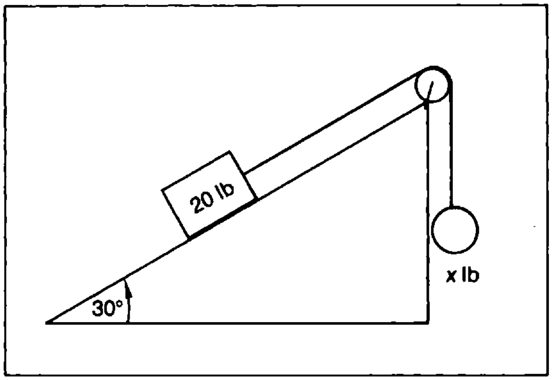


Figure 18-13

9. (a) A baseball player throws a baseball with a velocity of magnitude 96 ft/sec in a direction 30° above the horizontal. If the baseball leaves the thrower's hand at a height of 6 ft above the field, find the maximum height attained and the time it takes to reach the maximum height.
   (b) Find the horizontal distance traveled by the ball before it strikes the ground.
10. A basketball player releases a ball 6 ft above the floor at 60° above the horizontal. It passes through the basket 2.1 seconds later, after reaching a maximum height of 29 ft above the floor.
   (a) What was the speed of the ball as it left the player's hands?
   (b) How high above the floor was the basket?
11. (a) The angle of elevation of an anti-aircraft gun is 70°, and the muzzle velocity of the shell is 2700 ft/sec. If the shell is fused for 50 seconds, what is the greatest height the shell can reach before explosion?
   (b) What are the x- and y- coordinates of the shell 25 seconds after it was fired?
12. A golfer hits a ball and imparts to it an initial velocity 10 m/sec at an angle θ_0 with the horizontal. At the apex of its path, it just clears the top of a tall tree 50 m high. Find:
   (a) the time for half the total flight, assuming a level golf course and
   (b) how far away on the golf course the ball lands?
13. A bullet is fired from ground level. The horizontal component of the muzzle velocity is 1600 ft/sec and the vertical component is 960 ft/sec.
   (a) How long does it take for the bullet to reach the highest point in its trajectory?
   (b) To what maximum height will the bullet rise?
   (c) What is the horizontal range of the bullet?

14. A golf ball is driven with a velocity of 200 ft/sec at an angle of 37° above the horizontal. It strikes a green at a horizontal distance of 800 ft from the tee.
    (a) What was the elevation of the green above the tee?
    (b) What was the velocity of the ball when it struck the ground?
15. A car starts from rest on a circular race track of radius 1000 ft, and increases its speed at the rate of 4 ft/sec^2.
    (a) How long a time will be required to attain the speed at which the tangential and radial components of the car's acceleration are equal?
    (b) How far does the car move along the track before the radial and tangential components of its acceleration are equal?
16. A satellite travels in a circular orbit of radius 7.20×10^6 m with a constant speed of 7 m/sec.
    (a) What is its acceleration?
    (b) What is the average rate of change of speed?
17. A small box is placed 4 ft from the center of a horizontal rotatable platform. If the coefficient of static friction between the box and the platform is 0.50, what is the maximum number of revolutions per second the platform can make without having the box start to slide?
18. On a horizontal frictionless surface, a 500 gm body revolves in a circle whose radius is 90 cm. Find the magnitude of the centripetal force on the body if it makes one revolution per second.

19 Momentum

19-1 IMPULSE AND MOMENTUM

In the previous chapter, a number of physical situations were analyzed satisfactorily by direct application of Newton's laws of motion. In each case, it was possible to identify the forces involved without regarding them as explicit functions of time. For projectile motion, the force due to gravity was constant in magnitude and directed downward throughout the motion. In the case of uniform circular motion, the centripetal force was constant in magnitude and always directed toward the center of the circular path. Thus, it was possible to analyze the motion without regard to the fact that the centripetal force acts in a direction that varies with time.

The situation is less agreeable when we attempt to study cases involving forces that are explicit functions of time, especially those acting for very short time intervals and having magnitudes that vary markedly during the interval. Using the second law, as we did in Chapter 18, will then be difficult. This is so because now the acceleration must also become an explicit function of time, which means that to determine the equation of motion (**r** as a function of time) will require two integrations with respect to time. This can be seen when we consider a mass m subject to a resultant force $\mathbf{F}(t)$. Newton's second law of motion requires that

$$\sum_i \mathbf{F}_i(t) = \mathbf{F}(t) = m\mathbf{a},$$

$$\mathbf{a} = \frac{d\mathbf{v}}{dt} = \frac{\mathbf{F}(t)}{m}. \qquad (19\text{-}1)$$

We can now determine $\mathbf{r}(t)$ by two integrations with respect to time as indicated in Eqs. (14-18) and (14-19).

It is not hard to see that if $\mathbf{F}(t)$ is a complicated function of time, it will be difficult to determine the equation of motion. Furthermore, there are many instances for which it is not possible to state the time dependence of the resultant force in functional form, which additionally complicates the problem. As an example, consider the force imparted to a golf ball when it is struck by a golf club. Until the club contacts the ball, the magnitude of the force is zero. It then rises rapidly to a very large peak value corresponding to the most highly contracted or compressed state of the golf ball. As the ball expands and begins to move away from the club, the magnitude of the force falls rapidly to a value of zero as contact is lost. Thus, a graph of the magnitude of the force as a function of time might resemble Figure 19-1.

For forces of this type, we shall turn to an alternative approach for studying the motion. To begin, we return to the second law and integrate it with respect to time.

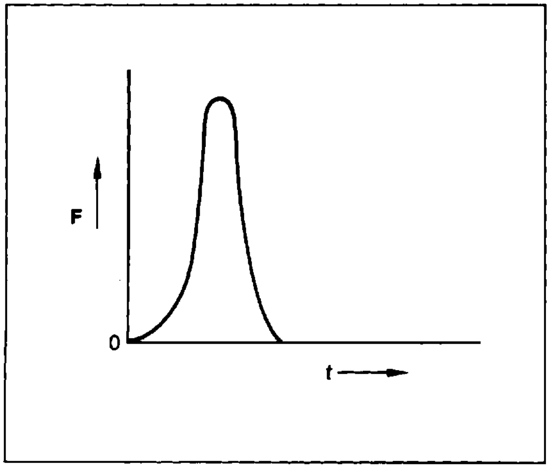


Figure 19-1 Schematic graph of an impulse force as a function of time.

$$\int_0^t \mathbf{F}(t)dt = \int_0^t m\,\mathbf{a}dt. \qquad (19\text{-}2)$$

The integral $\int_0^t \mathbf{F}(t)dt$, is defined to be the impulse of the resultant force $\mathbf{F}(t)$. As we have seen, this integral may be difficult to evaluate. The integral of the RHS offers no difficulty, however, for

$$\int_0^t m\,\mathbf{a}dt = m\int_0^t \frac{d\mathbf{v}}{dt}dt = m\int_{\mathbf{v}_0}^{\mathbf{v}} d\mathbf{v}$$
$$= m(\mathbf{v} - \mathbf{v}_0). \qquad (19\text{-}3)$$

The quantity $m\mathbf{v}$ is defined to be the momentum of a particle. In many texts, it is given the single symbol $\mathbf{p}$. Thus, we could write the relation

$$\mathbf{p} = m\mathbf{v}. \qquad (19\text{-}4)$$

In terms of momentum, Eq. (19-3) states that the impulse of the resultant force acting on a particle for a time t is equal to the resulting change in momentum of the particle during that time. It should be noted that both impulse and momentum are vectors.

Example 1. A golf ball of mass 0.05 kg is initially at rest on a tee. It is then struck by the head of a golf club which is in contact with the ball for 5×10^{-3} sec. If the ball leaves the tee with a speed of 80 m/sec find:

(a) the magnitude of the momentum for the ball just after the contact, and

(b) the impulse delivered to the ball by the club.

(c) Suppose the ball had acquired the same initial speed by means of a constant force F_c acting for the same time interval as the variable force due to the golf club. What would be the magnitude of the constant force?

SOLUTION

(a) From the definition of momentum, we obtain

$$p = mv = (0.05\text{ kg})(80\text{ m/sec}) = 4\text{ kg-m/sec}.$$

(b) By Eq. (19-3), we have

$$\int_0^t F dt = mv - 0$$
$$= 4\text{ kg m/sec}.$$

(c) If the final momentum is to be the same, then the change in momentum will also be the same. (The ball still starts from rest.) Therefore,

$$\int_0^t F_c\,dt = F_c\int_0^t dt = (F_c)(t) = 4\text{ kg m/sec}.$$

Hence,

$$F_c = \frac{4\text{ kg m/sec}}{5\times 10^{-3}\text{ sec}} = 800\text{ nt}.$$

19-2 CONSERVATION OF MOMENTUM

The relation between impulse and momentum [Eq. (19-3)] does not appear to have brought us any closer to our goal since it is in reality only a thinly disguised form of Newton's second law. To see how it can be put to good use, consider the collision of two particles A and B that have masses m_A and m_B, respectively. Initially, A and B have velocities $\mathbf{v}_A$ and $\mathbf{v}_B$, respectively. At time t after the collision, their velocities have become $\mathbf{V}_A$ and $\mathbf{V}_B$ as shown in Figure 19-2. If the only force acting on either particle is the resultant interaction force, then the force on A due to B, $\mathbf{F}_{AB}$, is (by Newton's third law of motion) equal in magnitude but opposite in direc-

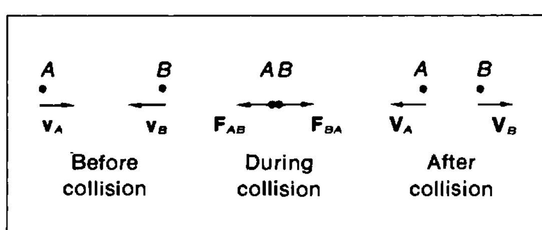


Figure 19-2 One-dimensional two-body collision.

tion to the force on B due to A, $\mathbf{F}_{BA}$. That is, $\mathbf{F}_{AB} = -\mathbf{F}_{BA}$. By Eq. (19-3), the impulse of $\mathbf{F}_{AB}$ is

$$\int_0^t \mathbf{F}_{AB}\, dt = m_A \mathbf{V}_A - m_A \mathbf{v}_A, \tag{19-5}$$

while the impulse of $\mathbf{F}_{BA}$ is

$$\int_0^t \mathbf{F}_{BA}\, dt = m_B \mathbf{V}_B - m_B \mathbf{v}_B. \tag{19-6}$$

Since

$$\int_0^t \mathbf{F}_{AB}\, dt = -\int_0^t \mathbf{F}_{BA}\, dt,$$

Eqs. (19-5) and (19-6) yield

$$m_A \mathbf{V}_A - m_A \mathbf{v}_A = m_B \mathbf{v}_B - m_B \mathbf{V}_B \tag{19-7}$$

or

$$m_A \mathbf{V}_A + m_B \mathbf{V}_B = m_A \mathbf{v}_A + m_B \mathbf{v}_B. \tag{19-8}$$

Equation (19-8) indicates that for a system of two particles subject only to an interaction force (no external forces) the total momentum of the system before the interaction is equal to the total momentum after the interaction. Since there has been no change in total momentum, Eq. (19-8) is a statement of what is called the law of conservation of momentum. Therefore, Newton's laws of motion require that, for a system of two particles subject to no external forces, the total momentum of the system must be constant (or conserved).

Example 2. A particle of mass $m_1 = 1$ kg and a velocity $v_0 = 5$ m/sec to the right collides with a second particle of mass $m_2 = 2$ kg, which is initially at rest. In this collision, mass m_1 is deflected from its original direction by an angle $\theta_1 = 53°$, and its speed after the collision is $v_{1f} = 3$ m/sec. Find:

(a) the angle θ_2 for the direction mass m_2 moves with respect to the original direction of m_1, and

(b) the magnitude of the velocity of the second mass (v_{2f}) after the collision.

SOLUTION

Before the collision, the total momentum of the system is $m_1 v_{10} = 5$ kg m/sec, and it is directed along the x-axis. From Eq. (19-8), therefore, the total momentum afterward must also be along the x-axis. However, v_{1f} has both x- and y-components, so it follows that v_{2f} must also have x- and y-components. Figure 19-3 illustrates the situation before and after the collision. We have shown m_2 moving downward when m_1 moves upward. This must be true if there is to be no y-component for the total momentum. If the final momenta for m_1 and m_2 are broken into x- and y-components, Eq. (19-8) and Figure 19-3 require that

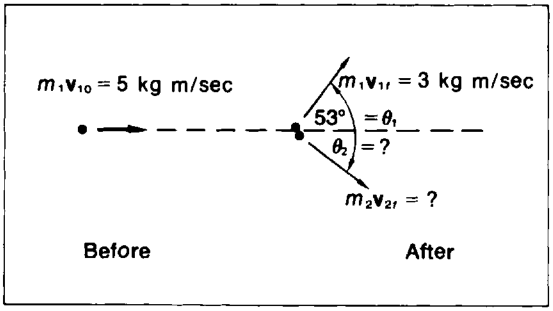


Figure 19-3 System diagram for Example 3.

$$m_1 v_{10} = m_1 v_{1f} \cos\theta_1 + m_2 v_{2f} \cos\theta_2 \tag{i}$$

and

$$0 = m_1 v_{1f} \sin\theta_1 - m_2 v_{2f} \sin\theta_2. \tag{ii}$$

Combining Eqs. (i) and (ii) yields

$$\tan\theta_2 = \frac{v_{1f} \sin\theta_1}{v_{10} - v_{1f}\cos\theta_1}. \tag{iii}$$

Substituting numerical values in Eq. (iii) shows that $\theta_2 = 37°$. Using this result in Eq. (ii) gives

$$v_{2f} = 2 \text{ m/sec}.$$

At this point, we digress briefly to mention that one of the goals of the physical scientist is to discover the various conservation laws governing the material world. Other conservation laws will be encountered as we continue our study. For example, it is found that, regardless of the nature of interaction involving a system of electric charges, the net (positive, zero, or negative) charge of the system is constant. We say, therefore, that the electric charge of a system of particles is conserved. (It is this principle that guides the chemist in balancing a chemical reaction involving a number of positive and negative ions.) By determining all the physical quantities that must be conserved in a given interaction, it may be possible to determine fully the later behavior of the system without direct application of Newton's laws of motion. What is more, the application of conservation laws not only may succeed where the direct approach fails, but will frequently be quite easy to accomplish.

At this point, the reader may have deduced another apparent conservation law involving colliding particles. We refer to the fact that when billiard balls collide their total mass remains constant during the collision, so that one is led to propose a law of conservation of mass for a system of particles subject to no external forces. In this instance, however, the conservation "law" is applicable only for particles whose speeds are small compared to that of light. It was deduced by Einstein in 1905 (and verified with terrible consequences in 1945 at Hiroshima) that the mass and energy of a system of particles are not separable quantities. Thus, it is possible for the mass of a system to decrease (or increase) provided that there is a simultaneous release (or absorption) of energy. Therefore, the correct conservation law in this instance is that the total mass-energy of a system subject to no external forces must be conserved. This follows from the theory of special relativity, but is not discussed further here.

In any case the search for conservation laws has been and continues to be a fruitful field of activity for physicists. We conclude this section with a cautionary note. The "laws" of nature represent a synthesis of experimental evidence and cause-effect relationships related to that evidence. As we noted in Section 9-1, our understanding of nature is subject to constant review (and revision when necessary). Hence, conservation laws obtained by formal manipulation of other "laws" of physics should be tested by careful experimentation to determine the extent of their validity. In this context, the law of conservation of momentum appears to be universally valid, while some of the other conservation laws are more limited in their applicability.

19-3 ELASTIC AND INELASTIC COLLISIONS

When two masses collide, their subsequent motion can be described by Eq. (19-3). If there are no external forces acting, the momentum of the system is conserved and Eq. (19-8) applies. Example 2 illustrates the fact that, for a two particle collision, it is not possible to determine the final momentum (and velocities) of each particle from a knowledge of the initial conditions and Eq. (19-8) only. However, it is possible to classify collisions. Specifying the type of collision is equivalent (for special cases) to providing an additional relationship between the particle velocities before and after the collision.

A *completely inelastic* collision occurs when the two particles stick together and thus move with a common final velocity. When this is true, Eq. (19-8) becomes

$$m_A\mathbf{v}_A + m_B\mathbf{v}_B = (m_A + m_B)\mathbf{V}, \qquad (19\text{-}9)$$

where **V** is the final velocity. Given the initial conditions, **V** can now be calculated.

For a *one-dimensional elastic* collision, the relative velocity of approach of two particles equals the negative of the relative velocity of separation. Using the notation of Eq. (19-8) for a one-dimensional collision, this requirement becomes

$$(v_B - v_A) = -(V_B - V_A), \qquad (19\text{-}10)$$

where

$(v_B - v_A)$ = velocity of approach of B relative to A

and

$(V_B - V_A)$ = velocity of separation of B relative to A.

An alternative form of the requirement that a collision is elastic is the following:

$$\frac{1}{2}m_A v_A^2 + \frac{1}{2}m_B v_B^2 = \frac{1}{2}m_A V_A^2 + \frac{1}{2}m_B V_B^2. \qquad (19\text{-}11)$$

To see that Eq. (19-11) and Eq. (19-10) represent the same requirement, we combine Eqs. (19-11) and (19-8). Thus, dividing Eq. (19-11)

$$m_A(v_A^2 - V_A^2) = m_B(V_B^2 - v_B^2)$$

by Eq. (19-8)

$$m_A(v_A - V_A) = m_B(V_B - v_B),$$

yields

$$v_A + V_A = V_B + v_B$$

or Eq. (19-10)

$$(v_B - v_A) = -(V_B - V_A)$$

as asserted.

The quantity $\frac{1}{2}mv^2$ is defined to be the *kinetic energy* of a particle of mass m and speed v. (The relationship between Newton's laws of motion and kinetic energy is discussed in Section 20-2.) Therefore, Eq. (19-11) states that in an elastic collision the kinetic energy of the two particle system is conserved. By contrast, kinetic energy is *not* conserved in an inelastic collision. For a one-dimensional elastic collision, it is clear that Eqs. (19-8), and (19-10) or (19-11) are sufficient to determine both final velocities. For a two-dimensional elastic collision of two particles, additional information will be required. This is because conservation of momentum and kinetic energy provide only three equations relating the velocities of the two particles, while each velocity has two components in a two-dimensional problem. As a result, there will be four unknown quantities, with only three equations relating them (an insoluble situation).

Example 3. Indicate whether the following collisions are elastic or inelastic.

(a) A mass of 1 kg with an initial velocity of 4 m/sec collides with an 8 kg mass, which is initially at rest. The final velocity of the 1 kg mass is 1 m/sec in the opposite direction.

(b) A mass of 1 kg with an initial velocity of 4 m/sec bounces off an initially stationary object of mass 3 kg. The 1 kg mass has a final velocity of 2.83 m/sec in a direction 90° from its initial direction.

SOLUTION

(a) From Eq. (19-8),

$$(1\text{ kg})(4\text{ m/sec}) = (1\text{ kg})(-1\text{ m/sec}) + (8\text{ kg})(v_{2f})$$

so that

$$v_{2f} = \frac{5}{8}\text{m/sec}.$$

For an elastic collision, Eq. (19-10) requires that

$$v_{10} - v_{20} = v_{2f} - v_{1f}.$$

In this case,

$$(4-0)\text{ m/sec} \stackrel{?}{=} \left[\frac{5}{8} - (-1)\right]\text{m/sec} = 1.625\text{ m/sec}.$$

Therefore, the collision in (a) is not elastic. Since $v_{2f} \neq v_{1f}$, it is not a completely inelastic collision either.

(b) Applying Eq. (19-8) in this case;

x-motion:

$$(1\text{ kg})(4\text{ m/sec}) = (1\text{ kg})(0\text{ m/sec}) + (3\text{ kg})(v_{2fx})$$

y-motion:

$$0 = (1\text{ kg})(2.83\text{ m/sec}) + (3\text{ kg})(v_{2fy}),$$

from which

$$v_{2fx} = \frac{4}{3}\text{m/sec}$$

$$v_{2fy} = -0.94\text{ m/sec}.$$

If the collision is elastic, then Eq. (19-11) should be satisfied. In this case,

$$\frac{1}{2}(1\text{ kg})(4\text{ m/sec})^2 + \frac{1}{2}(3\text{ kg})(0\text{ m/sec})^2 \stackrel{?}{=} \frac{1}{2}(1\text{ kg}) \times (2.83\text{ m/sec})^2 + \frac{1}{2}(3\text{ kg})\left[\left(\frac{4}{3}\right)^2 + (-0.94)^2\right](\text{m/sec})^2;$$

or, dropping the common factor of $\frac{1}{2}$ and the energy units,

$$16 \stackrel{?}{=} 8 + 3\left(\frac{16}{9} \times 0.88\right) = 8 + 7.95 = 15.95.$$

To the accuracy of the data the collision is elastic.

19-4 MOTION OF SYSTEMS WITH VARIABLE MASS

Let us consider in more detail the impulse-momentum relation,

$$\int_0^t \mathbf{F}(t)\,dt = m\mathbf{v} - m\mathbf{v}_0.$$

We have dealt thus far with systems for which the mass does not change during the time t. When this is true, Eq. (19-3) is completely equivalent to Newton's second law of motion in the form

$$\mathbf{F}(t) = \frac{d}{dt}(m\mathbf{v}). \qquad (19\text{-}12)$$

Now, however, consider a system for which the mass is not constant. Experiment shows that Eq. (19-12) is a valid statement, while the form

$$\mathbf{F}(t) = m\frac{d\mathbf{v}}{dt} = m\mathbf{a} \qquad (19\text{-}13)$$

is not. Since the rules of differentiation require that

$$\frac{d(m\mathbf{v})}{dt} = m\frac{d\mathbf{v}}{dt} + \frac{dm}{dt}\mathbf{v},$$

we see that the momentum concept has led us to a generalized form of the second law, which applies whether or not the mass is constant. It is a matter of history that Newton himself recognized the importance of momentum and expressed the second law of motion in the form of Eq. (19-12). As an example of a system of variable mass that is of considerable current interest, we consider in the next example the launching of a rocket. Here, the mass of the rocket decreases with time (as fuel is consumed to provide launching thrust due to the escaping gases).

Example 4. A rocket is to be launched vertically subject to the following conditions: at $t = 0$, it is at rest relative to the Earth; its mass (fuel plus rocket) is initially m_0; the rate of mass decrease dm/dt is assumed constant until a final mass m_f is attained. The velocity of the escaping gases relative to the rocket is $\mathbf{v}_e$, also assumed to be constant. Finally, we neglect air resistance, and assume that the "burn time" is short enough that the acceleration takes place under a constant acceleration due to surface gravity. Find the rocket velocity as a function of time during the "burn time."

SOLUTION

Consider the Earth to be an (approximate) inertial reference system with $\mathbf{v}$ (the velocity of the rocket relative to the Earth) and $\mathbf{p} = m\mathbf{v}$ (the corresponding momentum of the rocket) both at time t. In a small time interval dt, there is a mass decrease dm and a velocity increase $d\mathbf{v}$ relative to the Earth. Now $\mathbf{v}_g$, the velocity of the escaping gases relative to the Earth, is related to $\mathbf{v}_e$ and $\mathbf{v}$ by the equation

$$\mathbf{v}_g = \mathbf{v}_e - \mathbf{v}. \qquad (19\text{-}14)$$

At time $t + dt$, the momentum of the system relative to the Earth, $\mathbf{p} + d\mathbf{p}$ is in two parts:

$$\mathbf{p} + d\mathbf{p} = \underset{\text{Rocket}}{(m - dm)(\mathbf{v} + d\mathbf{v})} - \underset{\text{gases}}{dm\,\mathbf{v}_g}$$

or

$$\mathbf{p} + d\mathbf{p} = m\mathbf{v} - dm\,\mathbf{v} + m\,d\mathbf{v} - dm\,d\mathbf{v} - dm(\mathbf{v}_e - \mathbf{v}). \qquad (19\text{-}15)$$

Substituting Eq. (19-14) in Eq. (19-15), and neglecting the small quantity $dm\,d\mathbf{v}$ gives

$$\mathbf{p} + d\mathbf{p} = m\mathbf{v} + m\,d\mathbf{v} - dm\,\mathbf{v}_e. \qquad (19\text{-}16)$$

Using $\mathbf{p} = m\mathbf{v}$, and the fact that $d\mathbf{p}$ occurs in a time dt, we can write

$$\frac{d\mathbf{p}}{dt} = m\frac{d\mathbf{v}}{dt} - \frac{dm}{dt}\mathbf{v}_e. \qquad (19\text{-}17)$$

This change in momentum of the system is related to $\mathbf{F} = m\mathbf{g}$, the external force (due to gravity) by Eq. (19-2),

$$\mathbf{F} = \frac{d\mathbf{p}}{dt}$$

or

$$m\mathbf{g} = m\frac{d\mathbf{v}}{dt} - \frac{dm}{dt}\mathbf{v}_e. \qquad (19\text{-}18)$$

This can be re-arranged as follows:

$$\frac{d\mathbf{v}}{dt} = \mathbf{g} + \frac{1}{m}\frac{dm}{dt}\mathbf{v}_e.$$

Now, both $\mathbf{g}$ and $\mathbf{v}_e$ are directed downward, while $d\mathbf{v}/dt$ is upward (if launch is to occur!). Therefore,

$$dv = -g\,dt - \frac{1}{m}dm\,v_e. \qquad (19\text{-}19)$$

Equation (19-19) can now be integrated, noting that at $t = 0$, $v = 0$, and $m = m_0$, while at any later time (up to t_F, the end of the "burn") the corresponding values are $v(t)$ and m. Thus,

$$\int_0^{v(t)} dv = -g\int_0^t dt - v_e\int_{m_0}^m \frac{dm}{m},$$

which gives

$$v(t) = -gt - \left[\ln\left(\frac{m}{m_0}\right)\right]v_e. \qquad (19\text{-}20)$$

Using the fact that $m < m_0$, Eq. (19-20) becomes

$$v(t) = -gt + v_e \ln\left(\frac{m_0}{m}\right). \qquad (19\text{-}21)$$

As a numerical example, the following characteristics apply to the Centaur rocket†:

$$m_0 \approx 2.75 \times 10^6\,\text{kg},$$

$$m_F \approx 2.45 \times 10^6\,\text{kg},$$

$$\frac{dm}{dt} \approx 1300\,\text{kg/sec}.$$

Since

$$m = m_0 - m_F = \frac{dm}{dt}(t_F - 0),$$

it follows that

$$t_F = \frac{(0.30 \times 10^6\,\text{kg})}{(1300\,\text{kg/sec})} = 231\,\text{sec}.$$

A reasonable estimate for the magnitude of the exhaust velocity for this rocket is 5×10^4 m/sec. Thus, if $v = 0$ at $t = 0$, then the final velocity $v(t_F)$ will be, from Eq. (19-21),

$$v(t_F) = (-9.8)(231) + 5 \times 10^4 \ln 1.11$$
$$\approx 3000\,\text{m/sec},$$
$$\approx 6700\,\text{mi/hr}.$$

†The Centaur rocket is a [second-stage] rocket designed for use with the Saturn rockets used in the Apollo moon program.

19-5 FURTHER CONSIDERATIONS

It was noted in Section 19-4 that Eq. (19-12) is a more generally valid form of Newton's second law of motion than the form given by Eq. (19-13). The example of a rocket launch with a time-dependent mass was used as an illustration of the general form. The reader should deduce that if $dm/dt = 0$, Eqs. (19-12) and (19-13) become identical for typical types of motion.

The phrase "typical types of motion" is meant to refer to situations where the speeds involved are small compared to that of light. We saw, in Chapter 15, that the description of high speed kinematics requires the modifications introduced by Einstein in special relativity theory. There are additional effects that must be considered, the most significant of which we present here without proof†.

Suppose a particle of mass m_0 when at rest is caused to move at a speed v comparable to (but less than!) c. The theory of relativity requires, and experimental measurements confirm, that the mass m_0 must be replaced by a mass m related to m_0 by the expression

$$m = \frac{m_0}{\left[1 - \left(\frac{v}{c}\right)^2\right]^{1/2}}. \qquad (19\text{-}22)$$

If the speed depends upon the time, m will also be a function of time. As a result, the momentum for relativistic speeds is given by the expression mv and *not* by the expression m_0v. When this is done, it is found that Eq. (19-12) remains valid even for relativistic speeds.

It was asserted in Section 19-2 that the law of conservation of momentum is universally valid. Historically, the firm conviction that momentum conservation was universal led Wolfgang Pauli to postulate a particle that was to have no charge and no mass (or a very small mass). It was needed in order to account for momentum and energy apparently lost in a particular kind of nuclear decay process. The existence of this particle, called the neutrino ("little neutral one" in Italian) by Enrico Fermi, was not verified experimentally until 1958, about 25 years after its necessary properties were identified.

In view of the several radical modifications of the classical concepts of mechanics that special relativity requires, it is remarkable that a quantity as fundamental as momentum retains its identity regardless of the particle speed. This is not true for the kinetic energy.

†See any text on special relativity.

PROBLEMS

1. An 8 gm bullet is fired horizontally into a 90 gm block of wood that is free to move. The velocity of the combined block and bullet after impact is 4000 cm/sec. What was the initial velocity of the bullet?
2. If a bat is in contact with a ball for 0.02 sec, and it exerts an average force of 100 nt, what is the impulse given to the ball and the change in momentum of the bat?
3. A 16 gm body traveling with a velocity of 30 cm/sec collides inelastically head on with a 4 gm body traveling in the opposite direction with a velocity of 50 cm/sec. What is the magnitude of the velocity of the 16 gm body after collision?
4. A 3000 lb car traveling north and a 10,000 lb truck traveling east at 40 mph collide at an intersection; the combined pile of junk slides 30° north of east. Determine the speed of the car just before the collision.
5. A 32 lb steel disc hangs at the end of a rope. A bullet moving horizontally with a speed of 1600 ft/sec strikes it and drops down. Find the speed with which the steel disc begins to move, if the weight of the bullet is 1 oz.
6. A 200 lb skater traveling with a velocity of 20 ft/sec eastward collides with a 160 lb skater who is traveling northward at 15 ft/sec.
   (a) In what direction does the tangled heap of skaters slide?
   (b) What is the speed of the tangled skaters immediately after the collision?
7. A bullet of mass 100 gm moves horizontally eastward at 1000 m/sec, and has a perfectly inelastic collision with a block of mass 1 kg. The block is moving horizontally eastward at 10 m/sec at the moment of impact. If the coefficient of kinetic friction between the block and the plane is 0.51, how far does the block slide after collision?

8. A simple pendulum consists of an 800 gm bob hanging from a string. The length of the string is 130 cm. A 100 gm projectile moving horizontally toward the center of the bob strikes the bob and is imbedded in it. After impact, the velocity of the bob plus projectile is 120 cm/sec. Compute:
   (a) the height through which the bob plus projectile rises and
   (b) the magnitude of the velocity of the projectile immediately before impact.
9. An empty freight car weighing 10 tons rolls at 3 ft/sec along a level track and collides with a loaded car weighing 20 tons, standing at rest with the brakes released. If the two cars couple together, find:
   (a) their velocity after collision and
   (b) the impulse given to the loaded car.
10. A 96 lb boy running 6 ft/sec north jumps into a 112 lb boat moving 10 ft/sec in the opposite direction. Neglect friction between the boat and the water. Compute the velocity (magnitude and direction) of the boat after the boy arrives in it.
11. A lead bullet has a weight of 2 oz and a velocity of 1280 ft/sec. It hits the center of a block of wood and remains imbedded, the mass of the block and bullet being 8 lb. Find:
    (a) the velocity with which the block (with the bullet imbedded) starts to move,
    (b) the acceleration and the retarding force if the block is on ice when fired at, and comes to rest in 10 seconds, and
    (c) the coefficient of friction between the block and the ice.
12. A golf ball weighing $1\frac{2}{3}$ oz is driven from the tee with the initial velocity of 180 ft/sec. If the club is in contact with the ball for 0.0005 sec, compute the average force exerted on the ball.
13. A 5 kg body with an initial velocity of 20 m/sec travels 6 m while acted on by a force of 500 nt in the direction of its initial velocity. Find the final velocity and the time during which the force acts.
14. In accelerating a 0.50 lb hockey puck from rest to a speed of 80 ft/sec, the stick is in contact with the puck for 0.040 seconds. Compute:
    (a) the momentum of the puck as it leaves the stick and
    (b) the average force exerted on the puck by the stick.
15. A 150 lb man dives from a high springboard and reaches a maximum height above the surface of the water of 25 ft. After striking the water, he comes to rest in $\frac{1}{2}$ sec.
    (a) What is the momentum of the diver at the instant he strikes the water?
    (b) What average force does the water exert on the diver?

20 Work and Mechanical Energy

20-1 DEFINITION OF WORK

In this chapter, we develop another method for analyzing particle motion as an alternative to the direct application of Newton's laws of motion. It will be based upon these laws of motion but will involve the concept of work. It is first necessary to make clear what is meant by the term work in physics because it has so many meanings in our everyday language.

In physics, a particle that experiences a displacement $\Delta\mathbf{s}$ while subject to a constant force $\mathbf{F}$ that is directed at an angle ϕ relative to $\Delta\mathbf{s}$ has, by definition, experienced an amount of work ΔW given by

$$\Delta W = |\mathbf{F}||\Delta\mathbf{s}| \cos \phi. \qquad (20\text{-}1)$$

That is, the product of the magnitudes of the displacement and the component of force parallel to the displacement during the application of the force gives the amount of work done on the particle. As Eq. (20-1) shows, work is a scalar quantity even though it is defined in terms of two vector quantities. When two vectors are multiplied to yield a scalar quantity, the process of multiplication is called a scalar multiplication and yields a scalar product. This is indicated symbolically by the "dot" product notation

$$\mathbf{A} \cdot \mathbf{B} = |\mathbf{A}||\mathbf{B}| \cos \phi, \qquad (20\text{-}2)$$

where ϕ is the angle between $\mathbf{A}$ and $\mathbf{B}$. (The scalar [dot] product is also discussed in the Appendix.) With this definition, we may rewrite Eq. (20-1):

$$\Delta W = \mathbf{F} \cdot \Delta\mathbf{s}. \qquad (20\text{-}3)$$

If the particle follows a winding path while subject to a force that is not constant but depends upon the position s along the path, then Eq. (20-3) must be replaced. By considering only an infinitesimal displacement $d\mathbf{s}$ along the path and the force $\mathbf{F}(s)$ (which has a unique value when $d\mathbf{s}$ becomes vanishingly small), we can express the infinitesimal amount of work dW as

$$dW = \mathbf{F}(s) \cdot d\mathbf{s}. \qquad (20\text{-}4)$$

If the particle goes from point A to point B along the path, the total work $W_{A\to B}$ or W_{AB} will be the sum of all the individual dW's for displacements from A to B. Such a sum is, of course, achieved by integration, indicated symbolically as

$$W_{AB} = \int_A^B \mathbf{F}(s) \cdot d\mathbf{s}. \qquad (20\text{-}5)$$

This integral depends upon the form of the path from A to B *and* also involves a force vector whose magnitude and direction are functions of position along the path. Such an integral is called, a *line integral* to distinguish it from the simpler integrals encountered in determining **v** and **r** from a knowledge of the time dependence of **a**. (See Chapter 14.) When the component of $\mathbf{F}(s)$ tangent to the path is known as a function of s for a given path, the line integral

$$\int_A^B \mathbf{F}(s) \cdot d\mathbf{s}$$

becomes the simpler integral

$$\int_A^B F_t(s)\,ds,$$

where $F_t(s)$ is the component of $\mathbf{F}(s)$ tangent to the path element. In the following sections, illustrations of the usefulness of this expression are given.

Example 1. The graph of Figure 20-1 illustrates the variation of the tangential component of the force applied to a particle as a function of the particle displacement. Determine the work done as the particle moves from the origin to a point 50 m away.

SOLUTION

Since the graph illustrates $F_t(s)$ versus s, Eq. (20-5) becomes

$$W_{0\text{-}50} = \int_0^{50} F_t(s)\,ds$$

in terms of this data.

As we saw in Chapter 14, the value of such an integral is given by the area under the curve—here, $F_t(s)$ versus s. In this case, the area is conveniently split into three separate regions for which the force-displacement relations are

I $F_t(s) = 15s$ nt; $0 \le s \le 30$ m.

II $= 450$ nt; $30 \le s \le 40$ m.

III $= [450 - 35(s - 40)]$ nt; $40 \le s \le 50$ m
$= [1850 - 35s]$ nt.

From the graph,

I $W_{0\text{-}50} = \frac{1}{2}(30\text{ m})(450\text{ nt})$

II $+ (450\text{ nt})(10\text{ m})$

III $+ \frac{1}{2}(10\text{ m})(350\text{ nt}) + (100\text{ nt})(10\text{ m})$

$= 14 \times 10^3$ joules.

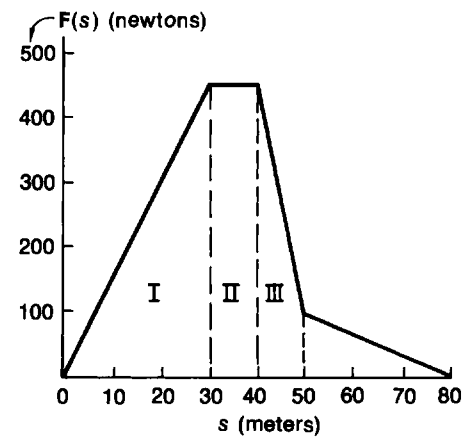


Figure 20-1 Force versus displacement for Example 1.

As a check, notice that the analytic integration gives

$$W_{0\text{-}50} = \int_0^{30} 15s\,ds + \int_{30}^{40} 450\,ds + \left[450s - 35\left(-40s + \frac{s^2}{2}\right)\right]\Bigg|_{40\text{m}}^{50\text{m}}$$

$$= \frac{15s^2}{2}\Bigg|_{0\text{m}}^{30\text{m}} + 450s\Bigg|_{30\text{m}}^{40\text{m}} + \left[450s - 35\left(-40s + \frac{s^2}{2}\right)\right]\Bigg|_{40\text{m}}^{50\text{m}}$$

$$= (450\text{ nt})(15\text{ m}) + (450\text{ nt})(10\text{ m}) + (450\text{ nt})(50 - 40)\text{m} - 35\left[(-40(10)) + \frac{1}{2}(2500 - 1600)\right]\text{ joules,}$$

$$= 14 \times 10^3\text{ joules,}$$

as before.

Work is a derived physical quantity, and its units are determined by its definition. In MKS units, the product of force and displacement units yields the newton-meter, a quantity given a single label: the joule. In cgs units, the dyne-centimeter is called an erg; the reader can show that 10^7 ergs = 1 joule.† In

†As an indication of the magnitude of an erg, it is approximately equal to the work required to lift a gnat 1 cm on the Earth's surface. Thus, the reader has only to lift 10 million gnats 1 cm—or 100,000 gnats 1 m—to do a joule of work.

the British engineering system of units the unit of work is called the ft-lb. (There is no single name for this unit.) As mentioned in the Introduction, the dimensions of work are ml^2t^{-2}. By using the conversions from meters and kilograms to feet and slugs, one finds that 1 joule = 0.7376 ft-lb.

In many applications, it is desirable to know the amount of work done per unit time at any given instant. The name given to this physical quantity is instantaneous power (P). It is defined by the relation

$$P = \frac{dW}{dt}, \tag{20-6}$$

If Eq. (20-4) is combined with Eq. (20-6), we find that

$$P = \mathbf{F} \cdot \frac{ds}{dt} = \mathbf{F} \cdot \mathbf{v}. \tag{20-7}$$

The average power $\bar{P}$ during a time interval t is just the total work done W divided by the time interval t, $\bar{P} = W/t$. From the point of view of an engineer, the rate at which work is done by a machine is of equal or greater significance than the total work done. The units of power are:

$$\text{MKS: } 1\frac{\text{nt-m}}{\text{sec}} = 1\frac{\text{joule}}{\text{sec}} = 1 \text{ watt}$$

$$\text{cgs: } 1\frac{\text{dyne-cm}}{\text{sec}} = 1\frac{\text{erg}}{\text{sec}}$$

$$\text{British Engineering: } 1\frac{\text{ft-lb}}{\text{sec}}$$

For historical reasons, English-speaking engineers adopted the horsepower (hp) as the British engineering unit of power. It is related to other units by the relations

$$1 \text{ hp} = 550\frac{\text{ft-lb}}{\text{sec}} = 746 \text{ watts}.$$

With the present high demands for power in industrialized nations, it is more common to speak of power in kilowatts (1 kw = 10^3 watts) or even megawatts (1 Mw = 10^6 watts). Thus, for example, residential electricity meters record the total electrical work delivered in kilowatt-hours (kwh), where 1 kwh = 3.6×10^6 joules. In the following sections, we will not be concerned further with power considerations in spite of their practical importance.

20-2 WORK AND KINETIC ENERGY

Consider a particle of mass m having a velocity $\mathbf{v}_A$ at point A and subject to a constant force $\mathbf{F}$ tangent to the path. The work due to this force will cause an increase in the velocity. To see this, first note that from Newton's second law;

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

so that

$$W_{AB} = \int_A^B \mathbf{F} \cdot d\mathbf{s} = \int_A^B F\,ds$$

$$= \int_A^B ma\,ds = mas_{AB}, \tag{20-8}$$

where s_{AB} is the magnitude of the displacement from A to B. [Since $\mathbf{F}$, $\mathbf{a}$, and $d\mathbf{s}$ are parallel, we need only consider magnitudes as indicated in Eq. (20-8).] Now, from the chain rule of differentiation, we could replace a by the expression

$$a = \frac{dv}{dt} = \frac{dv}{ds}\frac{ds}{dt} = v\frac{dv}{ds}, \tag{20-9}$$

and Eq. (20-8) becomes

$$W_{AB} = \int_A^B mv\,dv = \int_{v_A}^{v_B} mv\,dv$$

$$= \frac{1}{2}mv_B^2 - \frac{1}{2}mv_A^2. \tag{20-10}$$

By combining Eqs. (20-8) and (20-10), we get the relation

$$v_B^2 = v_A^2 + 2as_{AB}. \tag{20-11}$$

This is just the kinematical expression Eq. (14-15) relating velocity and displacement for a constant acceleration, now determined dynamically.

We have a new relation in Eq. (20-10), however. It states that the result of applying a constant force tangent to the path traveled by a particle is to change the value of a quantity $\frac{1}{2}mv_A^2$ to a value $\frac{1}{2}mv_B^2$, where v_B is the magnitude of the velocity of the particle at point B, as indicated by Eq. (20-11). If we define the translational kinetic (or motional) energy (K.E.) of the particle by the relation

$$\text{K.E.} = \frac{1}{2}mv^2, \tag{20-12}$$

then Eq. (20-10) simply becomes

$$W_{AB} = (\text{K.E.})_B - (\text{K.E.})_A, \tag{20-13}$$

$$= \Delta(\text{K.E.}),$$

or, in words, the difference between the final and initial kinetic energy values of the particle is equal to the work done on it by the force **F**.

It should be stressed that Eq. (20-10) is valid even when F (and, hence, **a**) is not constant [so that Eq. (20-8) is no longer applicable]. As an alternative to solving the vector relationships of Newton's laws of motion, it is now possible to apply the scalar relationship between the work and the change in kinetic energy.

Example 2. A 75 kg sled is made to move a horizontal distance of 60 m (starting from rest) by a constant horizontal force of 250 nt.

(a) Calculate the work done by the force.

(b) Determine the final velocity of the sled.

SOLUTION

(a) From Eq. (20-3),

$$W_{0\text{-}60} = (250\,\text{nt})(60\,\text{m}) = 15 \times 10^3\,\text{joules}.$$

(b) Since the initial velocity is zero, Eq. (20-10) becomes

$$(15 \times 10^3)_{\text{joules}} = \frac{1}{2}(75\,\text{kg})v_{60}^2,$$

$$v_{60}^2 = \left[\frac{30}{75} \times 10^3\right](\text{m/sec})^2 = [4 \times 10^2](\text{m/sec})^2,$$

$$v_{60} = 20\,\text{m/sec}.$$

Notice also that $a = F/m = 250\,\text{nt}/75\,\text{kg} = 10/3$ m/sec^2, and Eq. (14-15) yields $v_{60} = (2as_{AB})^{1/2} = 20$ m/sec, as it should.

Example 3. An unknown force causes the velocity of a particle of mass 5 kg to increase from 10 m/sec to 20 m/sec. Determine the work done by the force.

SOLUTION

From Eq. (20-10),

$$W_{AB} = \frac{1}{2}(5\,\text{kg})(20\,\text{m/sec})^2 - \frac{1}{2}(5\,\text{kg})(10\,\text{m/sec})^2$$

$$= 750\,\text{joules}.$$

20-3 CONSERVATIVE AND NON-CONSERVATIVE FORCES

In work-energy considerations, forces are divided into two classes—conservative and nonconservative. What distinguishes the two classes is the fact that the work done on a particle by a conservative force does not depend upon the path traversed but only upon the initial and final locations of the particle. Thus, if the work done in going from A to B by a given force is independent of the path, we can consider any two separate paths between A and B which necessarily involve the same amount of work W_{AB}. (See Figure 20-2.) Equation (20-5) requires that

$$W_{AB} = \int_A^B \mathbf{F} \cdot d\mathbf{s},$$

from which it follows that W_{BA}, the work done by the force **F** in moving the particle from B to A, will be given by

$$W_{BA} = \int_B^A \mathbf{F} \cdot d\mathbf{s} = -\int_A^B \mathbf{F} \cdot d\mathbf{s} = -W_{AB}. \quad (20\text{-}14)$$

If the particle is moved from A to B on the path I, and returned to A by path II, the total work done by the conservative force involved will be

$$W_{\text{total}} = W_{AB}(\text{I}) + W_{BA}(\text{II}).$$

For a conservative force, Eq. (20-14) requires that

$$W_{BA}(\text{II}) = -W_{AB}(\text{I}),$$

and the total work done is, thus,

$$W_{\text{total}} = W_{AB}(\text{I}) - W_{AB}(\text{I}) = 0.$$

We can, therefore, give as an alternative definition of a conservative force the requirement that the work done by the force on a particle in moving around a closed path must be zero. This requirement is written in equation form as

$$W_{\text{closed path}} = \oint \mathbf{F} \cdot d\mathbf{s} = 0 \quad \text{(conservative force)}, \quad (20\text{-}15)$$

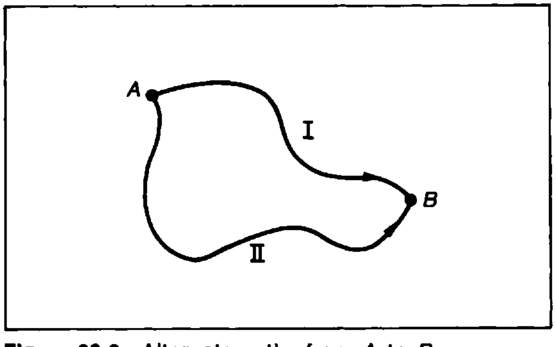


Figure 20-2 Alternate paths from A to B.

where the symbol $\oint$ denotes a line integral over a path beginning and ending at the same point.

There are a number of conservative forces that are of considerable interest, including gravitational, elastic, and electrostatic forces. In the rest of this chapter, we shall consider the first two, leaving a discussion of the third to Chapter 29.

Example 4. Figure 20-3 shows a particle of mass m initially located at a distance r_i from a particle of mass M. The particle of mass m is to be carried at constant speed around the closed path consisting of the segments labeled 1, 2, 3, and 4, respectively. Show that the work done by the force of gravitation over this path is zero.

SOLUTION

In this case, the gravitational force exerted by M on m has a magnitude given by $F = GMm/r^2$, and the direction of the force is radially inward from m toward M. (It is an attractive force, as discussed in Section 16-5.) Equation (20-15) can be written as

$$W_{\text{closed path}} = W_1 + W_2 + W_3 + W_4,$$

or, in words, the work done around the closed path is equal to the sum of the amounts of work done on the separate segments of the path.

Paths 1 and 3 are arcs of concentric circles, so that on these paths the gravitational force and the displacement vector will be perpendicular. Therefore, by Eq. (20-1) the work done on each of these segments will be zero. The work done along path 2 is readily found. On this path,

$$d\mathbf{s} = d\mathbf{r} \quad \text{and} \quad \mathbf{F} \cdot d\mathbf{s} = -\frac{GMm}{r^2}\,dr.$$

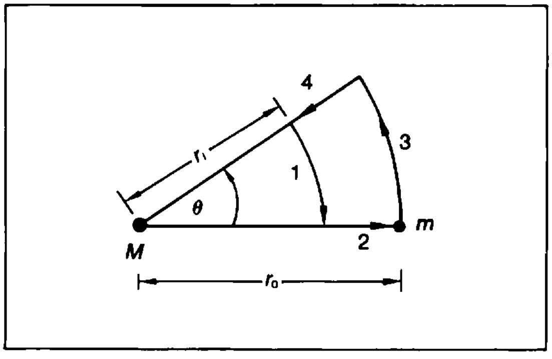


Figure 20-3 Diagram for Example 4.

(Since F is radially inward.) Therefore,

$$W_2 = -GMm\int_{r_i}^{r_0}\frac{dr}{r^2}$$
$$= (-GMm)\left[-\left(\frac{1}{r_0} - \frac{1}{r_i}\right)\right]$$
$$= GMm\left(\frac{1}{r_0} - \frac{1}{r_i}\right).$$

On path 4, we have $d\mathbf{s} = d\mathbf{r}$ and $\mathbf{F} \cdot d\mathbf{s} = -GMm/r^2\,dr$. As a result,

$$W_4 = -GMm\int_{r_0}^{r_i}\frac{dr}{r^2}$$
$$= (-GMm)\left[-\left(\frac{1}{r_i} - \frac{1}{r_0}\right)\right]$$
$$= -GMm\left(\frac{1}{r_0} - \frac{1}{r_i}\right).$$

Since $W_4 = -W_2$, we obtain the desired result,

$$W_{\text{closed path}} = W_1 + W_2 + W_3 + W_4$$
$$= 0 + W_2 + 0 - W_2 = 0.$$

Thus, the gravitational force is seen to be conservative, as asserted. Physically, what this analysis shows is that, in taking mass m along path 2, we must actually work against gravity, which acts inward. Along path 4, the amount of work done by gravity is equal in magnitude, but opposite in sign to that done along path 2. In other words, we can think of work put in along part of the path and taken out along another part. For a conservative force, the net work done must be zero, leading to no change in kinetic energy for the closed path. For a non-conservative force, Eq. (20-15) becomes

$$W_{\text{closed path}} = \oint \mathbf{F} \cdot d\mathbf{s} \neq 0 \quad \text{(non-conservative force).} \qquad (20\text{-}16)$$

Alternatively, we can define non-conservative forces by saying that the work done by such forces is *not* independent of the path chosen. In moving a particle from point A to point B in this case, the work done cannot be specified unless the path is also specified. The force of kinetic or sliding friction (discussed in Section 18-2) is an example of a non-conservative force.

Example 5. An object of mass m is to be moved horizontally at constant speed from its initial position A to a point B a distance s along the horizontal surface and back to A, as illustrated in Figure

20-4. As the figure shows, there is a sliding friction force of magnitude $\mu_K mg$. Show that this force is non-conservative.

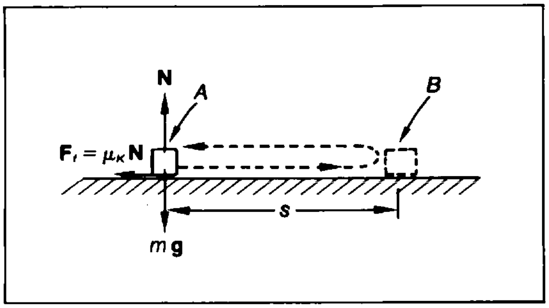


Figure 20-4 Force diagram for Example 5.

SOLUTION

If the friction force is non-conservative, $W_{\text{closed path}} \neq 0$. To show that this is the case, consider the closed path as composed of path 1 from A to B and the path 2 from B to A. For path 1,

$$W_1 = \int_A^B \mathbf{F}_f \cdot d\mathbf{s} = -\int_0^s (\mu_K mg)\,ds = -\mu_K mgs.$$

For path 2,

$$W_2 = \int_B^A \mathbf{F}_f' \cdot d\mathbf{s} = \int_s^0 (\mu_K mg)\,ds = -\mu_K mgs.$$

In words, for both segments of the closed path, the friction force acts in a direction opposite to the displacement. Therefore,

$$W_{\text{closed path}} = -2\mu_K mgs,$$

and we have shown that $W_{\text{closed path}} \neq 0$.

Example 6. Figure 20-5 shows an object of mass m located initially at point A on the Earth's surface. It is to be moved to a final position at point B via two alternative paths. Along path 1, it slides horizontally a distance s on a surface for which a coefficient of sliding friction μ_K can be specified. It is then raised a vertical distance y to point B against a gravitational force. Along path 2, it is raised a vertical distance y against the gravitational force, and then moved a horizontal distance s to the point B. Show that

$$W_2 \neq W_1.$$

SOLUTION

Along the horizontal portion of path 1, the forces acting are as shown in Figure 20-6. Since N and $m\mathbf{g}$ are perpendicular to this part of the path, they do not contribute to the work done. Along the vertical portion of the path, only the gravitational force is involved. Therefore,

$$\begin{aligned} W_1 &= W_{\text{horizontal}} + W_{\text{vertical}} \\ &= -\int_0^s \mu_K N\,ds - \int_0^y mg\,dy \\ &= -\mu_K Ns - mgy. \end{aligned}$$

Since $N = mg$, we can write

$$W_1 = -mg(\mu_K s + y).$$

For path 2, the work done on the vertical path is again $-mgy$ since only the gravitational force is involved, as before. In moving the horizontal distance s to point B to complete path 2, we now have a frictionless situation since the object is no longer in contact with the surface. Thus, no work is done on this horizontal path and

$$W_2 = -mgy.$$

We see, therefore, that

$$W_2 \neq W_1,$$

as required.

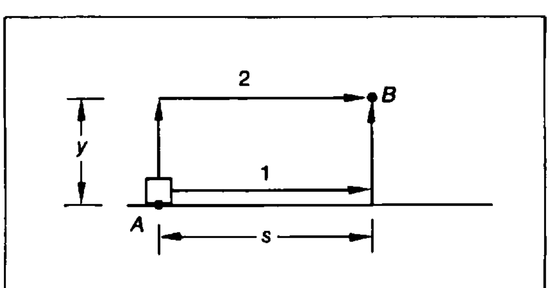


Figure 20-5 System diagram for Example 6.

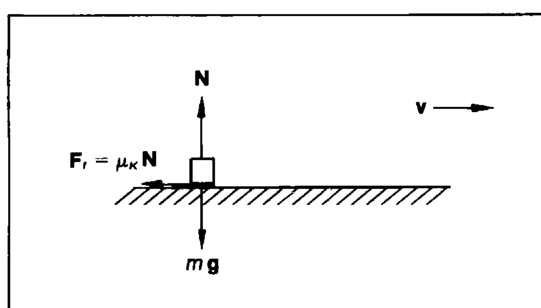


Figure 20-6 Force diagram for Example 6.

20-4 POTENTIAL ENERGY—GRAVITATIONAL AND ELASTIC

In the last section, examples were presented in which conservative forces did work on particles without changing the speed of the particles. Since this means that there is no change of kinetic energy, the particles must have experienced a change in energy associated not with motion but with a change in configuration of the particles. This change of energy is called a potential energy change, abbreviated as $\Delta(\text{P.E.})$

As an illustrative example, consider a particle of mass m subject to the gravitational force $\mathbf{F}_g$ due to a mass M. If the particle is to move a distance s at constant speed, an applied force $\mathbf{F}_{app}$ will be required while m is to move from r_A to r_B, as shown in Figure 20-7. Then Eq. (20-5) gives

$$W_{AB} = \int_A^B (\mathbf{F}_{app} + \mathbf{F}_g) \cdot d\mathbf{s} = \Delta(\text{K.E.}) = 0. \quad (20\text{-}17)$$

This requires that

$$\Delta(\text{P.E.}_g) = \int_A^B \mathbf{F}_{app} \cdot d\mathbf{s} = -\int_A^B \mathbf{F}_g \cdot d\mathbf{s}. \quad (20\text{-}18)$$

This change in potential energy, equal to the negative of the work done by the gravitational force and due to the action of the applied force, is called a gravitational potential energy change. Continuing,

$$\Delta(\text{P.E.}_g) = \int_A^B \mathbf{F}_{app} \cdot d\mathbf{s} = -\int_{r_A}^{r_B} \left(-\frac{GMm}{r^2}\right) dr$$
$$= -GMm\left(\frac{1}{r_B} - \frac{1}{r_A}\right) = (\text{P.E.}_g)_B - (\text{P.E.}_g)_A. \quad (20\text{-}19)$$

Since $r_A < r_B$, $\Delta(\text{P.E.}_g) > 0$, which is consistent with having to do work *on* m to move it further from M. If the applied force is removed ($\mathbf{F}_{app} = 0$), m will move toward M due to $\mathbf{F}_g$. This causes a change in kinetic energy $\Delta(\text{K.E.})$ equal to the work done *by* gravity. Thus,

$$\Delta(\text{K.E.}) = \int \mathbf{F}_g \cdot d\mathbf{s} = -\int_B^A \frac{GMm}{r^2} dr, \quad (20\text{-}20)$$

$$\Delta(\text{K.E.}) = -\int_{r_B}^{r_A} GMm\frac{dr}{r^2} = GMm\left(\frac{1}{r_A} - \frac{1}{r_B}\right). \quad (20\text{-}21)$$

Therefore, we can see that

$$\Delta(\text{K.E.})_{B\to A} = (\text{K.E.})_A - (\text{K.E.})_B = \Delta(\text{P.E.}_g)_{A\to B}. \quad (20\text{-}22)$$

Since

$$\Delta(\text{P.E.}_g)_{A\to B} = (\text{P.E.}_g)_B - (\text{P.E.}_g)_A,$$

we can rearrange Eq. (20-22) to give

$$(\text{K.E.})_A + (\text{P.E.}_g)_A = (\text{K.E.})_B + (\text{P.E.}_g)_B. \quad (20\text{-}23)$$

Equation (20-23) shows that if there is no work done due to applied forces, the sum of the kinetic and gravitational potential energies at A equals the sum of the same quantities at B. In other words, the total energy (kinetic plus potential) of the particle cannot change when no work is done on it by applied forces. This statement, which is called the *principle of conservation of mechanical energy*, is one of the keystones of classical physics. More will be said about it in Section 20-5.

From Eq. (20-19), we obtained

$$(\text{P.E.}_g)_B = -\frac{GMm}{r_B},$$
$$(\text{P.E.}_g)_A = -\frac{GMm}{r_A}. \quad (20\text{-}24)$$

The minus sign in Eq. (20-24) indicates that when m and M are at rest and separated by a finite distance, work must be supplied by an external applied force in order to cause a larger separation. At infinite separation ($r_B \to \infty$), on the other hand, the gravitational potential energy goes to zero.

Example 7. A rocket is launched from the Earth and reaches a maximum height of 8×10^6 m above the Earth.

(a) What was its change of potential energy between launch and this maximum separation, and

(b) what must have been the launch speed if its speed is zero at the height of the flight?

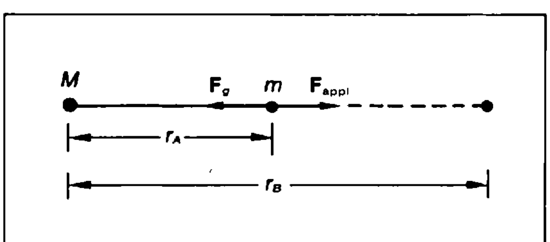


Figure 20-7 Force diagram for a particle undergoing motion at constant speed.

SOLUTION

(a) From Eq. (20-19),

$$\Delta(\text{P.E.}_g) = -\left[\frac{GM_E m}{R_E + \Delta R} - \frac{GM_E m}{R_E}\right]$$

$$= -\frac{GM_E m}{R_E}\left[\frac{1}{1+\dfrac{\Delta R}{R_E}} - 1\right] = +\frac{GM_E m\,\Delta R}{R_E^2\left(1+\dfrac{\Delta R}{R_E}\right)}.$$

In this example,

$$\frac{\Delta R}{R_E} = \frac{(8\times 10^6)\text{ m}}{(6.4\times 10^6)\text{ m}} = \frac{10}{8},$$

and

$$\frac{GM}{R_E^2} = g = 9.8\text{ m/sec}^2,$$

from Section 16-5. Therefore,

$$\Delta(\text{P.E.}_g) = \frac{mg\,\Delta R}{1+\dfrac{\Delta R}{R_E}} = \frac{(m\text{ kg})(9.8\text{ m/sec}^2)[(8\times 10^6)\text{ m}]}{\left[1+\dfrac{10\text{ m}}{8\text{ m}}\right]}$$

$$= [3.49\times 10^7\ m]\text{ joules}.$$

(b) From Eq. (20-23),

$$(\text{K.E.})_0 + (\text{P.E.}_g)_0 = (\text{K.E.})_F + (\text{P.E.}_g)_F$$

or

$$(\text{K.E.})_0 = 0 + (\text{P.E.}_g)_F - (\text{P.E.}_g)_0 = \Delta(\text{P.E.}_g).$$

Therefore,

$$\frac{1}{2}mv_0^2 = [3.49\times 10^7\ m]\text{ joules}$$

$$v_0 = \left[\frac{2(3.49\times 10^7\ m)\text{ joules}}{m\text{ kg}}\right]^{1/2} = 8.35\times 10^3\text{ m/sec}.$$

The reader should note that these calculations are unrealistic in the sense that the mass m of the rocket is *not* constant as assumed, since fuel (and, hence, mass) is expended to achieve the rocket flight. However, if it is reasonably assumed that the rocket "burn" is over in a distance small compared to R_E, then the results will not be greatly in error.

A particle subject to an elastic restoring force experiences a force that is linearly dependent upon the displacement of the particle. That is,

$$\mathbf{F} = -k(\mathbf{r} - \mathbf{r}_0). \qquad (20\text{-}25)$$

The minus sign indicates that, when the particle is displaced from its equilibrium (original) position, the elastic force is directed back toward the equilibrium position. The elastic constant (or spring constant) k has the dimensions of force/length. If one measures the restoring force as a function of displacement, k can be found as the slope of the force versus displacement curve (*if* the curve is linear).

As an example, a mass attached to a light spring and resting on a frictionless surface will experience no force if the spring is neither stretched nor compressed. Either stretching the spring or compressing it by an amount x will result in a force on the mass that is directed toward $x = 0$, the original equilibrium position. The work involved in moving the mass from position x_1 to position x_2 is, by Eqs. (20-5) and (20-25): (Note that $\mathbf{F}\cdot d\mathbf{s} = kx\,dx$.)

$$W_{1\to 2} = \int_{x_1}^{x_2} kx\,dx = \frac{1}{2}kx^2\Big|_{x_1}^{x_2}$$

$$= \frac{1}{2}kx_2^2 - \frac{1}{2}kx_1^2. \qquad (20\text{-}26)$$

If the displacement occurs at constant speed, we can write—as in Eq. (20-18)—that

$$W_{1\to 2} = \Delta(\text{P.E.})_{el} = \int_{x_1}^{x_2} kx\,dx = (\text{P.E.})_2 - (\text{P.E.})_1. \qquad (20\text{-}27)$$

Therefore, elastic potential energy satisfies the relation

$$(\text{P.E.})_{el} = \frac{1}{2}kx^2. \qquad (20\text{-}28)$$

As in the gravitational case, Eq. (20-23) is still valid if no external forces do work on the spring-mass system. Thus, a mass attached to a spring, displaced from $x = 0$ to $x = A$, and held at rest, has a total energy $\frac{1}{2}kA^2$. If it is then released, Eq. (20-23) requires that

$$\text{total energy} = \frac{1}{2}kA^2 = \frac{1}{2}kx^2 + \frac{1}{2}mv^2, \qquad (20\text{-}29)$$

for any other displacement x. We see, therefore, that the (P.E.) is a maximum ($\frac{1}{2}kA^2$) when the (K.E.) is zero, and the (K.E.) is a maximum ($\frac{1}{2}kA^2$) when the (P.E.) is zero (corresponding to $x = 0$). As the motion proceeds, there is a periodic change from maximum displacement to maximum speed as required by Eq. (20-29). The details of this periodic or oscillatory motion are presented in Chapter 22.

In concluding this section, it is worth emphasizing that potential energy is related to work done against conservative forces. By working against conservative forces, a change in the configuration of the system is accomplished and energy is in effect stored in the system and is *potentially* available for subsequent conversion to kinetic energy or to be dissipated as energy lost due to frictional or non-conservative forces.

20-5 CONSERVATION OF MECHANICAL ENERGY

In the discussion of Eq. (20-23), the sum of the kinetic and potential energies of a particle was called the total mechanical energy. Henceforth, we shall use the symbol E for this quantity. Thus,

$$E = (\text{K.E.}) + (\text{P.E.}). \tag{20-30}$$

In a situation where there are no non-conservative forces acting, the value of E will remain constant even though the kinetic and potential energies change. This information can be advantageously displayed graphically. Figure 20-8 shows the potential energy [P.E.(x)] as a function of position x for a particle executing one-dimensional motion. Several horizontal lines, E_1, E_2, and E_3, on the same graph represent different (constant) total mechanical energy states. As indicated by the ordinate scale, $E_1 < E_2 < E_3$. With such a potential energy curve and Eq. (20-30), it is possible to immediately construct the [K.E.(x)] versus x curve since

$$[\text{K.E.}(x)] = E - [\text{P.E.}(x)].$$

Notice that the total energy curves are shown dashed wherever $E < [\text{P.E.}(x)]$. If $E < [\text{P.E.}(x)]$, Eq. (20-30) requires that $[\text{K.E.}(x)] < 0$. Since this would require $\frac{1}{2}mv^2 < 0$, either the mass would have to be negative or v^2 would be negative (v must then be a purely imaginary number in the algebraic sense). Either possibility does not correspond to the physical world, so that we require that (for classical physics at least) $E \geq [\text{P.E.}(x)]$.

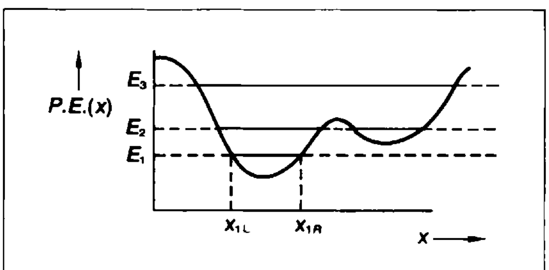


Figure 20-8 Schematic graph of potential energy as a function of position.

For E_1, there are two points (x_{1L} and x_{1R}) where $E_1 = [\text{P.E.}(x)]$ and (K.E.) = 0. Therefore, the velocity of the particle is zero at these points. Between these points (K.E.) > 0 (since $E > [\text{P.E.}(x)]$) and, since $(\text{K.E.}) = \frac{1}{2}\text{m}(\pm v)^2$, we can deduce that for total energy E_1 the particle can move back and forth along the x-axis between the *turning points* x_{1L}, x_{1R}, where $v = 0$. At x_{1L}, there must be a force acting to the right which slows down a particle approaching from the right and which speeds up a particle receding from x_{1L} to the right. Conversely, at x_{1R} the force must act to the left to stop a particle approaching from the left or to speed up a particle receding to the left from x_{1R}. When we calculate the potential energy change associated with the work done against a conservative force, we use the relation (for one-dimensional motion)

$$\int_A^B d[\text{P.E.}(x)] = \Delta(\text{P.E.}) = -\int_A^B F(x)dx. \tag{20-31}$$

This integral relationship is equivalent to the differential relationship

$$F(x) = -\frac{d[\text{P.E.}(x)]}{dx}. \tag{20-32}$$

With this relation, it is possible to sketch (qualitatively at least) the graph of $F(x)$ versus x if [P.E.(x)] versus x is known. Thus, at x_{1L}, $-d[\text{P.E.}(x)]/dx$ is positive (convince yourself this is so!) so that indeed $F(x)$ is directed to the right as we have already deduced on physical grounds. Similarly, we find $F(x)$ is directed to the left at x_{1R}. In summary, given [P.E.(x)] versus x and the value of E for a conservative system, the motion of the system is completely determined. Not only the kinetic energy but also the force acting (and the consequent acceleration) for all allowed positions can be found, frequently by simple graphical analysis.

Notice that for total mechanical energy E_2 there are two possible regions in which oscillatory motion can occur, with a "forbidden" region between them. Where the particle actually moves will depend upon the initial location of the particle, which would have to be specified if the motion is not to be ambiguous. It is left for the reader to deduce [K.E.(x)] versus x and $F(x)$ versus x curves for the energy states E_1, E_2, and E_3 of Figure 20-8.

20-6 THE WORK-ENERGY BALANCE

There are many non-conservative forces that can act upon particles in a physical situation. Thus, it is necessary to develop an expression more general than Eq. (20-30), but one which reduces to Eq. (20-30) when the vector sum of non-conservative forces is equal to zero. We have already associated the work due to conservative forces with a change in the potential energy so that (see Section 20-4)

$$\Delta E = \Delta(\text{K.E.}) + \Delta(\text{P.E.}) = 0. \qquad (20\text{-}33)$$

It seems reasonable, therefore, to assume that any work done by non-conservative forces must be reflected in a change in the total energy ΔE. That is, using W_N to denote work due to non-conservative forces,

$$W_N = \Delta E = \Delta(\text{K.E.}) + \Delta(\text{P.E.})$$

or

$$(\text{K.E.})_A + (\text{P.E.})_A + W_N = (\text{K.E.})_B + (\text{P.E.})_B. \qquad (20\text{-}34)$$

When W_N is positive (as, for example, when a proton experiences an increase in speed with each orbit in a cyclotron), the final energy will exceed the initial energy. For dissipative forces (friction), $W_N < 0$, and the final energy will be less than the original energy. Note that Eq. (20-34) reduces to Eq. (20-30) when $W_N = 0$. Since the conclusions obtained from Eq. (20-34) are physically acceptable, we accept it and refer to it as a work-energy balance. This emphasizes the fact that analyzing dynamical situations via the scalar work-energy balance reduces the problem essentially to a debit-credit type of bookkeeping problem. This is in contrast to the equally valid but usually more tedious analysis required in a direct application of the (vector) laws of motion.

Example 8. A 3 kg block initially at rest is released to slide along a surface of non-simple shape. When it is 2 m below its original position, its speed is observed to be 4 m/sec. (See Figure 20-9.) Find the work done by friction.

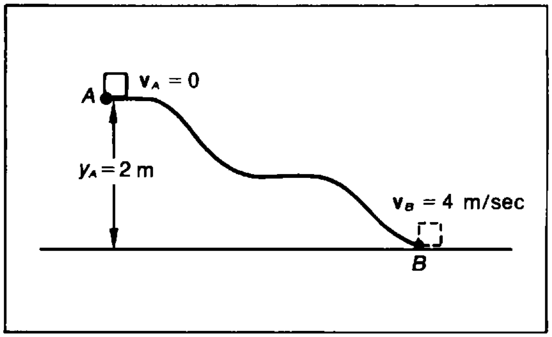


Figure 20-9 Diagram for Example 7.

SOLUTION

Using Eq. (20-34),

$$W_f = (\text{K.E.})_B - (\text{K.E.})_A + (\text{P.E.})_B - (\text{P.E.})_A$$

$$= \frac{1}{2}(3\text{ kg})(4\text{ m/sec})^2 - 0 + 0$$

$$- (3\text{ kg})(9.8\text{ m/sec}^2)(2\text{ m})$$

$$= (24 - 58.8)\text{ joules} = -34.8\text{ joules}.$$

The negative sign indicated that $E_B < E_A$, a result consistent with our experience with frictional systems. Notice that this problem could not be solved using Newton's laws of motion unless additional information were provided. Using Eq. (20-34), however, the solution was straightforward.

PROBLEMS

1. A 50 lb box is pulled 10 ft up a rough plane ($\mu_K = 0.10$) inclined 25° to the horizontal by a force of 100 lb acting parallel to the plane. Find:
   (a) the work done by the 100 lb force,
   (b) the work done against gravity,
   (c) the work done against friction, and
   (d) the increase in kinetic energy of the body.

2. A 40 lb box is given an initial velocity of 30 ft/sec up a surface inclined 30° with the horizontal. Consider the first 10 ft of motion up the incline. Determine:
   (a) the initial kinetic energy of the box,
   (b) the minimum possible change in kinetic energy (for 10 ft), and
   (c) the maximum possible final kinetic energy (at 10 ft).
3. A 10 lb block slides down the frictionless arc of radius 10 ft, starting from rest at the position shown in Figure 20-10. When the block reaches the horizontal portion of the track, it is brought to rest after traveling 20 ft.
   (a) What is the kinetic energy of the block just as it reaches the horizontal portion of the track?
   (b) What is the coefficient of friction between the block and the horizontal portion of the track?

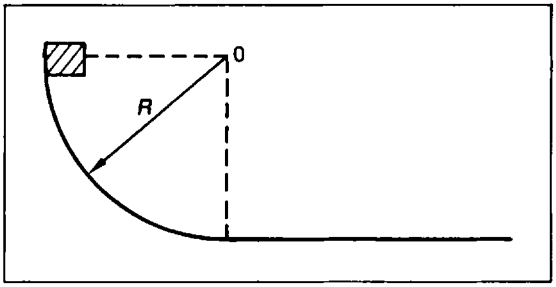


Figure 20-10

4. A force of 230 lb parallel to a plane is required to pull a 300 lb box at constant speed up a plane inclined 30° above the horizontal. The box is pulled a distance of 16 ft along the plane. Determine:
   (a) the total work done in moving the box,
   (b) the work done against friction, and
   (c) the efficiency of this inclined plane.
5. A gun at the edge of a cliff 100 m above sea level fires a 2 kg shell towards the sea at an angle of 53° above the horizontal with an initial velocity of 1000 m/sec. Find:
   (a) the initial kinetic energy and potential energy of the shell,
   (b) its kinetic and potential energy at the highest point of its path, and
   (c) its kinetic and potential energy as its strikes the sea.
6. One force continuously pulls 40 nt horizontally north on a 98 kg block, initially at rest on a frictionless, horizontal surface, while another force pulls 30 nt horizontally east on the same block. Compute the kinetic energy of the block when it has moved 20 m.
7. (a) A walking horse can pull a load equal to one-half his weight. If the horse weights 1600 lb and walks at the rate of 0.50 mi/hr, what horsepower does he develop?
   (b) Calculate the power output for a motor that pulls a car on a horizontal track at a constant velocity of 15 m/sec against a frictional force of 20 nt.
8. A 60 lb body is being pushed a distance 20 ft up an inclined plane by a force P of 50 lb parallel to the plane. The plane rises 2 ft for every 5 ft of incline, and the force of friction between the body and the plane is 18 lb. Compute:
   (a) the work done on the body by P,
   (b) the amount of energy transformed into heat,
   (c) the increase in potential energy of the body, and
   (d) the change in kinetic energy of the body.
9. A box weighing 120 lb is pulled 30 ft up a rough plane inclined 30° to the horizontal by a force of 90 lb parallel to the plane. The coefficient of friction between the box and the plane is 0.20. Find:
   (a) the total work done by the 90 lb force,
   (b) the work done against gravity,
   (c) the work done against friction, and
   (d) the increase in kinetic energy of the box.

10. A 100 lb box starting from rest at the top of a plane inclined 30° with the horizontal slides down the plane until it attains a velocity of 20 ft/sec and has then covered a distance of 22 ft measured along the plane. Find:
    (a) the kinetic energy gained by the box,
    (b) the potential energy lost by the box, and
    (c) the work done against friction while the box was sliding down the plane.
11. A tractor hitched to a load of hay hauls it from the ground to the loft of a barn in 1 minute. If the load weighs 500 lb and the height through which it is lifted is 20 ft, find the power developed in horsepower and watts.
12. A 20 kg block on a plane inclined at 37° with the horizontal is pulled 10 m up the plane by a force P of 250 nt parallel to the plane. Friction between the block and the plane is 40 nt.
    (a) Find the total work done by P.
    (b) Find the work done against friction.
    (c) Find the increase in potential energy of the block.
    (d) Without finding the velocity, find the increase in the kinetic energy of the block.
    (e) Find the velocity of the block when its kinetic energy is 1000 joules.
13. A 2 lb hammer moving with a velocity of 20 ft/sec strikes a nail and drives it in a plank a depth of 1 in.
    (a) Find the average force exerted upon the nail.
    (b) If the operation lasted only $\frac{1}{50}$ second, what was the horsepower of the blow?
14. On a plane making an angle of 37° with the horizontal, the point B is further down the plane than A by a distance of 6 m measured along the plane. A 100 kg block starting at A slides down past B. The coefficient of kinetic friction between the block and the plane is 0.250. Compute:
    (a) the decrease in potential energy of the block from A to B,
    (b) the work done against friction by the block from A to B, and
    (c) the kinetic energy of the block at B.
15. A plank that is 12 ft long has one end resting on the ground and the other end 4 ft higher. On the plank is a trunk weighing 125 lb, which is just kept from sliding by friction.
    (a) How great must the force of friction be?
    (b) If the trunk is pulled up the plane a distance of 10 ft, how much work is done?
    (c) If the previous amount of work was done in 10 seconds, how much horsepower was required?

21 Rotation

21-1 MOMENT OF FORCE† AND CENTER OF GRAVITY

It is appropriate to remind the reader that our discussion of particle motion in the previous seven chapters was limited at the outset to ideal or point-particles (see Section 14-1). As a result, any system of forces applied to such a particle is necessarily a concurrent one—each force acts at the same point, which is the location of the (point-) particle. When the resultant force is tangent to the direction of motion, straight line, rectilinear (or translational) motion results (Chapters 17 and 18). When the resultant force is instead perpendicular to the direction of motion, uniform circular motion results (Sections 18-4 to 18-6). In both cases, the point-particle concept allowed us to ignore the possibility of any intrinsic rotational motion (such as the rotation of the Earth on its axis as it executes orbital motion about the Sun).

If we now consider rigid objects (objects whose non-negligible size and shape does not change) we encounter the possibility that the system of applied forces may not all act at the same point (be concurrent). The resulting motion can then be more complex in that it may involve a combination of translation and rotation. This is the subject of this chapter.

Consider a point-particle of mass m attached to a weightless arm of length r, which is in turn attached at right angles to a long slender rod that can rotate freely about an axis along the rod, as illustrated in Figure 21-1. A force **F** is applied to the point mass m, in order to cause rotation about the rod axis. Now, any component of the force that is parallel to the rod axis or parallel to the supporting arm cannot cause rotation about the axis. On the other hand, an applied force will have maximum rotational effectiveness if it is applied perpendicular to both the supporting arm and the rod axis. Thus, rotation de-

†The term moment of force, like the term moment of momentum, involves in its definition the concept of a "moment arm." When the meaning of "moment arm" is clear, terms such as moment of force are quite descriptive and practical. From the standpoint of the algebra of vector quantities, a "moment arm" is directly related to a cross (or vector) product. In most modern literature, the term moment of force is replaced by the term torque. Similarly, moment of momentum is replaced by angular momentum. Although we mention the early terms for their descriptiveness, we shall conform to the modern usage in what follows.

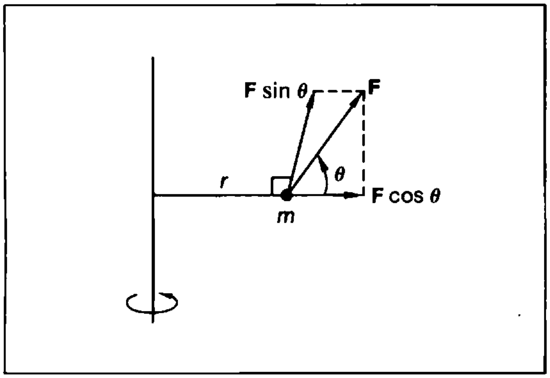


Figure 21-1 Diagram illustrating parameters involved in rotation about an axis.

pends not only on the force magnitude but also on the orientation of the line of action of the force. As a familiar example, the reader may consider the most effective way of causing a door to swing on its hinges. A force applied perpendicular to the plane of the door and to an axis through the hinges will be best.

This example is also useful in pointing out that the effectiveness of a given applied force is increased if the distance between the axis of rotation and the point of application of the force is increased. (The skeptic can test this by comparing the results of pushing normally on the doorknob and then pushing with a normal force of the same magnitude along the inner or hinge edge of the door.) Still another facet of this situation is the fact that reversing the direction of the applied force simply reverses the direction of the rotation around the same axis of rotation. Our problem then is to incorporate all these observations in a relationship that makes possible quantitative predictions for analogous systems and that is compatible with the mathematical requirements the vectors (force and particle displacement) must satisfy. By referring to Figure 21-1, it is seen that only the component of F that is perpendicular to the arm can cause rotation about the axis. If we call this component $F_\perp$, it follows from the figure that

$$F_\perp = F \sin \theta. \tag{21-1}$$

Relative to the rod axis, the point of application (the location of m) of the force **F** has a displacement **r** of magnitude r, directed outward along the arm from the rod axis to the mass m. Now we introduce the quantity *torque* (τ) to denote quantitatively the effectiveness of a force **F** in causing rotation about an axis. In order to display all the necessary properties, the definition of torque can be written as

$$\boldsymbol{\tau} = \mathbf{r} \times \mathbf{F}. \tag{21-2}$$

The expression $\mathbf{r} \times \mathbf{F}$ is a vector (or cross) product in contrast to the scalar (or dot) product we have encountered earlier. As is indicated in the Appendix, the vector product magnitude $\mathbf{r} \times \mathbf{F}$ is given by

$$|\mathbf{r} \times \mathbf{F}| = |\mathbf{r}|\,|\mathbf{F}| \sin \theta = |\mathbf{r}| \sin \theta\, |\mathbf{F}|. \tag{21-3}$$

(The quantity $|\mathbf{r}| \sin \theta$ is referred to in the older literature as the moment arm of the force **F**. Hence, the magnitude of the torque is just the product of the moment arm and the magnitude of the force. Thus, a force applied at the axis has a zero moment arm, yielding a zero torque as required.)

The direction of $\mathbf{r} \times \mathbf{F}$ is perpendicular to the plane containing **r** and **F** according to the convention that if the fingers of the right hand curl so that **r** swings toward **F**, then the thumb points along the direction of $\mathbf{r} \times \mathbf{F}$. It follows from this that reversing the direction of **F** reverses the direction of $\mathbf{r} \times \mathbf{F}$ along the rod axis. Thus, $\mathbf{r} \times \mathbf{F}$ is a product of vectors which yields a third vector perpendicular to the two vectors. We intend that this vector product is to relate to rotational motion. It is therefore clear that for torque to have a unique meaning it cannot lie in the plane containing **r** and **F**, for, once rotation begins, the directions of **r** and **F** can both change, while the sense of the rotational motion remains the same. In addition, when two or more different torques are applied to the mass m, one might expect that the resulting rotational motion is related to the vector sum of the applied torques, just as the resultant translational motion of an object is related to the vector sum of the applied forces. In fact, it is our intent to show in this chapter that (with suitably defined quantities pertinent to rotational motion) Newton's laws of motion are also adequate to describe rotational motion.

Next, we must introduce the quantities necessary to describe rotational kinematics. Not surprisingly, they are (analogous to rectilinear kinematical terms): ***angular displacement*** ($\boldsymbol{\theta}$), ***angular velocity*** ($\boldsymbol{\omega}$), and ***angular acceleration*** ($\boldsymbol{\alpha}$). If the angular displacement changes with time, the following relations are valid:

$$\boldsymbol{\omega} = \frac{d\boldsymbol{\theta}}{dt}. \tag{21-4}$$

$$\alpha = \frac{d\omega}{dt} = \frac{d^2\theta}{dt^2}. \qquad (21\text{-}5)$$

It should be stressed that the vectors $\boldsymbol{\theta}$, $\boldsymbol{\omega}$, $\boldsymbol{\alpha}$ are directed parallel to the axis of rotation about which the particle moves, rather than in the plane of motion.

Suppose the force **F** of Figure 21-1 (perpendicular to the axis of rotation and the supporting arm) remains constant in magnitude and always directed tangent to the circular path of radius r traversed by the mass m. This is shown in Figure 21-2. At any given instant, the tangential acceleration **a** and the force **F** are related by Newton's second law,

$$\mathbf{F} = m\mathbf{a}. \qquad (21\text{-}6)$$

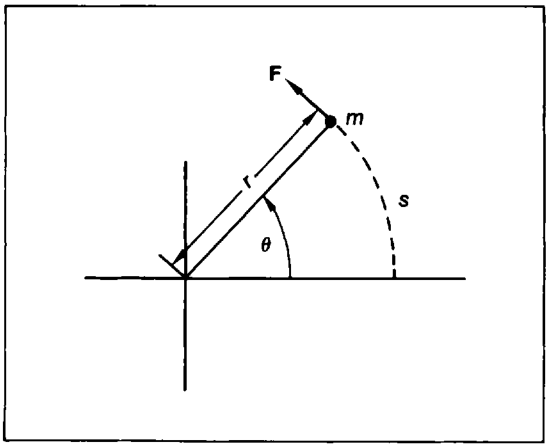


Figure 21-2 Circular motion of a particle about an axis due to a tangential force **F**.

For circular motion of fixed radius, the following scalar relations hold (the reader should verify them from the geometry of Figure 21-2):

$$s = r\theta. \qquad (21\text{-}7)$$

$$v = r\frac{d\theta}{dt} = r\omega. \qquad (21\text{-}8)$$

$$a = r\frac{d^2\theta}{dt^2} = r\frac{d\omega}{dt} = r\alpha. \qquad (21\text{-}9)$$

It is essential in these relations that the angular displacement be measured in radians—abbreviated, rad—with analogous units for ω and α. Often, however, a rotational displacement may be given in degrees—abbreviated, °. From the geometry of a circle:

$$c = 2\pi r,$$

where c is the circumference and r is the radius *in the same units*. Since 2π is the number of radians in one complete revolution—equivalent to 360°— we obtain the conversion relation

$$\theta\,(\text{rad}) = \theta(°) \times \frac{2\pi(\text{rad})}{360\,(°)}.$$

Thus, if the magnitude of **F** is constant, the mass m will experience tangential acceleration, which is also constant in magnitude. From Eq. (21-9), the angular acceleration is also constant in magnitude. Let us now determine the torque $\boldsymbol{\tau}$ due to **F**. Using Eqs. (21-2) and (21-6),

$$\boldsymbol{\tau} = \mathbf{r} \times \mathbf{F} = \mathbf{r} \times (m\mathbf{a}) = m\mathbf{r} \times \mathbf{a}. \qquad (21\text{-}10)$$

The right-hand rule for the cross product shows that $\mathbf{r} \times \mathbf{a}$, like $\boldsymbol{\tau}$, is perpendicular to the plane containing **r** and **F**. Equation (21-9), in addition, requires that the magnitude of the torque is given by

$$\tau = mr^2\alpha. \qquad (21\text{-}11)$$

Equation (21-11) expresses the fact that for a given mass m at fixed radial distance r from an axis of rotation, the angular acceleration and the applied torque are proportional. This is the desired analogy (for translational or rectilinear motion of a given mass m, the acceleration and the applied force are proportional). For rotational motion, the inertial quantity is seen to be dependent upon *both* the mass *and* its radial distance from the axis of rotation. If we introduce the term *moment of inertia* (I) with the relation

$$I = mr^2, \qquad (21\text{-}12)$$

then the rotational form of Newton's second law of motion becomes

$$\tau = I\alpha. \qquad (21\text{-}13)$$

If the net applied torque is zero, it follows that the angular acceleration is also zero. Then the angular velocity is constant (or zero as a special case), which is equivalent to saying that the particle is experiencing uniform circular motion (or else it is at rest).

The reader may recall that Section 17-1 discussed static and dynamic equilibrium of point-particles. We see now that for equilibrium of a real object to exist, we must require not only that the net applied force be zero but also that the net applied torque vanish.

Now consider a real object of weight **W**, which is to be supported at rest at the Earth's surface by a

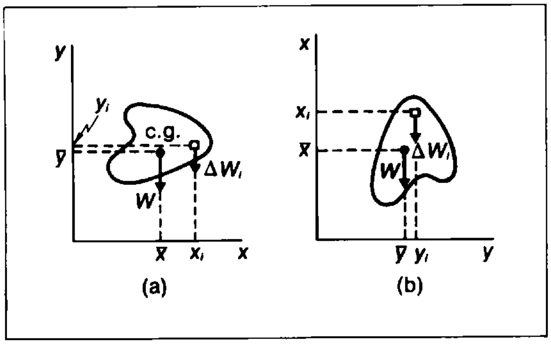


Figure 21-3 Diagrams illustrating the location of the x- and y-coordinates of the center of gravity of a body.

single force F. The problem is to locate where the force should be applied. It is clear from the discussion of Chapter 17 that F must be equal in magnitude but directed opposite to the weight **W**. For complete equilibrium, we must also require that **F** and **W** lie along the same line of action. If this were not the case, rotation about some axis would result, contrary to the conditions given. (Everyone has at one time or another experimentally solved this problem—for example, in the form of stacking blocks or balancing on a narrow rail.) A material object of total weight **W** on the Earth's surface can be regarded as a point-particle of weight **W** located at a point known as the center of gravity of the object. This point is located by determining the sum of torques due to the weight elements making up the body. Thus [see Figure 21-3(a)], we must require that the x-coordinate of the center of gravity $\bar{x}$ satisfy the relation

$$W\bar{x} = \sum_{i=1}^{N} (x_i \Delta W_i), \qquad (21\text{-}14)$$

where $W = \sum_{i=1}^{N} (\Delta W_i)$ = total weight of the object, and $x_i \Delta W_i$ is the torque of the ith element of the object about the z-axis. Similarly, Figure 21-3(b) illustrates the requirement satisfied by the y-coordinate of the center of gravity:

$$W\bar{y} = \sum_{i=1}^{N} (y_i \Delta W_i). \qquad (21\text{-}15)$$

Thus, if the object is supported by a force equal in magnitude to the weight passing through the center of gravity, it will remain at rest. Stated differently, Eqs. (21-14) and (21-15) simply require that the sum of clockwise torques balance the counterclockwise torques about the center of gravity in any orientation of the object.

If the weight of equal volume elements varies continuously throughout the object, Eqs. (21-14) and (21-15) become

$$W\bar{x} = \int x\,dW, \qquad (21\text{-}16)$$

$$W\bar{y} = \int y\,dW, \qquad (21\text{-}17)$$

with

$$W = \int dW, \qquad (21\text{-}18)$$

Example 1. A student has a quantity of laundry to be done in a machine which takes loads of 8 lb each. Although he does not own a set of bathroom scales, he possesses a uniform metal rod 1 m in length weighing 4 lb and a pair of shoes weighing 3 lb. To insure that his machine loads are the maximum weight, he attaches the laundry at one end of the rod, his pair of shoes at the other, and supports the rod at a point such that the rod rests in a horizontal position. How far from the laundry load end is this position?

SOLUTION

Since the rod is uniform, the center of gravity is at the center ($\frac{1}{2}$ m from each end). When the rod is supported at a position x meters from the 8 lb load (see Figure 21-4), it remains horizontal. In equilibrium, the counterclockwise torques must balance the clockwise torques. Thus,

$$(8\text{ lb})(x) = (4\text{ lb})\left[\left(\frac{1}{2} - x\right)\text{m}\right] + (3\text{ lb})[(1-x)\text{m}]$$

or

$$(15\text{ lb})(x) = 5\text{ lb-m}.$$

Therefore,

$$x = \frac{1}{3}\text{m from 8 lb load.}$$

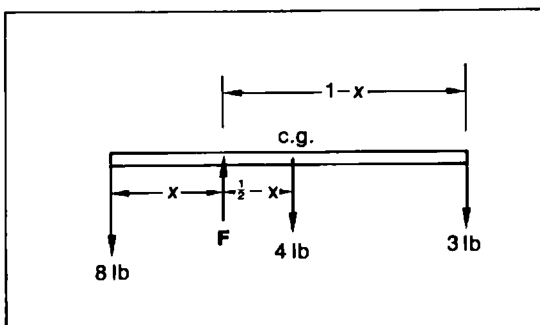


Figure 21-4 Force diagram for Example 1.

21-2 ROTATIONAL MOTION OF AN EXTENDED OBJECT

In the previous section, we analyzed the rotational motion of a point-particle about a fixed axis of rotation at a definite radial distance from the axis. If instead we wish to consider the rotation of an extended rigid object (for example, a thin flat plate), as in Figure 21-5, it is only necessary to subdivide the plate into small elements of mass (dm) and apply the previous analysis to each element. Thus, the magnitude of the torque $d\tau$ required to produce an angular acceleration α for the mass dm a distance r from the axis of rotation through 0 is given by

$$d\tau = dm\, r^2\alpha. \qquad (21\text{-}19)$$

The total torque required is then just the sum (an integral for continuous distributions of mass) of the individual contributions;

$$\tau = \int d\tau = \left(\int dm\, r^2\right)\alpha$$
$$= I\alpha, \qquad (21\text{-}20)$$

where the integration extends over the entire plate, yielding the moment of inertia of the plate. The significance of Eq. (21-20) is that, for a rigid body, the inertial and dynamical parts of the problem have been separated. The moment of inertia calculation is primarily a geometric question (for uniform density). When it is known, the torque required to produce a given angular acceleration (or vice versa) follows immediately.

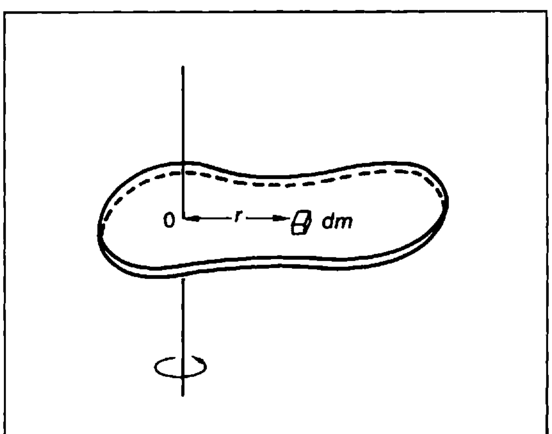


Figure 21-5 Rotational motion of a thin flat plate about an axis ⊥ to the plate.

Example 2. Figure 21-6 represents a uniform slender rod of length L meters, total mass M, and cross-sectional area A. If the rod is to rotate about the axis as indicated with an angular acceleration α, determine the magnitude of the torque required.

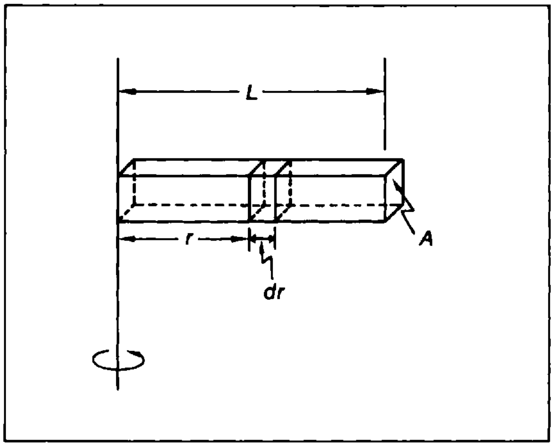


Figure 21-6 Diagram for Example 2.

SOLUTION

From Eq. (21-20) and the discussion preceding it, it is clear that we need only determine I, the moment of inertia, to obtain the desired result. The moment of inertia is given by

$$I = \int r^2\, dm.$$

From Figure 21-6,

$$dm = \frac{M}{V}dV = \frac{M}{LA}(A\,dr) = \frac{M}{L}dr.$$

Since the integral must include all elements of mass from $r = 0$ to $r = L$, we have

$$I = \int_0^L r^2\frac{M}{L}dr = \frac{M}{L}\int_0^L r^2\,dr$$
$$= \frac{M}{L}\frac{r^3}{3}\bigg|_0^L = \frac{ML^2}{3}.$$

Equation (21-23) then gives $\tau = \dfrac{ML^2}{3}\alpha.$

21-3 KINETIC ENERGY OF A ROTATING BODY

To continue the study of rotational motion, we next consider the motional or kinetic energy associated with rotation. Our starting point is the point-particle discussed in Section 21-1. Since the tangential velocity at a given instant has a magnitude v, the kinetic energy at the same instant is given by the rectilinear expression of Eq. 20-12:

$$\text{K.E.} = \frac{1}{2}mv^2,$$

or, equivalently in rotational quantities (since $v = r\omega$),

$$\frac{1}{2}m(r\omega)^2 = \frac{1}{2}mr^2\omega^2 = \frac{1}{2}I\omega^2. \qquad (21\text{-}21)$$

For the point-particle, the kinetic energy for both the rectilinear and rotational forms is thus seen to be one-half the product of an inertial factor and the square of the relevant velocity magnitude.

From Eq. (20-13), we can write

$$\text{K.E.} = \int_A^B \mathbf{F}\cdot d\mathbf{s},$$

where ds is an element of the circular path traversed by m due to the applied force $\mathbf{F}$. If $\mathbf{F}$ is tangent to ds,

$$\begin{aligned}\mathbf{F}\cdot d\mathbf{s} &= F\,ds = ma\,ds\\ &= (mr\alpha)(r\,d\theta) = (mr^2\alpha)(d\theta)\\ &= (I\alpha)\,d\theta = \tau d\theta = \boldsymbol{\tau}\cdot d\boldsymbol{\theta}. \qquad (21\text{-}22)\end{aligned}$$

For the point-particle of Section 21-1, we now can present a table of analogues for describing dynamic and kinematic quantities. Table 21-1 is such a table. It contains additional analogues than will be discussed in Section 21-6.

The reader should be able to extend the discussion of rigid bodies in Section 21-2 to see that the quantities and analogues of Table 21-1 remain valid for the case of extended rigid bodies. Thus, the reader who has mastered the mechanics of translational motion may now obtain twice as much utility from this knowledge by using the table of analogues.

Example 3. The moment of inertia of a wheel is 400 nt m^2. At a given instant, its angular velocity is 10 rad/sec. A constant torque is then applied which increases the angular velocity to 30 rad/sec after the wheel has rotated through an angular displacement of 100 rad. Calculate the magnitude of the applied torque and the work produced by it. Assume no frictional loss of energy.

SOLUTION

For an object of fixed moment of inertia subject to a constant torque, Eq. (21-13) requires that the angular acceleration be constant. Therefore, we must first determine the magnitude of $\boldsymbol{\alpha}$ if $\boldsymbol{\tau}$ is to be found. For constant α, Eqs. (21-4) and (21-5) can be used to find, in complete analogy with Eq. (14-15), that

$$\omega^2 = \omega_0{}^2 + 2\alpha\Delta\theta. \qquad (21\text{-}23)$$

Table 21-1 A Table of Translational and Rotational Analogues

Physical quantity	Translational		Rotational
Displacement	$\mathbf{r}$ (meters)		$\boldsymbol{\theta}$ (radians)
Velocity	$\mathbf{v} = \frac{d\mathbf{r}}{dt}$ (m/sec)		$\boldsymbol{\omega} = \frac{d\boldsymbol{\theta}}{dt}$ (rad/sec)
Acceleration	$\mathbf{a} = \frac{d^2\mathbf{r}}{dt^2}$ (m/sec^2)		$\boldsymbol{\alpha} = \frac{d^2\boldsymbol{\theta}}{dt^2}$ (rad/sec^2)
Inertia	m(Kg)		$I = mr^2$ (kg m^2)
Law of motion	$\mathbf{F} = m\mathbf{a}$ (nt)		$\boldsymbol{\tau} = I\boldsymbol{\alpha}$(nt-m) $= \mathbf{r}\times\mathbf{F}$
Kinetic energy	K.E. $= \frac{1}{2}mv^2$(joules)		K.E. $= \frac{1}{2}I\omega^2$(joules)
Work-kinetic energy relation	ΔK.E. $= \int_a^b \mathbf{F}\cdot d\mathbf{s}$		ΔK.E. $= \int_{\theta_1}^{\theta_2} \boldsymbol{\tau}\cdot d\boldsymbol{\theta}$
Momentum	$m\mathbf{v}$ (kg m/sec)		$I\boldsymbol{\omega}$ (kg m^2/sec)
Impulse-momentum relation	$\int_{t_1}^{t_2} \mathbf{F}\,dt = \Delta(m\mathbf{v})$		$\int_{t_1}^{t_2} \boldsymbol{\tau}\,dt = \Delta(I\boldsymbol{\omega})$
Equilibrium condition	$\sum_i \mathbf{F}_i = 0$	*and*	$\sum_i \boldsymbol{\tau}_i = 0$

Therefore,

$$\alpha = \frac{\omega^2 - \omega_0^2}{2\Delta\theta} = \frac{(30)^2 - (10)^2}{2(10^2)} = 4 \text{ rad/sec}^2,$$

and

$$\tau = I\alpha = 1.6 \times 10^3 \text{ nt-m}.$$

Now the work-energy relation:

$$\Delta\text{K.E.} = \text{work due to } \boldsymbol{\tau} = \int_{\theta_1}^{\theta_2} \boldsymbol{\tau} \cdot d\boldsymbol{\theta}.$$

Thus,

$$\text{work} = \frac{1}{2}I\omega^2 - \frac{1}{2}I\omega_0^2 = 1.6 \times 10^5 \text{ joules}.$$

The reader may already have noted that this result also follows from

$$\text{work} = \int_{\theta_1}^{\theta_2} \boldsymbol{\tau} \cdot d\boldsymbol{\theta} = \tau\Delta\theta$$

(since τ is constant)

$$= 1.6 \times 10^3 \times 10^2 = 1.6 \times 10^5 \text{ joules},$$

as before.

21-4 MOMENTS OF INERTIA

As indicated in Eq. (21-20), the determination of the moment of inertia of an extended rigid body is a calculus problem (assuming the variation of density with position in the body is known). Such calculations are in fact frequently used as practical examples of integration in an elementary calculus course. Thus, in principle, one could proceed in this fashion whenever necessary. However, the world of interest to scientists and engineers is one possessing many kinds of symmetry—cars have circular wheels, axles roll on spherical bearings, diatomic molecules can be viewed approximately as two spherically uniform masses connected by a massless rod (an ideal "dumbbell"), etc. The physicist has usually found that exploiting these symmetries can reduce or even eliminate the integration necessary to obtain a given moment of inertia.

Thus, in Example 2, we found the value of I for the slender rod of Figure 21-6 to be $\frac{1}{3}ML^2$ with respect to an axis perpendicular to the length of the rod and at one end of the rod. Suppose, instead, we have been required to determine the value of I_{CM} (for an axis parallel to the axis at the end but passing through the center of mass CM, of the rod normal to the length of the rod). We could, of course, do the same integration except that the limits of integration would now be from $-L/2$ to $L/2$ instead of from 0 to L. The reader can show that

$$I_{CM} = \frac{1}{12}ML^2.$$

Rather than repeating the first calculation with different limits, we proceed as follows. Recognize that a rod of length L with an axis through the center of mass normal to the rod length is the same as two rods of length $L' = L/2$ with an axis at one end and normal to the length of the rod. Using the result already obtained,

$$I_{CM} = 2I_{L'} = 2\left(\frac{1}{3}\right)\left(\frac{M}{2}\right)\left(\frac{L}{2}\right)^2$$

$$= \frac{1}{12}ML^2,$$

as before.

Now suppose that for a given body we had first calculated I_{CM} and then wished to determine I for an axis parallel to the one through the center of mass but located a distance h from it. We will show that the result is obtained by means of the "parallel-axis theorem,"

$$I = I_{CM} + Mh^2. \tag{21-24}$$

First, however let us test it for the example just presented. In that case, $h = L/2$; it is indeed true that

$$\frac{1}{3}ML^2 = \frac{1}{12}ML^2 + M\frac{L^2}{4}.$$

The proof is not difficult. Figure 21-7 is a cross-sectional view of an arbitrarily shaped object of mass M with the center of mass at point C. If the object rotates with angular speed ω about an axis through point P, its kinetic energy by Eq. (21-21), is

$$\text{K.E.} = \frac{1}{2}I\omega^2, \tag{21-25}$$

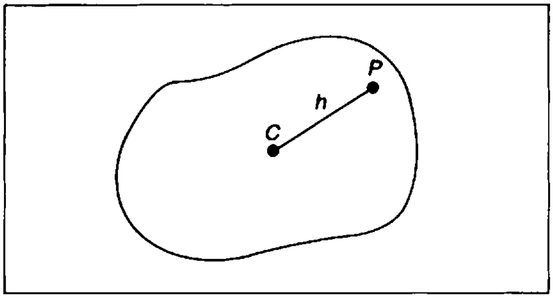


Figure 21-7 Cross-sectional view of an arbitrary object used to illustrate the parallel-axis theorem.

where I is the moment of inertia with respect to the axis through P.

Alternatively, we may view the kinetic energy of the rotating object as composed of two parts: (1) the kinetic energy of a particle containing the total mass M and moving with the center of mass (for which $v_{CM} = h\omega$) about point P, and (2) the rotational kinetic energy computed as though the body were in pure rotation about the center of mass.

For (1), we have

$$(\text{K.E.})_1 = \frac{1}{2}M(h\omega)^2; \qquad (21\text{-}26)$$

for (2),

$$(\text{K.E.})_2 = \frac{1}{2}I_{CM}\omega^2. \qquad (21\text{-}27)$$

Equating Eq. (21-25) with the sum of Eqs. (21-26) and (21-27) yields

$$\frac{1}{2}I\omega^2 = \frac{1}{2}Mh^2\omega^2 + \frac{1}{2}I_C\omega^2$$

or

$$I = Mh^2 + I_{CM},$$

as stated. The reader may wish to execute the proof replacing sums by the appropriate integrals.

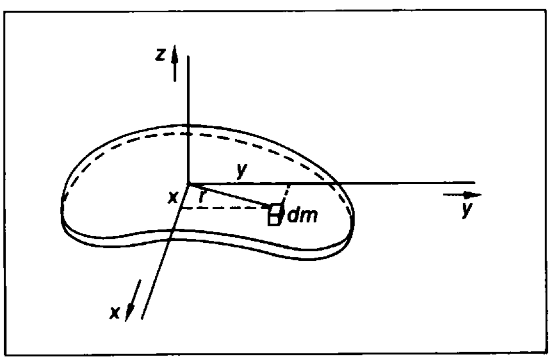


Figure 21-8 The flat plate used to illustrate the perpendicular-axis theorem.

Still another useful device to avoid integration is available for any rigid body whose thickness is negligible compared to its area. (Such bodies are called thin plates or laminar bodies.) The situation is shown in Figure 21-8. The problem is to determine I_z, the moment of inertia with respect to an axis (the z-axis) perpendicular to the plane of the body (the x-y plane). Because the body is thin (all $z_i \approx 0$),

$$\begin{aligned} I_z &= \int r^2\,dm \\ &= \int (x^2 + y^2)\,dm \\ &= \int x^2\,dm + \int y^2\,dm \\ &= I_x + I_y. \end{aligned} \qquad (21\text{-}28)$$

In words, this is a "perpendicular-axis" theorem in that, for laminar bodies, the moment of inertia with respect to an axis perpendicular to the plane of the body equals the sum of two moments of inertia in the plane of the body that are at right angles to each other.

Example 4. Calculate the moment of inertia of a thin uniform circular disk of radius R and thickness t with respect to an axis lying along a diameter of the disk and in the plane of the disk.

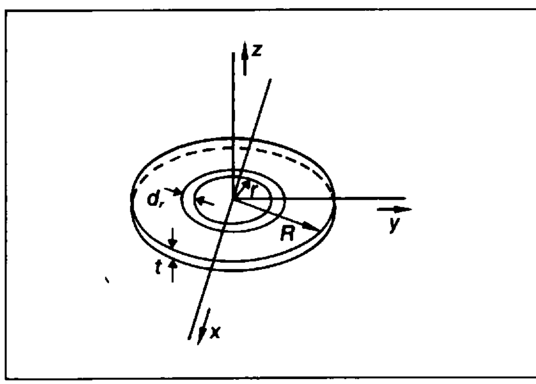


Figure 21-9 Diagram for Example 4.

SOLUTION

Figure 21-9 illustrates the problem. Here $I_x = I_y = I_D$, where I_D is the desired result. Using Eq. (21-28),

$$I_z = 2I_D$$

or

$$I_D = \frac{1}{2}I_z.$$

To calculate I_z, divide the disk into circular segments of thickness t, width dr and volume $2\pi rt\,dr$ at a radial distance r from the z axis. Then

$$dm = M\frac{dV}{V} = M\frac{2\pi rt\,dr}{\pi R^2 t} = \frac{2Mr\,dr}{R^2},$$

therefore,

$$I_z = \frac{2M}{R^2}\int_0^R r^2\,dm = \frac{2M}{R^2}\int_0^R r^3\,dr = \frac{2MR^4}{4R^2} = \frac{1}{2}MR^2.$$

Finally,

$$I_D = \frac{1}{4}MR^2. \qquad (21\text{-}29)$$

The reader may wish to compare this calculation with a direct calculation of I_D. The required choice of elements for dm is reasonably straightforward, but the integration involved should convince the skeptic of the practical value of Eq. (21-28).

In summary, it is possible to calculate many common moments of inertia using Eqs. (21-24) and (21-28) together with a limited number of results calculated directly from the integral definition. Table 21-2 is a modest list of such results.

Finally, just as it is often useful to replace an extended rigid body conceptually by a point-particle of mass M at the center of mass, so it is also useful to locate all the mass at a single distance from the axis of rotation such that the moment of inertia for this arrangement equals that of the extended body. This distance is called the radius of gyration k, which satisfies the defining equation

$$I = Mk^2. \qquad (21\text{-}30)$$

21-5 NEWTON'S LAWS OF MOTION FOR ROTATING BODIES

We are now ready to discuss the dynamics of a system involving a rotating body. Suppose the system also involves translational motion. It is then necessary to first subdivide the problem (in free-body diagrams) into its rotational part and the translational part. The translational part is analyzed using Newton's laws (Chapter 18). The rotational part, by analogy, is analyzed using Eq. (21-13). It should be stressed here that Eq. (21-13) is valid only with respect to an axis through the origin of an inertial reference frame.† Again this is analogous to translational dynamics in that the laws of motion are valid for inertial frames of references only as discussed in Chapter 18.

Consider the system shown in Figure 21-10, composed of a uniform circular cylinder of mass M and radius R, free to rotate about an axis through its center parallel to the axis of the cylinder. Two masses, m_1 and m_2, connected by an inextensible string of negligible mass are hung over the cylinder, and the system is released from rest. Assume that $m_1 > m_2$, the string does not slip on the cylinder, and the bearings supporting the rotation axis are frictionless. When m_1 has dropped a distance h below its original position, what are the linear speeds of m_1 and m_2 and the angular velocity of M?

As indicated in the figure, there are three free-body diagrams: two translational and one rota-

†In rotational motion this is equivalent to rotation about an axis fixed in space or an axis through the center of mass. When the axis of rotation is in translational motion (rolling motion) the analysis is somewhat more involved. We will discuss this case in Section 21-7.

Table 21-2 Moments of Inertia and Radii of Gyration for Some Simple Bodies of Uniform Density.

Body	z-axis	I_z	k_z^2
Slender rod, length L	$\perp$ rod at center	$\frac{1}{12}ML^2$	$\frac{L^2}{12}$
Thin hoop, radius R	$\perp$ plane of hoop, at center	MR^2	R^2
Thin circular ring, radii R_1, R_2	$\perp$ plane of ring, at center	$\frac{1}{2}M(R_1^2+R_2^2)$	$\frac{1}{2}(R_1^2+R_2^2)$
Sphere, radius R	through center	$\frac{2}{5}R^2$	$\frac{2}{5}MR^2$
Thin rectangular plate, sides W, L	$\parallel L$, at center	$\frac{1}{12}MW^2$	$\frac{W^2}{12}$
Thin rectangular plate, sides W, L	$\perp$ plate, at center	$\frac{1}{12}M(W^2+L^2)$	$\frac{(W^2+L^2)}{12}$
Right circular cylinder, radius R, length L	$\perp L$, through center	$M\left(\frac{R^2}{4}+\frac{L^2}{12}\right)$	$\frac{3R^2+L^2}{12}$

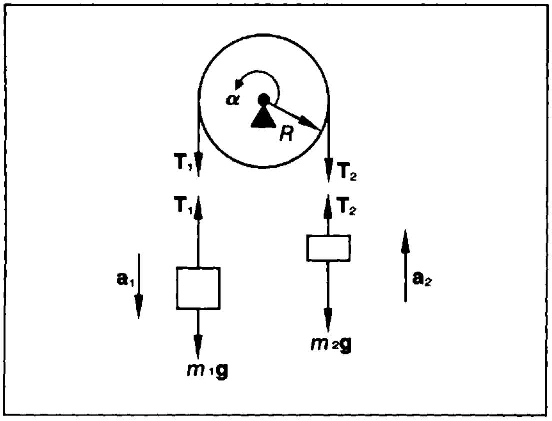


Figure 21-10 Free-body diagrams for a system showing both translational and rotational motion.

tional. Since $m_1 > m_2$, m_1 accelerates downward. Because the string cannot stretch, it follows that the magnitude of the acceleration of m_2 (upward) must be the same as that of m_1. The angular acceleration of the cylinder is similarly required to be counterclockwise, as shown. Using Newton's laws and Eq. (21-13), we obtain (the reader should verify this!):

$$m_1g - T_1 = m_1a, \quad (21\text{-}31)$$

$$T_2 - m_2g = m_2a, \quad (21\text{-}32)$$

$$T_1R - T_2R = I\alpha = \frac{1}{2}MR^2\alpha = \frac{1}{2}MRa. \quad (21\text{-}33)$$

Solving for a, we obtain the result

$$a = \frac{(m_1 - m_2)\,g}{\left(m_1 + m_2 + \frac{M}{2}\right)}. \quad (21\text{-}34)$$

Since the forces (and torques) are constant, we can write (since $v_{10} = v_{20} = 0$, $\omega_0 = 0$)

$$v_1^2 = 2ah = \frac{2(m_1 - m_2)\,gh}{\left(m_1 + m_2 + \frac{M}{2}\right)}$$

or

$$v_1 = -\left[\frac{2(m_1 - m_2)gh}{(m_1 + m_2 + \frac{M}{2}}\right]^{1/2} \quad (21\text{-}35)$$

and

$$v_2 = +\left[\frac{2(m_1 - m_2)gh}{m_1 + m_2 + \frac{M}{2}}\right]^{1/2}. \quad (21\text{-}36)$$

Finally, using either

$$v = R\omega$$

or

$$\omega^2 = 2\alpha\theta \quad \left(\alpha = \frac{a}{R} \text{ and } \theta = \frac{h}{R}\right),$$

yields

$$\omega = \left[\frac{2(m_1 - m_2)gh}{\left(m_1 + m_2 + \frac{M}{2}\right)R^2}\right]^{1/2}. \quad (21\text{-}37)$$

This shows that when $m_1 = m_2$, then $v_1 = v_2 = 0$ and $\omega = 0$ (equilibrium exists). It shows further that the more massive the cylinder when $m_1 > m_2$, the slower will be the motion—certainly a reasonable result.

It is left as an exercise for the reader to solve this problem using the conservation of energy and to show that the same result is obtained.

21-6 ANGULAR MOMENTUM AND ITS CONSERVATION

We saw in Chapter 19 that in some situations an analysis based on Newton's laws of motion was less convenient than an analysis making use of the impulse-momentum theorem. As indicated in Table 21-1, a rotational impulse-momentum theorem can be derived from the rotational second law, Eq. (21-13). Thus,

$$\int_{t_1}^{t_2} \tau\, dt = \int_{t_1}^{t_2} \frac{d(I\omega)}{dt}\, dt = \int_{(I\omega)_1}^{(I\omega)_2} d(I\omega)$$

$$= \Delta(I\omega) = (I\omega)_2 - (I\omega)_1. \quad (21\text{-}38)$$

Here, $\tau = I\alpha$ is replaced with the more fundamental relation

$$\tau = \frac{d(I\omega)}{dt},$$

in analogy with the translational second law discussion of Chapter 19.

Continuing the analogy, Eq. (21-38) states that, in the absence of any external torque on a system undergoing rotational motion, there can be no change in angular momentum. Thus, if there *is* a change in the moment of inertia of the system, it must be accompanied by an appropriate change in the angular velocity.

Example 5. A lecturer demonstrating the conservation of angular momentum stands upon a rotatable platform holding a 16 lb weight in each hand at his sides. The platform is set in rotation with an angular speed of 1 rev/sec. He then extends his arms horizontally, putting the weights at a larger radial distance from the axis of rotation. Assume that frictional effects can be ignored and that the lecturer's moment of inertia is 5 slug-ft^2 independent of the position of his arms. In addition, let the original radial distance of the weights be $\frac{1}{2}$ ft and their final distance 2 ft from the axis of rotation. What will be the final angular speed of the lecturer?

SOLUTION

There are no external torques acting; the forces involved in shifting the weights are radial and contribute no torques to the system. Therefore, from Eq. (21-38),

$$(I\omega)_1 = (I\omega)_2$$

or

$$I_1\omega_1 = I_2\omega_2.$$

Initially,

$$I_1 = I_{\text{lecturer}} + 2(m_{\text{ball}}r_1^2)$$
$$= 5 + 2\left(\frac{16}{32} \times \left(\frac{1}{2}\right)^2\right) = 5.25 \text{ slug-ft.}^2$$

Finally,

$$I_2 = I_{\text{lecturer}} + 2(m_{\text{ball}}r_2^2)$$
$$= 5 + 2\left(\frac{16}{32} \times (2)^2\right) = 9 \text{ slug-ft}^2.$$

Therefore,

$$\omega_2 = \frac{I_1}{I_2}\omega_1 = \frac{5.25}{9} \times 1$$
$$\approx 0.58 \text{ rev/sec.}$$

A more glamorous version of this example is the graceful pirouette executed by an ice skater who initiates a rotational motion with arms outstretched, then clasps her arms tightly to her body, and immediately experiences a much more rapid rate of rotation due to her lower moment of inertia. A similar analysis can be made of the way a diver executes a one turn (or multiple turn) somersault from a diving board. The reader can supply the details in this case.

21-7 ROLLING MOTION

To conclude this chapter, we now consider rotation about an axis that is in translational motion, or what is commonly called rolling motion. It is necessary to use both the translational and the rotational relationships simultaneously. Either type of motion can be analyzed from the point of view of the laws of motion, the impulse-momentum theorem, or energy considerations. It should not be surprising, therefore, that rolling motion can be analyzed in several alternative ways. To illustrate some of the possibilities, consider the motion of a uniform cylinder of mass M and radius R, released from rest, to roll without slipping down an inclined plane of incline angle θ, as in Figure 21-11. What will the acceleration of the center of mass of the cylinder be when it has moved a distance s down the incline?

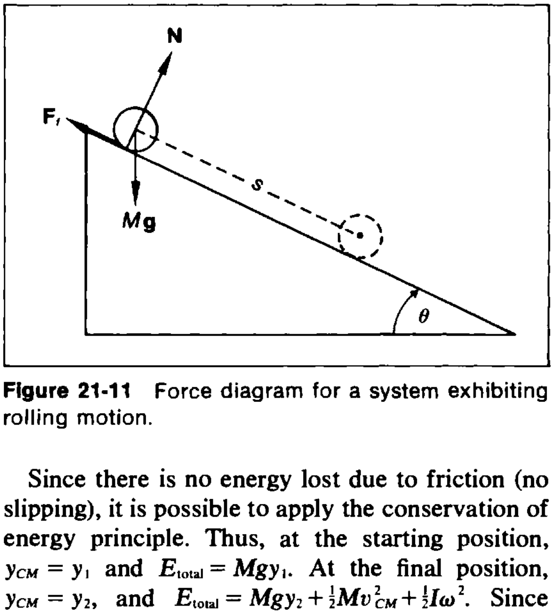


Figure 21-11 Force diagram for a system exhibiting rolling motion.

Since there is no energy lost due to friction (no slipping), it is possible to apply the conservation of energy principle. Thus, at the starting position, $y_{CM} = y_1$ and $E_{\text{total}} = Mgy_1$. At the final position, $y_{CM} = y_2$, and $E_{\text{total}} = Mgy_2 + \frac{1}{2}Mv_{CM}^2 + \frac{1}{2}I\omega^2$. Since energy is conserved,

$$Mg(y_1 - y_2) = \frac{1}{2}Mv_{CM}^2 + \frac{1}{2}\left(\frac{1}{2}MR^2\omega^2\right). \quad (21\text{-}39)$$

The reader can show from Figure 21-11 that

$$y_2 - y_1 = s \sin \theta \quad (21\text{-}40)$$

and

$$v_{CM} = R\omega. \quad (21\text{-}41)$$

By combining Eqs. (21-39), (21-40), and (21-41), we obtain

$$v_{CM}^2 = 2\left(\frac{2}{3}g \sin \theta\right)s. \quad (21\text{-}42)$$

This relation is in a familiar form, which relates an initial velocity of the center of mass ($v_{CM_0} = 0$ in this case), the final velocity of the center of mass, the displacement of the center of mass, and the acceleration *down the incline* of the center of mass. That is,

$$v^2 = v_0^2 + 2as,$$

from which we conclude that

$$a_{CM} = \frac{2}{3} g \sin\theta. \qquad (21\text{-}43)$$

(Notice that had the cylinder slid without rolling, the acceleration would have been $g \sin\theta$, a larger result.)

Next, consider the same situation from the context of the laws of motion. Considering rotation about the center of mass, Eq. (21-13) requires that

$$F_f R = I\alpha \qquad (21\text{-}44)$$

(we have chosen clockwise torques as positive). For the translational motion, we have:

perpendicular to the incline the sum of forces is

$$N - Mg\cos\theta = 0, \qquad (21\text{-}45)$$

parallel to the incline the sum of forces is

$$Mg\sin\theta - F_f = Ma_{CM}, \qquad (21\text{-}46)$$

with *down* the incline taken as positive. By differentiating Eq. (21-41) with respect to t, we obtain

$$a_{CM} = R\frac{d\omega}{dt} = R\alpha. \qquad (21\text{-}47)$$

Combining Eqs. (21-44), (21-46), and (21-47),

$$Mg\sin\theta - \frac{Ia_{CM}}{R^2} = Ma_{CM} \qquad (21\text{-}48)$$

or

$$a_{CM} = \frac{Mg\sin\theta}{M + \frac{I}{R^2}} = \frac{Mg\sin\theta}{M + \frac{\frac{1}{2}MR^2}{R^2}}$$
$$= \frac{2}{3} g \sin\theta,$$

as before.

Notice in this calculation that no use was made of Eq. (21-45). It is *not* possible, for example, to calculate the coefficient of static friction since this would imply that slipping was about to occur and there is no evidence either favorable or unfavorable to support such an assumption.

PROBLEMS

1. A wheel of 10 cm in radius turns through an angle of 500 rad in 10 seconds starting from rest. The acceleration is constant. Calculate:
   (a) the average angular velocity,
   (b) the angular acceleration,
   (c) the final angular velocity after 10 seconds,
   (d) the tangential acceleration,
   (e) the tangential velocity of a particle on the rim after 10 seconds, and
   (f) the centripetal acceleration of the particle after 10 seconds.
2. A car goes around a curve of 121 ft radius at a speed of 30 mi/hr. What must be the coefficient of friction between the wheels and the level pavement in order that the car does not skid?
3. A string is wound around the rim of a cylindrical grindstone of mass 2 slugs and radius 8 in. By pulling on the end of the string, a man exerts a constant tangential force of 3 lb. Assume the bearings are frictionless.
   (a) Find the magnitude of the torque applied to the grindstone.
   (b) Find the magnitude of the angular acceleration of the grindstone.
   (c) Find the angular velocity of the grindstone 6 seconds after it starts from rest.
   (d) What is the kinetic energy of the grindstone 6 seconds after it starts from rest?
   (e) How far does the man pull the free end of the string in 6 seconds starting from rest?
4. (a) On a horizontal frictionless surface, a 500 gm body revolves in a circle whose radius is 90 cm. Find the magnitude of the centripetal force on the body if it makes 1 rev/sec.
   (b) If the body in (a) should accelerate 0.50 rad/sec^2 from the given velocity, what would its angular velocity be at the end of 10 seconds?

5. A figure skater with arms outstretched starts to rotate with angular velocity of 1 rad/sec and then immediately brings in her arms. Given that her initial and final radii of gyration are 20 cm and 5 cm, respectively, what is her resulting angular velocity?
6. On a horizontal turntable a 100 gm lead block lies whose center of gravity is 40 cm from the axis. The angular velocity of the turntable is very slowly increased until it becomes 0.50 rev/sec, at which instant the block begins to slip. Find:
   (a) the radial acceleration of the block at this time,
   (b) the centripetal force on the block at this time, and
   (c) the coefficient of static friction between the block and the turntable.
7. A car starts from rest on a circular race track of radius 1000 ft and increases its speed at the rate of 4 ft/sec^2.
   (a) How long a time will be required to attain the speed at which the tangential and radial components of the car's acceleration are equal?
   (b) How far does the car move along the track before the radial and tangential components of its acceleration are equal?
8. A 40 gm ball on the end of a string revolves in a horizontal circle of 50 cm radius on smooth ice at a uniform speed of 60 cm/sec.
   (a) Find the angular velocity of the ball.
   (b) Compute the tension in the string.
   (c) If the ball should decelerate uniformly at 0.03 rad/sec^2, through how many radians would it travel before it stopped?
9. A light rigid rod has point masses of 5 kg and 10 kg at 3 m and 1 m, respectively, from the supporting frictionless pin P about which it is free to rotate in a vertical plane.
   (a) If the rod is released with negligible angular velocity from the position shown in Figure 21-12, find the angular velocity with which the 5 kg mass passes the horizontal line through P.
   (b) Calculate the angular acceleration of the rod at the moment that the rod becomes horizontal.

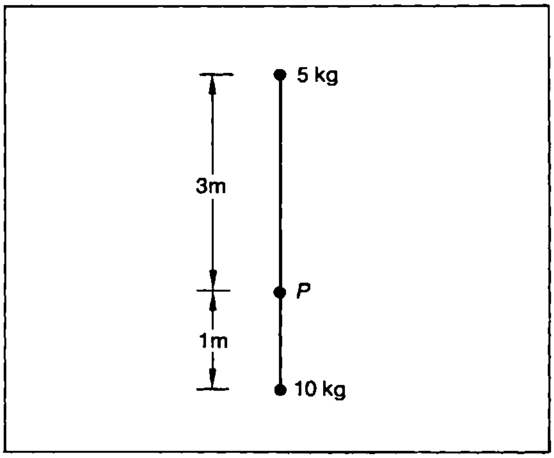


Figure 21-12

10. A wheel 5 ft in diameter starts from rest and acquires an angular speed of 720 rev/min in 12 seconds. A rivet on the rim weighs $\frac{1}{4}$ lb. Compute:
    (a) the angular acceleration,
    (b) the linear acceleration of the rivet,
    (c) the instantaneous tangential velocity of the rivet 12 seconds from the start,
    (d) the radial acceleration of the wheel 12 seconds from the start, and
    (e) the centripetal force exerted on the rivet.
11. (a) An airplane in turning should not be subjected to an acceleration greater than 7 g. Find the radius of the smallest possible circle when flying at 350 mi/hr.

(b) A car with wheels of 2 ft diameter is traveling 20 ft/sec. Find the centripetal force necessary to prevent 1 lb of mud from leaving the rim of the wheel.

12. Disks A and B are mounted on the same shaft so that they may rotate at different speeds or may be connected together by a clutch to rotate at the same speed. Initially, A has an angular velocity ω_0 rad/sec and a kinetic energy of 400 ft-lb, while B is stationary. Then they are coupled together, and heat is produced in the clutch. The moment of inertia of B is 3 times the moment of inertia of A. (a) Find the ratio of the final angular velocity of the two disks to the initial angular velocity of A. (b) Compute in ft-lb the heat produced in the clutch.

13. A stationary disk of mass 0.10 kg and radius 0.10 m, free to rotate about an axis through its center, has a little projection on its circumference. A 0.02 kg bit of putty with a velocity of 5 m/sec strikes the projection tangentially and sticks to it. What is the angular velocity given to the disk?

14. In Figure 21-13, a weight W is attached to a light string wrapped around a solid cylinder of mass M, mounted on a frictionless axle at 0. If the weight starts from rest and falls a distance h, show that its tangential speed is given by

$$v = \sqrt{\frac{2Wh}{\frac{W}{g}+\frac{M}{2}}}.$$

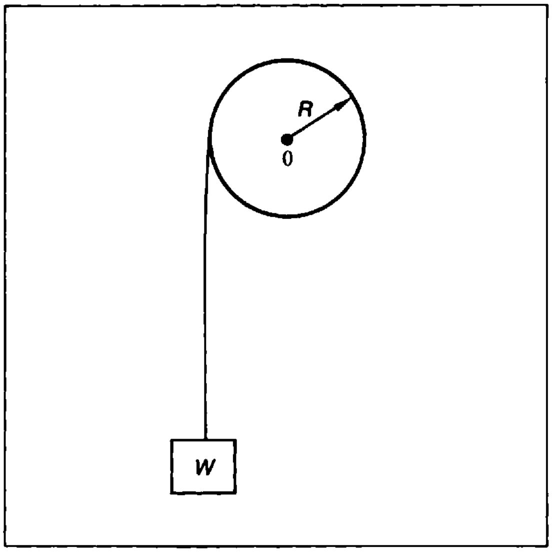


Figure 21-13

15. Starting from rest, a small car of mass m rolls down the incline of the "loop-the-loop" shown in Figure 21-14. It starts from a height H of 32 ft, where $R = 8$ ft is the radius of the loop. Find the magnitude of the velocity of the car as it passes point A on the top of the loop. Neglect friction.

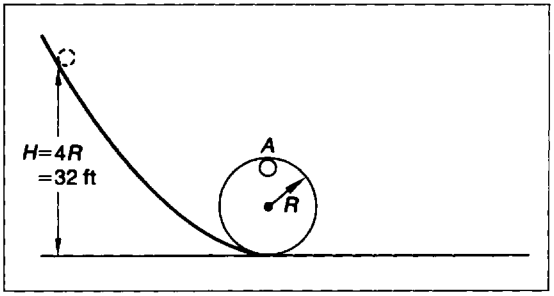


Figure 21-14

16. (a) Find the moment of inertia about an axis perpendicular to the axis of a slender rod of uniform cross-sectional area A and length L, if the density varies according to the relation

$$\rho = a + br,$$

where a and b are constants, and r is the distance along the rod from one end.

(b) A flywheel 3 ft in diameter is pivoted on a horizontal axis. A rope is wrapped around the outside of the flywheel and a steady pull of 10 lb is exerted on the rope. It is found that 24 ft of rope are unwound in 4 seconds. What was the moment of inertia of the flywheel?

17. Consider a circular ring of outer radius R_1, inner radius R_2, with a thickness t, and of uniform density ρ. As indicated in Table 21-2, the moment of inertia about an axis ⊥ to the plane of the ring and through the center is given by $\frac{1}{2}M(R_1^2 + R_2^2)$. Show that this result can be obtained without integration by noting (as in Example 4) that for a solid circular disk ($R_2 = 0$), $I = \frac{1}{2}MR_1^2$, and considering the ring as a solid disk of radius R_1 from which another solid disk of radius R_2 has been removed. That is, $I_{\text{Ring}} = I_{\text{disk}}(R_1) - I_{\text{disk}}(R_2)$. Note: the masses of the two rings must be calculated from the expression $M = \rho V = \rho\pi R^2 t$.
18. Use the perpendicular-axis theorem to verify the expression for the moment of inertia of a thin rectangular plate of sides W and L about an axis ⊥ to the plate and through the center, given the validity of the preceding result in Table 21-2.
19. The moment of inertia of a sphere of radius R, mass M, about an axis through the center, is given by $I_{CM} = \frac{2}{5}MR^2$. Show that the moment of inertia for the sphere about an axis tangent to the sphere is $I = \frac{7}{5}MR^2$.

22 Harmonic Motion

22-1 PERIODIC MOTION

One of the physical concepts discussed in a number of earlier chapters (beginning with Chapter 10) had to do with repetitive phenomena. Thus, a transverse sinusoidal wave propagating along a string has the principal feature that any given particle of the string will take on a continuous sequence of displacements that lie between two extreme values symmetrically located on either side of the equilibrium position. When the particle has gone through the entire sequence of displacement values, it repeats the sequence or cycle of events again and again, as long as the same wave motion is maintained. As we have seen, the extreme displacement value is called the amplitude of the wave motion and the time required for one cycle of the motion is the period.

In the present chapter, we will discuss the nature of forces that can produce such repetitive motion, which is generally referred to as periodic motion. We shall be concerned primarily with sinusoidal motion, but the reader should understand that a more general periodic motion such as the "saw-tooth" wave form illustrated in Figure 22-1 can be discussed similarly (although the mathematics is necessarily more involved). This can be partially understood if it is recalled from Section 10-2 that any periodic disturbance can be expressed as an infinite series of sinusoidal disturbances. Such

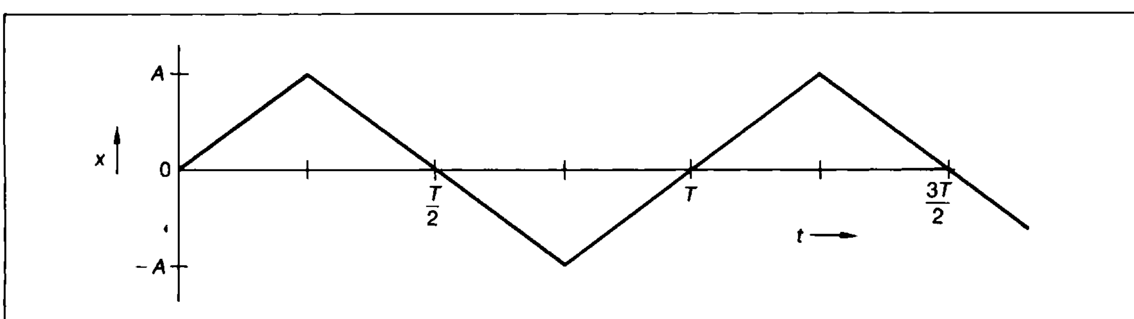


Figure 22-1 Displacement versus time for a "saw-tooth" wave.

series are called Fourier series, and they are extremely useful in the analysis of general periodic motion.

22-2 SIMPLE HARMONIC MOTION

A particle is said to experience simple harmonic motion (SHM) if the displacement-time relationship is a sinusoidal one. For example, one-dimensional SHM is given by

$$x = A \sin\left(\frac{2\pi t}{T} + \phi\right), \qquad (22\text{-}1)$$

where ϕ is the phase angle and is dependent upon initial conditions. Since $T = 1/f$, where f is the frequency, and $2\pi f = \omega$, Eq. (22-1) can also be written

$$x = A \sin(\omega t + \phi). \qquad (22\text{-}2)$$

We wish to know the kind of force that could cause a particle of mass m to experience the SHM of Eq. (22-2). This is accomplished by recalling that the velocity and acceleration of the particle are given by dx/dt and d^2x/dt^2, respectively. That is,

$$v_x = \frac{dx}{dt} = \omega A \cos(\omega t + \phi) \qquad (22\text{-}3)$$

and

$$a_x = \frac{dv_x}{dt} = \frac{d^2x}{dt^2} = -\omega^2 A \sin(\omega t + \phi), \qquad (22\text{-}4)$$

or

$$a_x = -\omega^2 x, \qquad (22\text{-}5)$$

where Eq. (22-5) follows from Eqs. (22-2) and (22-4). The force F_x, which is responsible for the motion described by Eq. (22-2), can be determined from Newton's second law:

$$F_x = ma_x = -m\omega^2 x. \qquad (22\text{-}6)$$

Since $m\omega^2$ is a constant, we can rewrite Eq. (22-6) as

$$F_x = -kx, \qquad (22\text{-}7)$$

where

$$k = m\omega^2. \qquad (22\text{-}8)$$

The meaning of Eq. (22-7) is clear: a particle experiencing SHM does so because of the existence of a force F which is proportional, but directed opposite, to the displacement of the particle. That is, the direction of F_x is always toward the equilibrium position, and is greater in magnitude the further from equilibrium the particle is located. This kind of force is known as a linear restoring force or a Hooke's law force because the magnitude of the force is a linear function of the displacement and because Robert Hooke first stated the relationship. Very few physical systems exhibit true SHM. (Frictional effects in most systems cause the oscillations to die out.) However, the simple harmonic oscillator has proved to be a reasonable starting approximation for a wide variety of situations, from a simple pendulum to the small oscillations of one atom in a diatomic molecule due to the force exerted on it by the other atom.

22-3 EXAMPLES OF SIMPLE HARMONIC MOTION

One of the simplest examples of SHM that can be studied experimentally consists of a mass m connected to a helically coiled spring (of negligible mass), fastened to a rigid support as in Figure 22-2. (See Section 20-4.) It is assumed that the surface on which the mass moves is frictionless, and that the amplitude of the motion of the mass is small, so that the restoring force due to the spring is not altered by distortion of the spring. (For large amplitude, a coiled spring will suffer a permanent stretch, as a result of which the restoring force is also changed.) At equilibrium, the position of the mass relative to the origin at the wall at left is x_0, so that the displacement from equilibrium is given by $x' = x - x_0$, where x is the displacement of m from the origin. If $x' = A$ at $t = 0$, then the motion is described by

$$x' = A \cos \omega t, \qquad (22\text{-}9)$$

where ω is given by Eq. (22-8). For a spring, the constant k is known as the force constant of the spring, or simply the spring constant. Experimentally, it is determined by measuring the force P necessary to cause a given stretch x' of the spring,

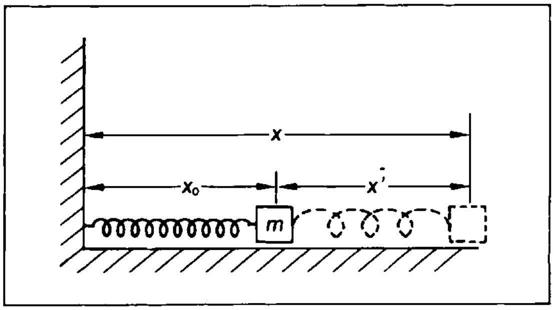


Figure 22-2 An example of SHM.

since the applied force is equal in magnitude but opposite in direction to the restoring force due to the coiled spring. It follows from Eq. (22-7) that

$$k = \frac{-F_x}{x'} = \frac{P}{x'}. \tag{22-10}$$

An experimental test for the suitability of SHM conditions can be made by comparing the value of k determined statically by Eq. (22-10) and the value obtained dynamically by finding the period τ of the motion and using $\omega = 2\pi/\tau$ in Eq. (22-8).

As another example, consider the simple pendulum shown in Figure 22-3. For a simple pendulum, it is understood that the dimensions of the mass m are small compared to the length l of the supporting cord, which has a negligible mass. Applying Newton's second law along the string and perpendicular to it, we have

$$T - mg\cos\theta = 0,$$
$$-mg\sin\theta = F_s = ma_s. \tag{22-11}$$

For motion along the arc of the circular path, we can write

$$s = l\theta, \tag{22-12}$$

from which

$$a_s = l\frac{d^2\theta}{dt^2}, \tag{22-13}$$

since l is a constant. Combining Eqs. (22-11) and (22-13) gives

$$ml\frac{d^2\theta}{dt^2} = -mg\sin\theta$$

or

$$\frac{d^2\theta}{dt^2} = -\frac{g}{l}\sin\theta. \tag{22-14}$$

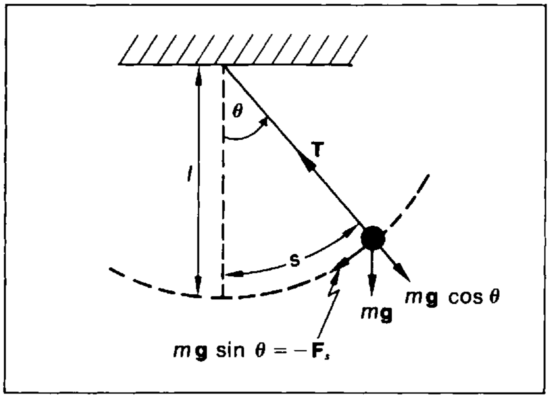


Figure 22-3 Force diagram of a simple pendulum.

This equation is complicated and leads to equally complicated solutions without further assumptions. Therefore, let us consider oscillations for which θ is small enough to permit use of the approximation $\sin\theta \approx \theta$. If this is done, Eq. (22-14) becomes

$$\frac{d^2\theta}{dt^2} \approx -\left(\frac{g}{l}\right)\theta, \tag{22-15}$$

which is identical in form to Eq. (22-5). It follows, therefore, that the solution to Eq. (22-15) is

$$\theta = \theta_m \sin(\omega t + \phi), \tag{22-16}$$

where θ_m is the amplitude of this *rotational* SHM, and

$$\omega = \frac{2\pi}{\tau} = \sqrt{\frac{g}{l}} \tag{22-17}$$

by comparison with Eqs. (22-5) and (22-8). [The skeptical reader is invited to show for himself by substitution that Eq. (22-16) is the solution of Eq. (22-15).] Rearranging Eq. (22-17) leads to the result that

$$\tau = 2\pi\sqrt{\frac{l}{g}}. \tag{22-18}$$

We conclude, therefore, that the period of a simple pendulum is independent of the amplitude of the oscillations *for small oscillations*. A more sophisticated analysis shows that the solution to Eq. (22-14) can be written in the form

$$\tau = 2\pi\sqrt{\frac{l}{g}}\left\{1 + \frac{1}{4}\sin^2\theta_m + \frac{9}{64}\sin^4\theta_m + \cdots\right\}. \tag{22-19}$$

It is left to the reader to verify that for amplitudes up to $\theta_m \le 25°$, Eq. (22-18) is in error by less than 1%. Because the period of the simple pendulum is independent of the amplitude of oscillation, it is a convenient device for keeping time. Alternatively, by means of an auxiliary time-keeping device, the simple pendulum can be used to determine the value of g, the acceleration due to gravity.

22-4 ENERGY RELATIONSHIPS

Throughout the discussions above, it has been assumed that there are no frictional (dissipative) forces. When this is true, it is clear from Chapter 20 that the principle of conservation of mechanical energy can be applied. For the examples of the last section, this means that the total energy of a system exhibiting SHM must remain a constant. For the mass attached to the spring of Figure 22-2, the total energy is partly kinetic energy and partly potential energy stored in the coiled spring, in amounts that are sinusoidal functions of time.

Thus, the kinetic energy is given by

$$\text{K.E.} = \frac{1}{2}mv_{x'}^2 = \frac{1}{2}m\left(\frac{dx'}{dt}\right)^2$$

$$= \frac{1}{2}m\omega^2 A^2 \sin^2 \omega t$$

$$= \frac{1}{2}mv_m^2 \sin^2 \omega t = (\text{K.E.})_{\max} \sin^2 \omega t \qquad (22\text{-}20)$$

from Eq. (22-9). In the last expression, v_m is the maximum speed of the mass m.

From Eq. (20-28), the elastic potential energy of a stretched spring is given by

$$\text{P.E.} = \frac{1}{2}kx'^2. \qquad (22\text{-}21)$$

Therefore, from Eqs. (22-8) and (22-9), we obtain

$$\text{P.E.} = \frac{1}{2}kA^2 \cos^2 \omega t = \frac{1}{2}m\omega^2 A^2 \cos^2 \omega t$$

$$= (\text{P.E.})_{\max} \cos^2 \omega t. \qquad (22\text{-}22)$$

The total energy E for the system is the sum of kinetic and potential energy expressions, or

$$E = \text{K.E.} + \text{P.E.}$$

$$= \frac{1}{2}m\omega^2 A^2(\sin^2 \omega t + \cos^2 \omega t)$$

$$= \frac{1}{2}m\omega^2 A^2 = (\text{K.E.})_{\max} = (\text{P.E.})_{\max}, \qquad (22\text{-}23)$$

which is a constant for fixed m and k.

It is clear, then, that the transformation of energy from a kinetic form to a potential form is beautifully symmetric in cases of SHM. For maximum displacement from equilibrium, the system is at rest and the total energy is potential in form. At the equilibrium position, there is no restoring force so the potential energy vanishes while the speed of the system as it goes through the equilibrium position becomes a maximum, giving maximum kinetic energy. At any other position, the total energy is distributed in the two forms. Alternatively, we may say that the kinetic and potential energies differ in phase by $\pi/2$ or 90°.

Example 1. A mass $m = 0.1$ kg is attached to a helical spring of spring constant $k = 4$ nt/m as in Figure 22-2. If the mass is pulled out 0.10 m from the equilibrium position and released, determine:

(a) the period of the resulting SHM,

(b) the total energy E of the system,

(c) the earliest time for which the kinetic energy equals the potential energy, and

(d) the displacement from equilibrium for which the kinetic energy is one-half as large as the potential energy.

SOLUTION

(a) From Eq. (22-8),

$$\tau = \frac{2\pi}{\omega} = 2\pi\sqrt{\frac{m}{k}} = 2\pi\sqrt{\frac{0.1\text{ kg}}{4\text{ nt/m}}} = 1\text{ sec}.$$

(b) $E = \frac{1}{2}kA^2 = \frac{1}{2}(4\text{ nt/m})(0.1\text{ m})^2 = 0.02$ joule.

(c) K.E. = P.E. = $E/2$, when $\sin^2 \omega t = \cos^2 \omega t = \frac{1}{2}$ or $\omega t = \pi/4$. Therefore, $t = \frac{1}{8}\tau \approx 0.12$ sec.

(d) If K.E. = $\frac{1}{2}$P.E., then

$$E = \frac{1}{2}\text{P.E.} + \text{P.E.} = \frac{3}{2}\text{P.E.}$$

As a result, $\frac{1}{2}kA^2 = \frac{3}{2}(\frac{1}{2}kx^2)$. Therefore,

$$x^2 = \frac{2}{3}A^2$$

or

$$x = \pm\sqrt{\frac{2}{3}}A \approx \pm 0.82A = \pm 0.082\text{ m},$$

since $A = 0.1$ m.

22-5 DAMPED OSCILLATIONS

Let us now turn to the effect of friction forces on a system subject to a linear restoring force. Since most freely oscillating systems do not continue to oscillate indefinitely, it is clear that some dissipative mechanism must be included to account for the damping (steady reduction of oscillation amplitude) of the motion. The exact nature of the friction force is a complex problem depending to a great extent on the specific situation involved. Thus, a mass suspended from a helical spring and immersed in a beaker of oil will not experience the same frictional forces as a mass fastened to a similar helical spring and moving horizontally along a surface that is rough enough to provide a drag on the mass. However, it is an experimental fact that in many cases one can obtain a satisfactory analysis of the motion by assuming that the friction force is proportional to the speed of the oscillating object and is opposite in direction to the motion of the object. Therefore, we express the frictional force $\mathbf{F}_f$ in the form

$$\mathbf{F}_f = -K\frac{d\mathbf{r}}{dt}, \qquad (22\text{-}24)$$

where K is a proportionality constant that is strongly dependent upon the system being considered, and $d\mathbf{r}/dt$ is the velocity of the particle. For motion in one dimension, we can omit the vector notation and write, for example,

$$F_f = -K\frac{dx}{dt}. \qquad (22\text{-}25)$$

Applying Newton's second law to this situation gives the equation

$$m\frac{d^2x}{dt^2} = -kx - K\frac{dx}{dt}. \qquad (22\text{-}26)$$

The solution of this equation is a straightforward exercise in differential equations, but we prefer to use physical arguments to infer the form of the possible solution(s) to it. Thus, let us suppose that the linear restoring force could somehow be neglected in comparison with the friction force. This could be done, for example, if the system involved a mass suspended from a helical spring with a weak spring constant and immersed in a highly resisting fluid such as molasses. If this is true, it is not difficult to deduce that the particle will approach the equilibrium position exponentially with time. That is,

$$x = Ae^{-ct}, \qquad (22\text{-}27)$$

where c is a constant, is a solution to the equation

$$m\frac{d^2x}{dt^2} \approx -K\frac{dx}{dt}.$$

On the other hand, we have seen that SHM results if the frictional force is neglected. It is tempting to suppose that a solution to Eq. (22-26) would be a trigonometric expression (for the SHM) multiplied by a decreasing exponential (due to the damping force). Let us, therefore, try the expression

$$x = Ae^{-ct}\sin(\omega t + \delta) \qquad (22\text{-}28)$$

as a possible solution. Substituting in Eq. (22-26), and equating the coefficients of $\sin(\omega t + \delta)$ and $\cos(\omega t + \delta)$ on both sides of the equation, shows that if

$$c = \frac{K}{2m} \qquad (22\text{-}29)$$

and

$$\omega = \left[\frac{k}{m} - \left(\frac{K}{2m}\right)^2\right]^{1/2}, \qquad (22\text{-}30)$$

then Eq. (22-28) is indeed a possible solution. It should be noted that if $K = 0$, Eq. (22-30) reduces to Eq. (22-8), and Eq. (22-28) becomes Eq. (22-2), as they should. We have obtained much more, however. Thus, Eq. (22-30) predicts that for damped oscillatory motion the frequency is a constant depending upon k, K, and m. The maximum frequency occurs for the undamped ($K = 0$) situation. For $K \neq 0$, the frequency is reduced to a lower value, but the motion is still oscillatory. This case is called *underdamped* motion. If $(K/2m)^2 = k/m$, $\omega = 0$ and the motion ceases to be oscillatory, becoming exponential instead—a situation referred to as *critically damped* motion. For $(K/2m)^2 \gtrsim k/m$, the motion is still non-oscillatory (since ω cannot be negative), but the exponential decay proceeds much more slowly, giving rise to *overdamped* motion. If $(K/2m)^2 \gg k/m$, Eqs. (22-28) to (22-30) no longer apply, but the motion is still exponential in form. We will not pursue this point. Figure 22-4 is a graph of x versus t for the three cases. Finally, we remark that the exponential term can be rewritten in the form $e^{-t/\tau}$, where

$$\tau = \frac{2m}{K} \qquad (22\text{-}31)$$

is called the decay time for the system. That is, in a time equal to the decay time the amplitude of the motion decreases to $1/e$ of its original value. The reader should find physical arguments for the decay time increasing with increasing mass m or decreasing frictional effects (decreasing K).

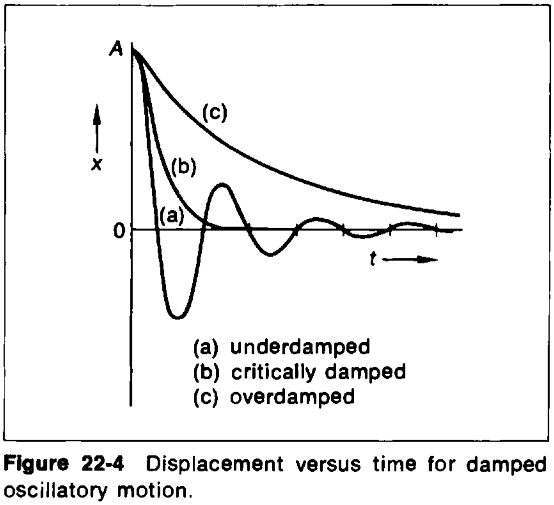


Figure 22-4 Displacement versus time for damped oscillatory motion.

22-6 FORCED OSCILLATIONS

If the oscillations of a damped system are to remain constant in amplitude, it is clear from the last section that energy must be supplied by some external agency to compensate for the energy lost because of the frictional force. This can be done by applying an external force to the system. To be effective, however, it is necessary that the force applied be an oscillatory function of time. This is because a constant force would resist the motion of the system whenever the system was moving in the direction opposite to the applied force. Furthermore, it is not difficult to see that an applied force with a frequency equal to that of the oscillatory system will be more effective than a force with a different frequency. An example of this is a child being pushed in a swing. Only when the swing is moving in the direction of the applied force will that force have its maximum effect.

The mathematics of a driven, damped oscillator is substantially more complex than the damped free oscillator. To begin with, when the driving force is initially applied, there will be an initial motion described by an equation of the form of Eq. (22-28). This motion dies out rather quickly (that is, for $t \gtrapprox 4\tau$), and is therefore referred to as the *transient* part of the solution. The more important part of the solution is the *steady-state* part, which describes the motion for all time once the transient part has damped out. In the analysis that follows, we will discuss only the steady-state solution for the case of an applied force varying sinusoidally with time. That is, we assume a driving force given by the equation

$$F_d = F_0 \sin \omega_d t, \qquad (22\text{-}32)$$

where ω_d is the angular frequency of the applied force. In this case, $\Sigma F = ma$ leads to

$$m\frac{d^2x}{dt^2} = F_0 \sin \omega_d t - kx - K\frac{dx}{dt}, \qquad (22\text{-}33)$$

where the various quantities retain the meanings assigned to them in the previous section.

Since we are interested in the steady-state motion, it is certainly reasonable to assume that the time dependence will be sinusoidal with the frequency of the driving force, ω_d. On the other hand, if $\omega_d \neq \omega = (k/m)^{1/2}$, the natural frequency of the undamped oscillator, we should expect that the amplitude of the oscillations will not be a maximum and will be frequency dependent. To test these assumptions, we substitute a solution of the form

$$x = A(\omega_d) \sin (\omega_d t + \delta) \qquad (22\text{-}34)$$

in Eq. (22-33). The result is

$$-m\omega_d^2 A(\omega_d) \sin (\omega_d t + \delta) = F_0 \sin \omega_d t - kA(\omega_d) \sin (\omega_d t + \delta) - K\omega_d A(\omega_d) \cos (\omega_d t + \delta)$$

or

$$\left(\frac{k}{m} - \omega_d^2\right) A(\omega_d) \sin (\omega_d t + \delta) = \frac{F_0}{m} \sin \omega_d t - \frac{K\omega_d}{m} A(\omega_d) \cos (\omega_d t + \delta). \qquad (22\text{-}35)$$

This can be simplified by using the identities

$$\sin (\alpha + \beta) = \sin \alpha \cos \beta + \cos \alpha \sin \beta$$

and

$$\cos (\alpha + \beta) = \cos \alpha \cos \beta - \sin \alpha \sin \beta,$$

and rearranging terms to give

$$\left[A(\omega_d)\left\{(\omega^2 - \omega_d^2) \cos \delta + \frac{K\omega_d}{m} \sin \delta\right\} - \frac{F_0}{m}\right] \sin \omega_d t = \left[-(\omega^2 - \omega_d^2) \sin \delta - \frac{K\omega_d}{m} \cos \delta\right] A(\omega_d) \cos \omega_d t, \qquad (22\text{-}36)$$

where $\omega^2 = k/m$, from Eq. (22-8). Since $\cos \omega_d t \neq \sin \omega_d t$, Eq. (22-36) is satisfied only if the coefficients of $\sin \omega_d t$ and $\cos \omega_d t$ vanish independently. We then have

$$A(\omega_d) = \frac{\dfrac{F_0}{m}}{\left\{(\omega^2 - \omega_d^2) \cos \delta + \dfrac{K\omega_d}{m} \sin \delta\right\}} \qquad (22\text{-}37)$$

and

$$\tan \delta = \frac{\sin \delta}{\cos \delta} = -\frac{\dfrac{K\omega_d}{m}}{(\omega^2 - \omega_d^2)}. \qquad (22\text{-}38)$$

From Eq. (22-38), it follows that

$$\sin \delta = \frac{-\dfrac{K\omega_d}{m}}{\left[(\omega^2 - \omega_d^2)^2 + \left(\dfrac{K\omega_d}{m}\right)^2\right]^{1/2}},$$

$$\cos \delta = \frac{(\omega^2 - \omega_d^2)}{\left[(\omega^2 - \omega_d^2)^2 + \left(\dfrac{K\omega_d}{m}\right)^2\right]^{1/2}}.$$

When these expressions are substituted into Eq. (22-37), we obtain

$$A(\omega_d) = \frac{\left(\frac{F_0}{m}\right)}{\left[(\omega^2 - \omega_d^2)^2 + \left(\frac{K\omega_d}{m}\right)^2\right]^{1/2}}. \qquad (22\text{-}39)$$

These somewhat laborious manipulations show that, as we had anticipated, the steady-state does indeed possess the frequency of the driving force. The amplitude of the oscillations depend not only upon the driving frequency and the natural frequency of the undamped oscillator but also upon the friction force constant K. The displacement of the oscillating object differs in phase from the applied force by the phase angle δ, given by Eq. (22-38).

We can draw additional conclusions from our solution. As asserted above, $A(\omega_d)$ will be a maximum when $\omega_d = \omega$. In fact, for an undamped oscillator ($K = 0$), Eq. (22-39) indicates that the amplitude increases without limit as $\omega_d \to \omega$. This is a situation known as resonance. At resonance an oscillating system absorbs a maximum amount of energy from the driving force mechanism during each cycle of the motion, and the amplitude increases until some portion of the system breaks down. The torsional oscillations induced in the Tacoma Narrows suspension bridge by winds of resonant frequency resulted in gigantic oscillations, which ultimately led to a collapse of the bridge. Figure 22-5 is a plot of amplitude versus frequency for undamped and damped driven oscillations.

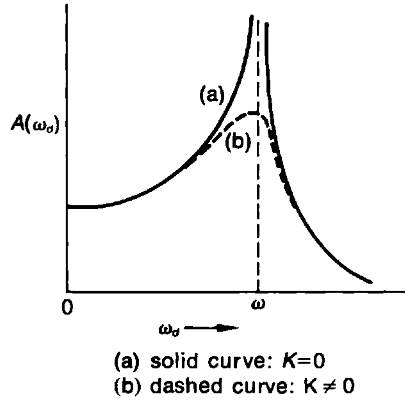


Figure 22-5 Amplitude versus driving frequency.

PROBLEMS

1. A body of mass 0.10 kg hangs from a long spiral spring. When pulled down 0.10 m below its equilibrium position and released, it vibrates with a period of 2 seconds.
   (a) What is its velocity as it passes through the equilibrium position?
   (b) What is its acceleration when it is 0.05 m above the equilibrium position?
   (c) What is the force constant of the spring?
2. A spiral spring of negligible mass hangs vertically with a 200 gm lead ball fastened to its lower end. The ball is then pulled down 4 cm and released. In 60 seconds it completes 150 vibrations or cycles. For the vibrating ball, compute:
   (a) its maximum speed and
   (b) the magnitude of its maximum acceleration.
   When the ball is 2 cm below its equilibrium position, compute the magnitude of:
   (c) its speed and
   (d) its acceleration.
3. A 3 kg mass is attached to a spring and set into oscillation. The mass executes 9 oscillations in 40 seconds. The total energy associated with the simple harmonic motion is 2 joules. Determine:
   (a) the maximum velocity of the mass and
   (b) the amplitude of the motion.
4. For a body undergoing simple harmonic motion,
   (a) when the displacement is one-half the amplitude, what fraction of the energy is kinetic and what fraction is potential?
   (b) At what displacement is the energy half kinetic and half potential?

5. A fifty cent coin is placed at the end of a rough plank that is vibrating horizontally with simple harmonic motion of an amplitude of 10 cm and a period of 2 seconds. Compute the coefficient of friction between the coin and the plank if the coin is just on the verge of slipping.
6. A body oscillates with simple harmonic motion according to the equation

$$y = 10 \sin\left(5\pi t + \frac{\pi}{6}\right) \text{ m},$$

   where t is in seconds. Find:
   (a) the displacement, velocity, and acceleration when $t = 2$ seconds and
   (b) the amplitude, frequency, and period of the motion.
7. A 0.070 kg mass is attached to a spring and placed on a horizontal frictionless plane as shown in Figure 22-6. The mass is displaced 0.12 m from equilibrium and released. The force constant of the spring is 8.47 nt/m. Determine the following quantities:
   (a) the frequency of oscillation,
   (b) the total energy of motion,
   (c) the maximum speed of the mass, and
   (d) the maximum acceleration of the mass.

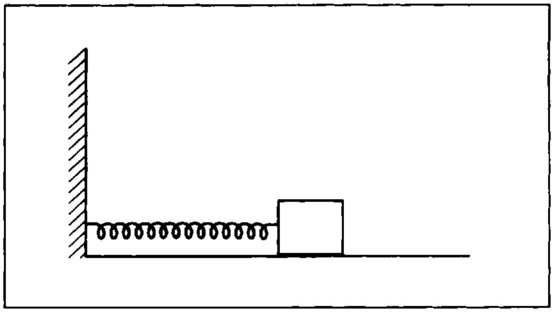


Figure 22-6

8. For a mass that is oscillating harmonically, plot the kinetic energy as a function of time and the potential energy as a function of time. Assume that the mass is displaced an amount A and released from rest at $t = 0$.
9. (a) A particle oscillates with a frequency of 1000 cycles/sec if the amplitude is 0.010 cm and its mass is 0.10 gms. What is the maximum restoring force on the particle?
   (b) Write the equation for a sinusoidal transverse wave of wavelength 2 m, traveling in the positive x-direction with a period of 0.05 seconds.
10. A body is vibrating with simple harmonic motion of amplitude 10 cm and frequency 3 vibrations/sec. Compute:
   (a) the maximum values of acceleration and velocity,
   (b) the acceleration and velocity when the displacement is 5 cm, and
   (c) the time required for the body to move from the equilibrium point to a point 8 cm distant from it.
11. A point at the end of one prong of a tuning fork makes 128 vibrations/sec, and the amplitude is 0.10 cm. Calculate the maximum velocity and maximum acceleration of this point.
12. In what ratio will the frequency, maximum potential energy, and maximum kinetic energy of simple harmonic motion in a spiral spring be altered if the load mass is quadrupled? If the amplitude is doubled?

23 Properties of Matter

23-1 ELASTICITY

When a substance is described as being rather elastic, one commonly pictures something like a rubber band or even a steel band which will, if deformed by stretching or bending, "snap back" to its original configuration. This qualitative behavior is similar to the Hooke's law situation of the last chapter. Therefore, if we wish to develop a more quantitative means of classifying substances that exhibit elastic behavior, it seems reasonable to proceed by discussing them in terms of SHM.

No real substance is completely elastic, if this means that it exhibits perfect SHM or is completely restored to its undisturbed state with no increase or decrease in energy content. Thus, the helical spring was idealized to permit the discussion in Chapter 22. It is easily demonstrated that such a system, if set in motion, will eventually come to rest because of internal dissipative forces that cannot entirely be eliminated. Furthermore, if such a spring is extended beyond a particular limiting extension, it becomes permanently deformed and assumes new characteristics. From these observations one can postulate a number of things:

1. Within rather restricted limits, many substances may exhibit properties that approximate those related to SHM.
2. Beyond those limits, the properties exhibited could vary markedly in a manner dependent upon both the material and the extent to which limiting values have been exceeded.
3. It might be useful to classify materials in general in terms of Hooke's law and to give some indication of the manner in which they fail to satisfy the ideal behavior requirements.

Ultimately, one would prefer a microscopic or detailed understanding of the interatomic or intermolecular forces that make it possible for a material to appear to behave like a harmonic oscillator. To attempt such a program here is not desirable for several reasons, although it does make an interesting study. Instead, we present a macroscopic description relating the application of external forces to changes in the size or shape of isotropic material and the forces produced by these changes.

Consider a metal wire, which is subjected to a tensile or stretching force parallel to the length of the wire. Measuring the applied force and the resulting elongation or extension of the wire for a wide range of values, we can illustrate the results graphically. It is customary to speak of the tensile stress, which is the applied force divided by the original cross-sectional area of the wire upon which it acts normally. The strain is given by the elonga-

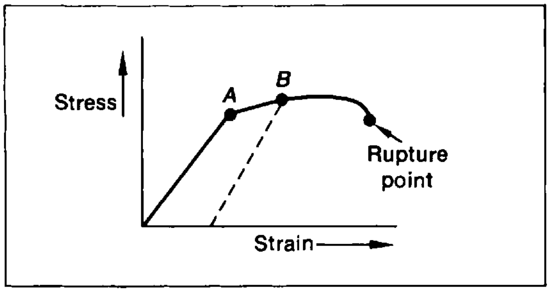


Figure 23-1 Stress-strain relation for a metal wire.

tion divided by the original length. Since the original area and the original length are constants, it is clear that an equivalent statement to Hooke's law is that the tensile stress is proportional to the strain. If the relevant data is plotted in this fashion, the results are as shown in Figure 23-1. Point A, the proportional limit, indicates where non-linearity begins on the curve. For metals, this corresponds to a strain of less than 1%. Above A, at point B, a permanent "set" or deformation takes place, so that reducing the stress to zero yields the dashed curve instead of the original curve. Beyond point B, a very slight increase in stress produces a marked increase in the strain. This region of the curve is referred to as the plastic region. The rupture point, as the name implies, is the point representing the stress that causes the sample to fail.

If we limit our discussion to stresses falling in the proportional region, we can expect that the material will behave in a manner approximating the SHM of Chapter 22. In the next section, the constant of proportionality in Hooke's law (in stress-strain form) will be identified for the common types of deformation that can occur at low stress. In Section 23-3, we discuss the relationships that exist between these elastic moduli, and present a table of moduli values for some common materials. Section 23-4 gives a brief indication of phenomena for which the SHM approximation is not a satisfactory description. The propagation of a longitudinal elastic wave along a metallic rod is the subject of Section 23-5.

23-2 STRESS, STRAIN, AND ELASTIC MODULI

If Hooke's law is to be written in the form discussed above (stress $\propto$ strain), our attention should be directed to the evaluation of the proportionality constant that is required if Hooke's law is to be an equation. This proportionality constant is called an elastic modulus and is equal to the ratio of stress to strain, *provided* that the material has not been subjected to stresses exceeding the proportional limit.

It is important to distinguish between types of deformation that can occur and to properly define the stress and strain related to these types. Three cases are presented here: the tensile case discussed above, tangential deformation or shear, and bulk or volume deformation. Figure 23-2 illustrates the three cases, with solid lines indicating the initial geometry and dashed lines indicating the geometry resulting from the application of the appropriate stress.

As indicated in Figure 23-2(a), the tensile force F_t is applied perpendicular to the cross-sectional area

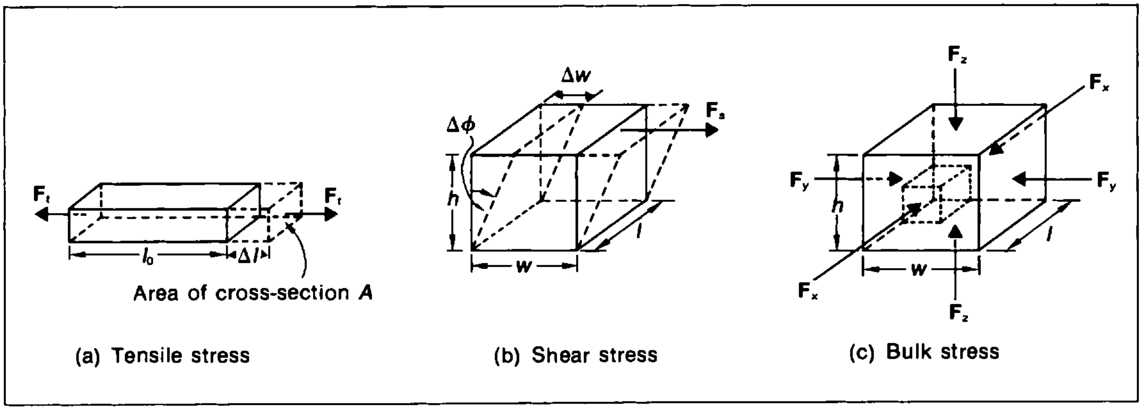


Figure 23-2 Stress-strain geometry for tensile stress, shear stress, and bulk stress.

A. The tensile stress or normal stress is thus defined to be

$$\text{tensile stress} = F_t/A, \tag{23-1}$$

and the related strain is given by

$$\text{tensile strain} = \frac{\Delta l}{l_0}. \tag{23-2}$$

The elastic modulus for this case is called Young's modulus and is indicated by *Y*. Thus, Hooke's law for this case becomes

$$\frac{F_t}{A} = Y\frac{\Delta l}{l_0}, \tag{23-3}$$

so that

$$F_t = \frac{YA}{l_0}\Delta l \tag{23-4}$$

can be compared with

$$F = k\Delta l, \tag{23-5}$$

which follows from Eq. (22-8) for the extension of a material obeying Hooke's law. We see, therefore, that by determining Young's modulus it is possible to calculate a force constant for the material. This constant represents an estimate of the force per unit separation exerted by one atom on another in a metal sample. Typical values of k obtained in this manner are of the order of 10 nt/m. One would expect, therefore, that any microscopic theory of the mechanical behavior of metal wires should predict values of this magnitude.

For the shear case, Figure 23-2(b) indicates that by applying the force F_s tangentially to the surface area $A = lw$, the solid is deformed linearly through an angle $\Delta\phi$. For this case, the shear stress is defined by the relation

$$\text{shear stress} = F_s/A = F_s/lw, \tag{23-6}$$

and the shear strain is taken to be

$$\text{shear strain} = \frac{\Delta w}{h} = \tan\Delta\phi. \tag{23-7}$$

Since we are concerned with strains of less than 1% (<0.01), we can use the relation

$$\tan\Delta\phi \approx \Delta\phi \quad \text{(in radians)},$$

so that

$$\text{shear strain} = \frac{\Delta w}{h} \approx \Delta\phi. \tag{23-8}$$

If the elastic modulus in the case of shear is given the symbol *S*, Hooke's law becomes

$$\frac{F_s}{A} \approx S\Delta\phi \tag{23-9}$$

or

$$S = \frac{F_s}{A}\Big/\Delta\phi, \qquad = \frac{F_s h}{lw\Delta w},$$

for the geometry of the figure.

Figure 23-2(c) illustrates a rectangular block of material subject to forces that are all normal to the surface to which they are applied. To avoid motion of the sample as a whole, it is clear that the forces applied to the surfaces must be equal in magnitude. Like the tensile case, it is natural to define the stress in this case as the ratio of the force applied normally to the surface divided by the area of the surface. The strain is the sum of the strains. Therefore, we write

$$\text{bulk strain} = \frac{\Delta w}{w} + \frac{\Delta h}{h} + \frac{\Delta l}{l}, \tag{23-10}$$

where w = total change in the width due to the bulk stress, etc. Since the volume of the solid is given by

$$V = lwh, \tag{23-11}$$

$$+\frac{\Delta V}{V} = +\frac{\Delta l}{l} + \frac{\Delta w}{w} + \frac{\Delta h}{h}. \tag{23-12}$$

Therefore, the bulk strain is equal to the fractional change in volume produced by the bulk stress. The elastic modulus for this situation is given the symbol *B*, and we define it by the relation

$$\text{bulk stress} = B\left(-\frac{\Delta V}{V}\right), \tag{23-13}$$

where the minus sign is introduced to insure that the bulk modulus (*B*) will be a positive number. An increase in applied force causes a *decrease* in volume.

In order to determine the bulk stress, we first introduce the concept of fluid pressure. An extended discussion of fluids is given in Chapter 24, so that we merely note here that a fluid is not capable of resisting shear or tangential stresses. When a gas or a liquid is subjected to a shear stress, slip motion occurs freely until the removal of the stress. Now consider a fluid enclosed in a container equipped with a tightly fitting piston. If the fluid is subjected to a compressive force by moving the piston, this force will be transmitted uniformly throughout the fluid if static equilibrium is to be maintained. Figure 23-3 shows a wedge-shaped portion of the fluid that has been isolated from the remaining fluid in the piston. Since the fluid is in equilibrium and shear forces do not exist, the forces F_1, F_2, and F_3 on the surfaces of the wedge due to the surrounding fluid must act normal to the respective surfaces of area A_1, A_2, and A_3. This means that F_1 must be equal in magnitude to F_{2x}, and F_3 is equal in magnitude to F_{2y} (see the decomposition of force F_2 into component forces F_{2x} and F_{2y} in the figure). As a result, the magnitude of the stresses

$$\left|\frac{F_1}{A_1}\right| = \left|\frac{F_{2x}}{A_1}\right| = \left|\frac{F_2 \sin\theta}{A_2 \sin\theta}\right| = \left|\frac{F_2}{A_2}\right|$$

and

$$\left|\frac{F_3}{A_3}\right| = \left|\frac{F_{2y}}{A_3}\right| = \left|\frac{F_2 \cos\theta}{A_2 \cos\theta}\right| = \left|\frac{F_2}{A_2}\right|.$$

We see, therefore, that the magnitudes of the stresses are equal for an arbitrary wedge angle θ; the stress must, therefore, be independent of orientation of the fluid wedge. This compressive stress is called the fluid pressure Δp (or hydrostatic pressure in the case of a liquid), and it follows that the pressure is the same on all the surfaces of the wedge.

Returning to Eq. (23-13), we can now write

$$\text{bulk stress} = \Delta p \tag{23-14}$$

and, therefore,

$$B = -\Delta p\left(\frac{V}{\Delta V}\right). \tag{23-15}$$

In the case of solids and liquids, the fractional change due to a given pressure increase is very small, so that the value of B is truly constant. For gases, however, the volume changes markedly for rather small changes in pressure. In this case, it is useful to use the relation

$$B = -V\frac{dp}{dV} \tag{23-16}$$

together with the gas law relating pressure and volume (and temperature). In fact, it is possible to obtain an infinite number of values for B, depending upon how the temperature varies during the variation of pressure that produces the volume change. To avoid confusion or ambiguity, it is customary to quote only values of the adiabatic bulk modulus, B_{ad}, and the isothermal bulk modulus, B_T. B_{ad} is the value obtained when the pressure volume change takes place in an insulated situation, so that no heat energy can enter or leave the system. Equivalent results are obtained if the change takes place so rapidly that heat energy does not have time to enter or leave the system before the change is complete. It is this modulus that is required in discussing the propagation of a sound wave in a gas (Chapter 28), since the periodic variations of pressure due to the sound wave take place too rapidly for heat energy to enter or leave a given region of the gas during the variation. The isothermal modulus on the other hand is the value obtained when the temperature is held fixed during the change in pressure.

Finally, it should be noted that some tables present values not of the bulk modulus but rather of the compressibility k, where

$$k = \frac{1}{B} = -\frac{1}{V}\left(\frac{\Delta V}{\Delta p}\right). \tag{23-17}$$

Thus, the compressibility gives the fractional decrease in volume per unit increase in pressure, and is usually quoted in reciprocal atmospheres or other (pressure units)$^{-1}$. As an example a one atmosphere increase in pressure on a volume of water will produce a decrease in volume of less than 0.01%.

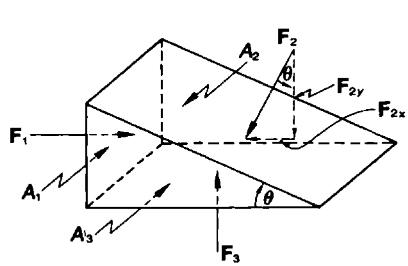


Figure 23-3 Segment of a fluid subject to a compressive force.

23-3 INTERRELATIONSHIPS OF THE ELASTIC MODULI

It was assumed in our discussions of Section 23-2 that no changes occur in the areas upon which the deforming forces are applied. It is known experimentally, however, that in the tensile case an increase in length is accompanied by a decrease in cross-sectional area at right angles to the length increase. If the substance is isotropic (as we have assumed in this chapter), the connection between the length strain and the linear strains associated with the fractional area change can be written

$$-\sigma\frac{\Delta l}{l}=\frac{\Delta w}{w}=\frac{\Delta h}{h}, \qquad (23\text{-}18)$$

where σ is called Poisson's ratio, and the cross-sectional area $A = wh$. Since an increase in l is accompanied by decreases in w and h, this definition assures that σ is positive. Experimentally, σ is found to be about 0.3. Theory shows that it cannot be negative and must be less than $\frac{1}{2}$ for isotropic materials. Furthermore, it is possible to show that for homogeneous isotropic materials a knowledge of σ and Y will provide the values of S and B. The relationships connecting them are:

$$Y = 2S(1+\sigma) \qquad (23\text{-}19)$$

$$Y = 3B(1-2\sigma). \qquad (23\text{-}20)$$

Example 1. Derive Eq. (23-20).

SOLUTION

Consider a rectangular solid of length l, and transverse dimensions w and h, which is placed in a pressure tank and subjected to a uniform hydrostatic pressure Δp on all surfaces. As a result, compressional forces will produce corresponding compressional strains in all three directions. From the definition of Poisson's ratio (σ), however, it is clear that a compressional strain in the (w) (or h) directions will be related to a tensile strain in the l direction. Let us therefore determine the total strain in the l direction and relate this to the volume strain.

First of all, the compressional stress in the (l) direction (Δp) produces a strain given by

$$\Delta p = -Y\left(\frac{\Delta l}{l}\right)_l \qquad \text{(i)}$$

or

$$\left(\frac{\Delta l}{l}\right)_l = -\frac{\Delta p}{Y}. \qquad \text{(ii)}$$

For the (w) direction, we can write

$$\left(\frac{\Delta w}{w}\right) = -\frac{\Delta p}{Y}. \qquad \text{(iii)}$$

Since

$$\left(\frac{\Delta w}{w}\right) = -\sigma\left(\frac{\Delta l}{l}\right)_w, \qquad \text{(iv)}$$

we have

$$\left(\frac{\Delta l}{l}\right)_w = \sigma\frac{\Delta p}{Y}. \qquad \text{(v)}$$

Similarly, in the (h) direction we obtain

$$\left(\frac{\Delta l}{l}\right)_h = \sigma\frac{\Delta p}{Y}. \qquad \text{(vi)}$$

Now the total strain in the (l) direction becomes

$$\frac{\Delta l}{l} = \left(\frac{\Delta l}{l}\right)_l + \left(\frac{\Delta l}{l}\right)_w + \left(\frac{\Delta l}{l}\right)_h = -\frac{\Delta p}{Y}(1-2\sigma). \qquad \text{(vii)}$$

Since the material is isotropic and the applied pressure is uniform, the same result is obtained if we calculate the total strain in the (w) or (h) directions. That is,

$$\frac{\Delta l}{l}=\frac{\Delta w}{w}=\frac{\Delta h}{h}= -\frac{\Delta p}{Y}(1-2\sigma). \qquad \text{(viii)}$$

Notice now that the volume of the solid is

$$V = lwh, \qquad \text{(ix)}$$

so that the fractional change in volume due to the applied pressure is just the volume strain;

$$\frac{\Delta V}{V}=\frac{\Delta l}{l}+\frac{\Delta w}{w}+\frac{\Delta h}{h}. \qquad \text{(x)}$$

Substituting (viii) in (x), we obtain

$$\frac{\Delta V}{V} = -\frac{3\Delta p}{Y}(1-2\sigma) \qquad \text{(xi)}$$

or

$$Y = -\left(\frac{\Delta p}{\frac{\Delta V}{V}}\right)\{3(1-2\sigma)\}. \qquad \text{(xii)}$$

Equation (23-13) then leads to the desired result

$$Y = 3B(1-2\sigma). \qquad \text{(xiii)}$$

It should be noted that Eq. (xi) requires that $\sigma < \frac{1}{2}$, since a greater value would result in an expansion when the solid is subjected to a compression, which is contrary to experiment.

It should be emphasized that the assumption of a homogeneous isotropic material is an idealization. This is because no real material is microscopically homogeneous (there is empty space between the atoms constituting the material). Furthermore, most materials are not isotropic. Their properties in fact vary markedly with direction in the material, due to the forces holding the material together.

In Table 23-1, approximate values of the moduli and Poisson's ratio are given for some materials. The reader can determine readily the extent to which Eqs. (23-19) and (23-20) are satisfied. Even where the equations fail, the lack of validity is not generally a gross failure. As a result, these equations can be used for estimation purposes when it is not desirable to measure all the moduli for materials whose values are not tabulated.

Table 23-1 Approximate Elastic Moduli and Poisson's Ratio for Selected Materials.

Substance	$Y\left(10^{10}\frac{\text{nt}}{\text{m}^2}\right)$	$S\left(10^{10}\frac{\text{nt}}{\text{m}^2}\right)$	$B\left(10^{10}\frac{\text{nt}}{\text{m}^2}\right)$	σ
Aluminum	7.0	2.8	7.0	0.34
Copper	12.0	4.2	12.0	0.34
Iron	20.0	8.2	16.0	0.30
Lead	1.16	0.5	3.3	0.40
Silver	8.0	8.0	10.0	0.38

23-4 INELASTIC PROPERTIES OF SOLIDS

It is possible to demonstrate experimentally that the elastic properties discussed in the preceding sections provide a satisfactory description of real materials. For example, consider a mass m suspended from a wire and set into small amplitude vertical oscillatory motion. The observed frequency of vibration is consistent with the value $\omega = (k/m)^{1/2}$ calculated from a knowledge of the mass m and the value of k obtained from a comparison of Eqs. (23-4) and (23-5). Other examples of such tests are given in the problems. In the next section, the propagation of longitudinal elastic waves in solids is developed in terms of elastic moduli. The results can similarly be shown to be in accord with physical reality (within the limits imposed on the development at the outset).

As a result, it is tempting to speculate that the world around us is primarily governed by linear restoring forces which produce oscillatory motions that are essentially symmetric about the equilibrium position. Providing the amplitudes of the motions are suitably limited, this speculation is in fact essentially correct. On the other hand, some situations cannot be described by an elastic model for solids. When this is the case, the force law governing the motion is found to have a more complicated form, a principal feature of which is a departure from the symmetry of Hooke's law. For example, when the displacement Δl of a solid rod is negative, the restoring force is substantially larger than the restoring force for a corresponding positive displacement, and the disparity becomes more pronounced at larger displacement magnitudes. This behavior is not unreasonable from a microscopic point of view. As the atoms or molecules of a solid are brought closer together, the restoring forces might be expected to increase sharply, especially when the atoms are essentially in contact with one another. Conversely, as they become more widely separated, one would expect the restoring forces to diminish steadily.

Physical evidence favoring this picture of inelastic or asymmetric forces is not hard to find, although a complete discussion of the various phenomena belongs in a study of the solid state. We shall now discuss qualitatively the situation involved in the thermal expansion of solids and merely list some other inelastic properties. For additional details, the reader can consult any introductory text devoted to the solid state.

In Chapter 3, the length of a solid rod as a function of temperature was given by Eq. (3-2),

$$l_t = l_0(1 + \alpha \Delta t),$$

where

l_t = length at temperature t,
l_0 = length at temperature t_0 (usually 0°C),
α = coefficient of linear expansion,
$\Delta t = t - t_0$ = change in temperature.

This relationship can also be written in the form of Eq. (3-1):

$$\alpha = \frac{\Delta l}{l_0 \Delta t}$$

or

$$\Delta l = \alpha l_0 \Delta t. \qquad (23\text{-}21)$$

This latter form can be interpreted as a linear relationship between the displacement (Δl) from the equilibrium configuration (l_0) and the temperature (measured from the reference temperature t_0). If the solid is viewed microscopically as a collection of atoms each oscillating about an equilibrium position, it is reasonable to assume that the higher the temperature the greater will be the range of displacements from the equilibrium positions. Furthermore, it is also reasonable to consider the displacement (Δl) in Eq. (23-21) as the average displacement resulting from the oscillations of the atoms of the solid. For an atom oscillating in SHM, the symmetry of the motion would produce an average displacement of zero magnitude. In other words, the elastic model predicts no thermal expansion. On the other hand, an inelastic model predicts that the oscillating atom would vibrate to larger positive displacements than negative displacements from the equilibrium position for a given total energy because of the lack of symmetry of the restoring forces, as mentioned earlier. As a result, the average displacement of each atom will be positive, corresponding to an increased separation of the atoms. If the total energy is regarded as a linear function of temperature, it is plausible to assert that the inelastic model predicts a thermal expansion that increases linearly with temperature, in agreement with Table 3-1. From the values indicated in that table for α, it is clear that the linear expansion is very small for small changes in temperature ($\Delta t \approx 0$), giving further support to the assumption of elastic forces for small displacements.

Finally, other properties requiring an inelastic force model are: (1) the dependence of the elastic constants of a solid on temperature and pressure; (2) the temperature dependence of the thermal conductivity of a solid; and (3) the temperature dependence of the heat capacity of a solid at high temperatures.

23-5 ELASTIC WAVES†

The discussion of SHM in Chapter 22 considered the motion of a mass under the influence of an elastic or linear restoring force. Such a force can be provided by a helical spring or, as mentioned in the previous section, by suspending the mass from a slender wire, which is given an initial extension, and then releasing it. In these two situations, it was tacitly assumed that the mass undergoing the motion was large enough to permit neglecting the mass of the elastic material. In this section, we consider the motion of an elastic medium itself.

As a specific example, we will consider the propagation of longitudinal waves in an isotropic solid in the form of a long slender rod of uniform cross section. For simplicity, the cross-sectional area is assumed to be small enough that transverse vibrational effects (related to Poisson's ratio considerations) are negligible. It is not difficult to visualize such a system. For example, by striking the end of the rod periodically with a hammer, a wave (or train of pulses) will be produced that travels along the rod. At any given thin section of the rod, the displacement from its equilibrium position depends upon time and upon position along the rod. Figure 23-4 illustrates the situation. We now write Newton's second law of motion as it applies to the thin section whose volume is $A\,dx$. For a uniform density ρ, the mass of this section is $\rho A\,dx$. If u, the displacement of the segment from equilibrium, is small, then we may neglect the small change in size of the section without appreciable error. As indicated in the diagram, the force on the section directed to the right is then given by‡

$$\left[\frac{F_t}{A} \times \frac{\partial(F_t/A)dx}{\partial x}\right]A = F_t + \frac{\partial F dx}{\partial x},$$

while the force on the section directed to the left is

$$\left(\frac{F_t}{A}\right)A = F_t.$$

As a result, the net force becomes

$$\Sigma F_x = \frac{\partial F_t}{\partial x}\,dx,$$

and the second law becomes

$$\frac{\partial F_t}{\partial x}\,dx = \rho A\,dx\frac{\partial^2 u}{\partial t^2}. \tag{23-22}$$

But from Hooke's law, we have

$$\frac{F_t}{A} = Y\frac{\partial u}{\partial x}, \tag{23-23}$$

where $\partial u/\partial x$ is the strain. Note again that the partial derivative notation is required, since u also depends upon both x and t. From Eq. (23-23), we see that

$$\frac{\partial F_t}{\partial x}\,dx = YA\,dx\frac{\partial}{\partial x}\left(\frac{\partial u}{\partial x}\right) = YA\,dx\frac{\partial^2 u}{\partial x^2}. \tag{23-24}$$

Combining Eqs. (23-22) and (23-24), we get

$$\frac{\partial^2 u}{\partial t^2} = \frac{Y}{\rho}\frac{\partial^2 u}{\partial x^2}. \tag{23-25}$$

†This section is not essential in the remainder of the text and may be omitted if desired.)

‡The notation $\partial F_t/\partial x$ indicates a partial derivative. That is, since F_t depends upon both position and time, $\partial F_t/\partial x$ indicates that the derivative with respect to x is to be determined, while the dependence on time is not allowed to vary. Similarly, $\partial F_t/\partial t$ indicates a differentiation with respect to time, while the x dependence is held fixed. For example, suppose $F_t = 4xt^2$. Then $\partial F_t/\partial x = 4t^2$, and $\partial F_t/\partial t = 8xt$.

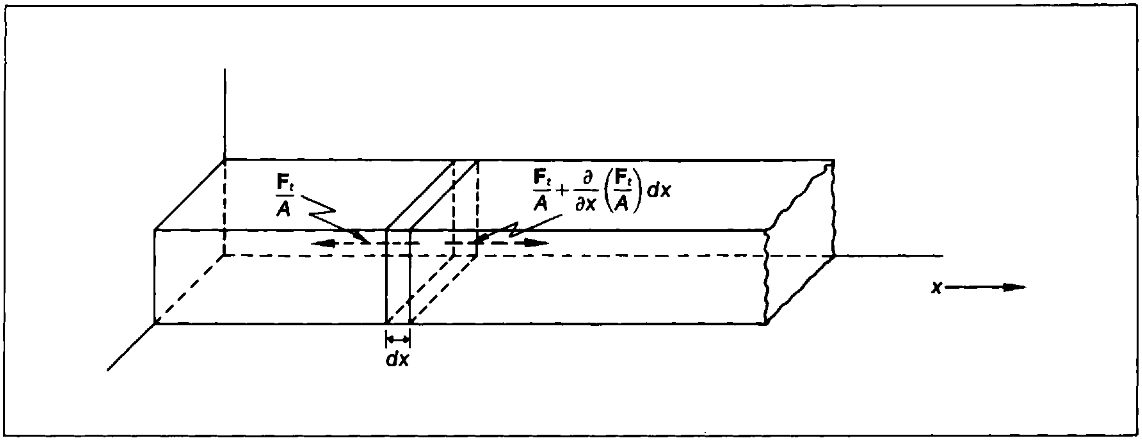


Figure 23-4 Longitudinal stresses acting upon a thin section of a long rod.

Equation (23-25) is a one-dimensional wave equation. Its solution is clearly a rather special kind of function in that differentiating it twice with respect to time must be equal to a constant times the result of differentiating twice with respect to position.

Rather than considering this problem from a strictly mathematical point of view, it is more appropriate to be guided by the physical considerations presented in Chapter 10, where we first discussed traveling waves. Thus, let us assume in analogy with Eq. (10-5) that the solution we seek can be written in the form

$$u = A \sin\frac{2\pi}{\lambda}(x - vt). \tag{23-26}$$

Equation (23-26) leads to

$$\frac{\partial u}{\partial x} = \left(\frac{2\pi}{\lambda}\right) A \cos\frac{2\pi}{\lambda}(x - vt),$$

$$\frac{\partial^2 u}{\partial x^2} = -\left(\frac{2\pi}{\lambda}\right)^2 A \sin\frac{2\pi}{\lambda}(x - vt)$$

$$= -\left(\frac{2\pi}{\lambda}\right)^2 u; \tag{23-27}$$

$$\frac{\partial u}{\partial t} = -\left(\frac{2\pi}{\lambda}v\right) A \cos\frac{2\pi}{\lambda}(x - vt);$$

$$\frac{\partial^2 u}{\partial t^2} = -\left(\frac{2\pi v}{\lambda}\right)^2 A \sin\frac{2\pi}{\lambda}(x - vt)$$

$$= -\left(\frac{2\pi v}{\lambda}\right)^2 u. \tag{23-28}$$

Substituting Eqs. (23-27) and (23-28) in Eq. (23-25), we obtain

$$-\left(\frac{2\pi}{\lambda}\right)^2 v^2 = -\left(\frac{2\pi}{\lambda}\right)^2 \frac{Y}{\rho}$$

or

$$v^2 = \frac{Y}{\rho}. \tag{23-29}$$

We see, therefore, that our assumed solution, Eq. (23-26), is satisfactory provided that the velocity of propagation is related to the elastic and mechanical properties of the rod by Eq. (23-29). It is left for the reader to show that the expression $(Y/\rho)^{1/2}$ has the dimensions of velocity. Since $Y \approx 10^{11}\,\mathrm{nt/m^2}$ and $\rho \approx 10^4\,\mathrm{kg/m^3}$, $v \approx 3\times 10^3$ m/sec for a solid rod.

Although we shall not do so here, it is not difficult to show by similar reasoning that similar wave equations result for transverse waves upon a string, longitudinal waves in gases, or longitudinal *and* transverse waves in solids of unlimited extent. In every case, it is found that the velocity of propagation is proportional to the square root of the appropriate elastic modulus divided by the density of the medium.

PROBLEMS

1. A 100 kg weight W is attached by means of a hook to a horizontal rod 2 m long, suspended by two wires of equal length, as shown in Figure 23-5. Wire A has a cross-sectional area of 1 mm^2, and wire B has a cross-sectional area of 3 mm^2. The weight of the rod may be neglected. At what distance from A should the weight be placed to give equal stresses in wires A and B?

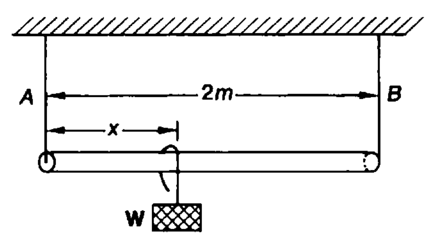


Figure 23-5

2. A wire of 0.5 mm in diameter and 10 m long hangs from the ceiling, and an added weight of 2 kg causes an extension of 1 mm. Find the Young's modulus for the wire.

3. Find the extension caused in a steel wire 1 mm in diameter and 5 m in length when an extra weight of 900 gm is hung on it. $Y_{steel} = 22 \times 10^{10}$ nt/m^2.

4. When a 5 kg block is attached to the end of a spring hanging vertically, the spring experiences an elongation of 10 cm. Find the force constant of the spring.

5. A hollow cylindrical iron column 12 ft high, 10 in. external diameter, and 8 in. internal diameter is anchored in a vertical position to a rigid base. The column is then subjected to a load of 220 tons placed at its upper end. $Y_{iron} = 2 \times 10^7$ lb/in^2 for this specimen.
   (a) Find the amount by which the column is shortened.
   (b) Find the work done in shortening the column.

6. A 9.8 kg weight fastened to the end of a steel wire of unstretched length 1 m is whirled in a vertical circle with an angular velocity of 1 rev/sec at the bottom of the circle. The cross section of the wire is 1 mm^2. Calculate the elongation of the wire when the weight is at the lowest point of its path. $Y_{steel} = 22 \times 10^{10}$ nt/m^2.

7. A mass of 5 kg attached to a spring causes it to stretch 1 cm. What is the work done in stretching the spring?

8. A rod of elastic material is elongated 2% by a stress of 10^9 nt/m^2, assuming this to be within the elastic limit. What is the Young's modulus of the material?

9. The force constant of a certain spring is 1 nt/cm. What is the potential energy of the spring when it is extended 10 cm?

10. A spiral spring with a force constant of 0.50 lb/ft is 3 ft long when unstretched. What is its elastic potential energy when it is stretched to a length of 4.20 ft?

11. How much force is necessary to stretch a steel wire 2 mm^2 in cross-sectional area and 2 m long a distance of 0.50 mm? $Y_{steel} = 22 \times 10^{10}$ nt/m^2.

12. A high pressure apparatus can produce pressures as high as 2×10^9 nt/m^2. What change in volume will a cube of quartz 1 cm on a side undergo if subjected to this pressure? $B_{quartz} = 2.7 \times 10^{10}$ nt/m^2.

13. Suppose a rubber ball 4 in. in diameter is immersed in water until the pressure on its surface is 20 lb/in^2. What is the change in volume of the ball? $B_{rubber} = 1 \times 10^7$ nt/m^2.

14. What is the shear modulus of a rod 1 m long and 2 mm in diameter if the torsion constant of the rod is 5000 m–nt/rad?

15. The tension in a rod, whose cross-sectional area is 25 in², is 1000 lb. What is the shearing stress on the inclined section of the rod shown in Figure 23-6?

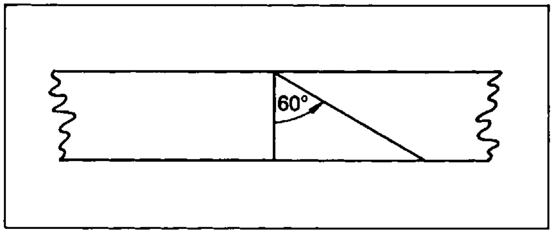


Figure 23-6

24 Fluid Mechanics

24-1 IDEAL AND REAL FLUIDS

A fluid is a substance that has the ability to flow or alter its shape to conform more closely to that of its container. Thus, solids are non-fluid, but both gases and liquids are classified as fluids. Although glass exhibits an extremely slow flow, it behaves in many respects more like a solid and is not usually considered to be a fluid. Another way of describing fluids is to say that they have little or no rigidity. From Chapter 23, a rigid substance is one that can resist a change in shape when acted upon by tangential or shearing forces. All real substances must exhibit some tendency toward rigidity.† This is due to the existence of the forces that the molecules of the substance exert upon each other and upon the molecules of any containing vessel. For solids, these forces are quite large, but for fluids they are much weaker and in some circumstances can be neglected.

We can distinguish between a real fluid and the concept of an ideal fluid in terms of this rigidity. A real fluid exhibits some rigidity while an ideal fluid does not. This absence of rigidity simplifies the discussion of fluids in static equilibrium (Sections 24-2 and 24-3) and in motion (Sections 24-4 and 24-5). Because of the mathematical complexities involved, we will not discuss modifications that are required in dealing with the motion of real fluids.

24-2 FLUID STATICS

If a given body of fluid is at rest, Newton's first law of motion requires that at any point in the fluid the resultant force at the point is zero. Furthermore, for any small surface area element about a point in the fluid, any force acting on that surface must be acting in a direction perpendicular to the surface. This assertion follows from our assumption that the fluid is ideal and so cannot respond to tangential or shearing forces except by flowing freely, which would not be a static situation. Even for real fluids, there can be no tangential forces if static equilibrium exists. It simplifies the application of Newton's laws of motion to ideal fluids if the concept of pressure is used. Pressure is a scalar quantity, which is equal to the magnitude of the normal force per unit surface area acting on a given element of surface area. If we consider a small surface area A in the fluid, a total force F due to the fluid on one side of A will be exerted on the fluid on the other side of A (and, conversely, by Newton's

†We do not include superfluids in this discussion. Their properties are discussed in the June 1958 issue of *Scientific American.*

third law). In terms of these quantities, the average pressure $\bar{p}$ is given by

$$\bar{p} = \frac{F}{A}. \qquad (24\text{-}1)$$

The pressure at a given point in the fluid is given by the ratio of the force to the surface area as the surface area (bounding the point in question) becomes vanishingly small. That is,

$$p = \lim_{A \to 0} \frac{F}{A}. \qquad (24\text{-}2)$$

Additional simplification becomes possible by using the density of the fluid at the point in question instead of the mass of a given part of the fluid. Density is defined as the mass per unit volume of the substance,

$$\rho = \frac{m}{V}. \qquad (24\text{-}3)$$

As we have already seen in Chapter 3, a change in temperature can alter the dimensions (and thus the volume) of a substance. Since the mass of the substance will not be altered by temperature (unless the substance somehow disintegrates and disperses or else undergoes a chemical reaction), it is reasonable (and correct!) to assume that the density of a substance depends upon the temperature. There are other factors that can influence the density of a substance which vary in importance for the different phases of matter. Because of their very low compressibilities, solids and liquids show little change in density as the pressure varies, in marked contrast to the situation for gases. Densities of various substances are given in Table 24-1.

Let us now apply the condition of static equilibrium to determine the variation of pressure with depth in a fluid assumed to be at rest. Consider a small volume of fluid at a depth y below the surface, as in Figure 24-1. The element of fluid has a horizontal surface area A and a thickness dy. For

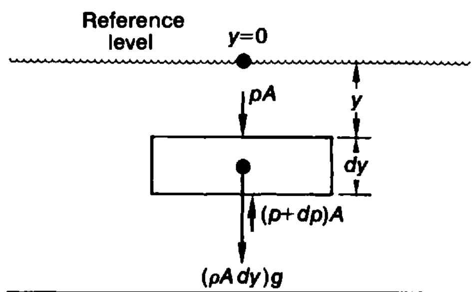


Figure 24-1 Pressure variation with depth in a fluid.

fluid density ρ, the mass of fluid in the volume is $\rho A\,dy$, and its weight is therefore $(\rho A\,dy)g$. Since the fluid is at rest, the sum of all forces must be zero. The horizontal forces on the volume element must cancel one another in pairs. Since the horizontal forces are due to the pressure exerted by the surrounding fluid, symmetry requires that the pressure be the same at all points in a given horizontal plane. In equilibrium, the vertical forces must also vanish. From Figure 24-1, it is clear that

$$\Sigma F_y = 0 = (p + dp)A - pA - (\rho A\,dy)g,$$

$$dp = \rho g\,dy \qquad (24\text{-}4)$$

or

$$\frac{dp}{dy} = \rho g. \qquad (24\text{-}5)$$

From Eq. (24-5), we see that there is a downward pressure gradient; that is, the pressure increases with increasing depth below the surface. To determine the pressure difference between two points, it is necessary to integrate Eq. (24-4). Thus, if p_1 is the pressure at depth y_1, and p_2 is the pressure at depth y_2, we have

$$\int_{p_1}^{p_2} dp = \int_{y_1}^{y_2} \rho g\,dy$$

Table 24-1 Approximate Densities of Solids and Liquids (values are given for 20°C except when otherwise stated).

Liquids (10^3 kg/m^3)		Metallic solids (10^3 kg/m^3)		Non-metallic solids (10^3 kg/m^3)	
Ethyl alcohol	0.789	Steel	7.80	Ice	0.992 (0°C)
Carbon tetrachloride	1.59	Copper	8.90	Concrete	2.30
Sea water	1.03	Brass	8.70	Glass	2.60
Pure water	1.000 (4°C)	Gold	19.3	Cork	0.240
Mercury	13.6 (0°C)	Lead	11.3	Balsa wood	0.130
Benzene	0.880	Aluminum	2.70	Oak	0.720
				Ebony	1.20

or

$$p_2 - p_1 = \Delta p = \int_{y_1}^{y_2} \rho g\, dy. \qquad (24\text{-}6)$$

The integral on the right-hand side of Eq. (24-6) cannot be performed unless the relationship between density and depth is known. Furthermore, one must be sure that there is no variation of the value of g in the depth interval involved. For liquids on the Earth's surface, this is true, but in the Earth's atmosphere there is a non-negligible variation of density with height (negative depth) above the Earth's surface, and, for sufficiently large differences in height, g varies also.

Example 1. Find the relationship between pressure and depth below the surface of water.

SOLUTION

At the surface ($y_1 = 0$), the pressure $p_1 = p_0$ is that due to the atmosphere, at $y_2 = d$, the pressure is $p_2 = p_0 + \Delta p$. For an incompressible fluid and assuming no variation in g, Eq. (24-6) gives

$$p_2 - p_0 = p_0 + \Delta p - p_0 = \rho g (y_2 - y_1) = \rho g d,$$

$$\Delta p = \rho g d$$

or, alternatively,

$$p_2 = p_0 + \rho g d. \qquad (24\text{-}7)$$

The weight density (ρg) of fresh water is about 62.4 lb/ft^3. Equation (24-7), therefore, indicates that the pressure increases by 62.4 lb/ft^2 for every foot of depth below the surface in a lake. Since atmospheric pressure at sea level is about 15 lb/in^2, the reader can show that an increase in pressure equal to atmospheric pressure corresponds to a depth of approximately 35 ft in a fresh water lake.

24-3 ARCHIMEDES' PRINCIPLE

Over 2000 years ago, Archimedes deduced that a body immersed in a fluid should experience a lifting (or buoyant) force equal to the weight of fluid displaced by the body and acting in a vertical direction through the original center of gravity of the displaced fluid. To see that this must be so, consider a volume of the undisturbed fluid having the same shape as the solid to be immersed in the fluid. From the discussion of the previous section, the net force exerted on the volume of fluid by the surrounding fluid must be just equal in magnitude and opposite in direction to the weight of that volume of fluid for equilibrium to exist. In addition, that force must act through the center of gravity, as does the weight of the volume of fluid involved. If the volume of fluid is replaced by a body with the same shape and volume, the remaining fluid will still exert the same force as it did on the volume of fluid, since no change has been made in the remaining fluid. If the body remains at rest when immersed, it follows that its average density must correspond to that of the displaced fluid.

When the buoyant force is greater than that of the weight of the body (density of the fluid greater than density of the body), the body floats in a liquid or gains altitude in a gas. Conversely, when the buoyant force is less than the weight of the body (density of the fluid less than density of the body), the body sinks in a liquid or loses altitude in a gas. For example, if a gas-filled ballon is released in still air, it will rise to a height at which the density of the supporting air equals the average density of the balloon and the gas inside. At that height, the buoyant force will equal the weight of the balloon and its contents; no further increase in height will occur.

Buoyant forces must be taken into account if one makes precision "weighings" with a sensitive equal arm analytical balance when the density of the object being measured differs appreciably from the density of the standard "weights" used with the balance. For an object of volume and density V_x and ρ_x, and using "weights" of density ρ_w and balancing volume V_w in air of density ρ_a, a balanced situation requires that

$$(\rho_x V_x - \rho_a V_x) g = (\rho_w V_w - \rho_a V_w) g. \qquad (24\text{-}8)$$

The desired quantity is the actual mass ($\rho_x V_x$) of the object, which from Eq. (24-8) is

$$\rho_x V_x = \rho_w V_w + \rho_a (V_x - V_w). \qquad (24\text{-}9)$$

If ρ_x differs greatly from ρ_w, then V_x and V_w will also differ markedly, and the buoyant force exerted by the air cannot be neglected.

24-4 APPLICATIONS OF FLUID STATICS

The mercury barometer can be used to measure atmospheric pressure. As shown in Figure 24-2, it consists of a glass tube filled with mercury. The tube is then inverted in an open dish of mercury. The mercury in the tube then drops, until the weight of the remaining column of mercury exerts the same pressure on a unit cross section of surface of mercury at point 2 as does the atmosphere outside

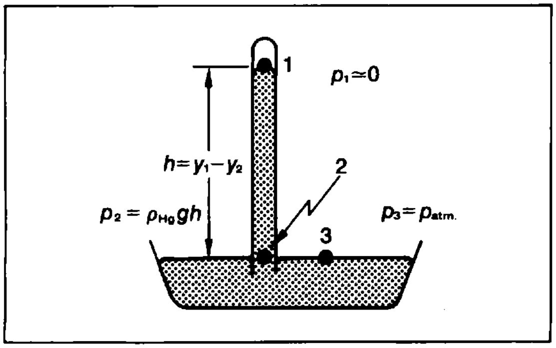


Figure 24-2 The mercury barometer.

the tube on a unit cross section of the surface at point 3. (The small amount of mercury vapor in the evacuated space above the column is negligible at ordinary temperatures.) By Eq. (24-7),

$$p_2 = p_1 + \rho_{Hg} gh$$
$$= \rho_{Hg} gd.$$

Since points 2 and 3 lie in the same horizontal plane, symmetry requires that the two points are at the same pressure. Therefore,

$$p_{atm} = \rho_{Hg} gh.$$

From Table 24-1, the density of mercury is $13.6 \times 10^3\,kg/m^3$. At 0°C at sea level, where $g = 9.81\,m/sec^2$, a pressure of one atmosphere is taken to be equal to exactly 0.76 m of mercury. Thus, one atmosphere is equivalent to

$$p_{atm} = (13.6 \times 10^3\,kg/m^3)(9.81\,m/sec^2)(0.76\,m)$$
$$= 1.013 \times 10^5\,nt/m^2$$
$$= 14.70\,lb/in^2$$
$$= 1\,atm.$$

It is common to quote pressures in centimeters or inches of mercury, but it should be understood that this is done for convenience. The example above indicates the calculation that is required to convert centimeters of mercury to actual units of force per unit area. If a liquid of density ρ_x (and negligible vapor pressure) is used in the barometer instead of mercury, it should be clear that equilibrium with atmospheric pressure requires that

$$\rho_x h_x = \rho_{Hg} h_{Hg}$$

or

$$h_x = \frac{\rho_{Hg}}{\rho x} h_{Hg}. \qquad (24\text{-}10)$$

Example 2. What is the height of a column of water in a water barometer on a day when atmospheric pressure is one atmosphere?

SOLUTION

From Eq. (24-10),

$$h_w = \frac{(13.60 \times 10^3\,kg/m^3)}{(1 \times 10^3\,kg/m^3)}(0.76\,m)$$
$$= 10.34\,m \text{ of water}$$
$$= 33.9\,ft \text{ of water}.$$

As another example of the application of the conditions of static equilibrium to fluids at rest, consider the situation shown schematically in Figure 24-3. An upright gate of width L and height d, hinged along its base, serves as part of a dam. If the water surface is level with the top of the gate: (a) what is the total horizontal force acting on the gate, and (b) what is the torque about the hinge due to the body of water?

To begin with, we can neglect atmospheric pressure since it acts on both sides of the gate, directly on the one side and as an increase in hydrostatic pressure on the upstream side. From Eq. (24-7), neglecting P_{Atm}, the pressure p at depth y is

$$p = \rho g y,$$

so that the resulting horizontal force on the element of area

$$dA = L\,dy$$

is

$$dF = p\,dA = \rho g L y\,dy.$$

The total force will be

$$F = \int_0^F dF = \int_0^d \rho g L y\,dy = \frac{\rho g L d^2}{2}.$$

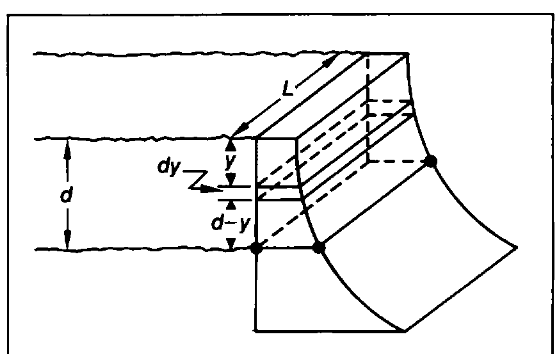


Figure 24-3 A hinged vertical gate in a dam.

The torque acting about the hinge due to the pressure p on the area element dA is

$$d\tau = (p\,dA)(d-y) = \rho gLy(d-y)dy,$$

and the total torque about the hinge will be

$$\tau = \int_0^\tau d\tau = \int_0^d \rho gLy(d-y)dy = \frac{1}{6}\rho gLd^3. \quad (24\text{-}11)$$

Let us examine more closely the assertion in the example above that the effect of the external (atmospheric) pressure on the surface of the water is to increase the total pressure at any point by exactly the same amount. This follows directly from Eq. (24-7), and is known as Pascal's principle since it was stated by Blaise Pascal in the mid-seventeenth century. A major application of the principle can be found in the operation of a hydraulic jack or hydraulic press. As indicated in Figure 24-4, a small force F_s is applied to a piston of cross-sectional area A_s, which thus exerts a pressure $p = F_s/A_s$ on the liquid in the device (oil, for example). By Pascal's principle, this pressure is transmitted undiminished to a larger piston of area A_L. Since the increase in pressure is the same at both pistons, the force F_L exerted on the larger piston because of the applied force is found from the relation

$$p = \frac{F_s}{A_s} = \frac{F_L}{A_L}$$

or

$$F_L = F_s\frac{A_L}{A_s}.$$

Thus, the hydraulic press is a force multiplying device that depends upon the ratio of the piston areas for its multiplying factor.

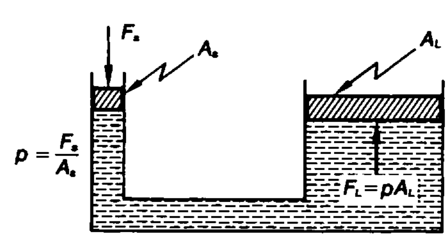


Figure 24-4 The hydraulic press.

24-5 FLUID DYNAMICS—BERNOULLI'S EQUATION

We turn now to fluid dynamics, the study of fluids in motion. To begin our study, we continue to discuss only ideal fluids which are free from dissipative forces. For simplicity, it is also assumed that the flow pattern is static, which means that at any given point in the fluid the velocity remains constant even though the particles of the fluid are always in motion from point to point. When this is true, one can sketch representative paths followed by the particles. Such paths are called streamlines, and the velocity vectors of the various particles along the streamline are tangent to the streamline. This latter requirement means that streamlines cannot cross, since to do so would involve points of intersection at which the velocity vector would have to be double-valued in order to be simultaneously tangent to both streamlines. In considering the fluid confined between two surfaces formed by sets of adjacent streamlines, the absence of intersecting streamlines leads to a view of the fluid flow as occurring in sheets or layers between the two streamline surfaces. For this reason, laminar flow is an alternative description of streamline flow.

A tube of flow exists if the boundary of a given portion of flowing fluid is a set of adjacent streamlines. The assumption of an ideal fluid means that a tube of flow and the flow pattern in the containing pipe should be the same at any given cross section of the pipe, since there are no dissipative forces between the pipe and fluid or between adjacent layers of fluid. For the moment, let us also assume the fluid is compressible and that at point 1 all fluid particles have a velocity of magnitude v_1 directed perpendicular to the cross-sectional area A_1 of the pipe, while at point 2 the particle velocities are all of magnitude v_2 directed perpendicular to the cross-sectional area A_2 of the pipe, as in Figure 24-5. In a short time interval Δt, the particles that were crossing A_1 will move a distance $v_1\Delta t$ and those that were crossing A_2 will move a distance $v_2\Delta t$. If the density of the fluid at point 1 is ρ_1, the mass of fluid crossing A_1 in Δt will be approximately given by

$$\Delta m_1 = \rho_1 A_1 v_1 \Delta t_1,$$

and the mass of fluid crossing A_2 in Δt will similarly be

$$\Delta m_2 = \rho_2 A_2 v_2 \Delta t.$$

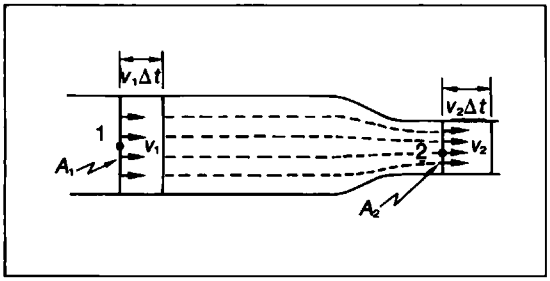


Figure 24-5 Streamline flow in a pipe of varying cross section.

If the time interval becomes infinitesimally small ($\Delta t \to 0$), the mass flux at 1 and 2 become

$$\frac{dm_1}{dt} = \rho_1 A_1 v_1$$

and

$$\frac{dm_2}{dt} = \rho_2 A_2 v_2.$$

In the pipe between points 1 and 2, there are no locations where mass can be removed from the flow (sinks) nor are there locations where mass can be introduced (sources). As a result, the mass flux at any point in the pipe must be identical to that at any other point in the pipe. Therefore, we can write

$$\frac{dm_1}{dt} = \frac{dm_2}{dt}.$$

This result is known as the equation of continuity, and is expressed in the form

$$\rho_1 A_1 v_1 = \rho_2 A_2 v_2. \qquad (24\text{-}12)$$

For incompressible fluids, $\rho_1 = \rho_2$, and the equation of continuity simplifies to the form

$$A_1 v_1 = A_2 v_2. \qquad (24\text{-}13)$$

For liquids then, Eq. (24-13) indicates that where the cross section increases the velocity must decrease, and conversely. In terms of the streamlines of Figure 24-5, it is clear that in regions where the streamlines are close together the fluid velocity is greater than in regions of wider spacing.

Now let us consider the conservation of energy for an ideal *liquid* in streamline or laminar flow. Instead of the pipe of Figure 24-5, we shall use the modification shown in Figure 24-6. In a short time interval Δt, the net work done on the fluid will be

$$W = F_1 v_1 \Delta t - F_2 v_2 \Delta t = (p_1 A_1 v_1 - p_2 A_2 v_2)\Delta t. \qquad (24\text{-}14)$$

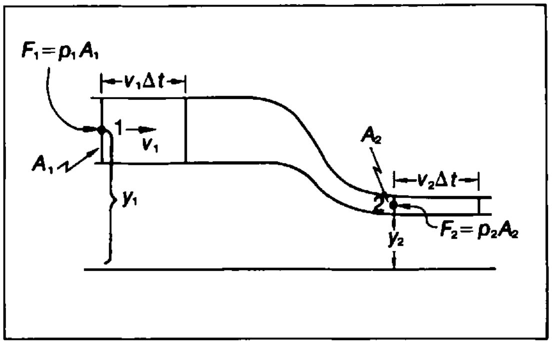


Figure 24-6 Streamline flow in a pipe of varying cross section and elevation.

The work-energy theorem states that

$$W = (\text{K.E.})_F - (\text{K.E.})_I + (\text{P.E.})_F - (\text{P.E.})_I$$
$$= \Delta(\text{K.E.}) + \Delta(\text{P.E.}). \qquad (24\text{-}15)$$

For the system of Figure 24-6,

$$\Delta(\text{K.E.}) = \frac{1}{2}\rho A_2 v_2^3 \Delta t - \frac{1}{2}\rho A_1 v_1^3 \Delta t$$
$$= \frac{1}{2}\rho(A_2 v_2^3 - A_1 v_1^3)\Delta t$$

and

$$\Delta(\text{P.E.}) = \rho A_2 v_2 \Delta t g y_2 - \rho A_1 v_1 \Delta t g y_1$$
$$= \rho g (A_2 v_2 y_2 - A_1 v_1 y_1)\Delta t.$$

Using Eq. (24-13),

$$\Delta(\text{K.E.}) = \frac{1}{2}\rho(v_2^2 - v_1^2)(A_1 v_1 \Delta t),$$
$$\Delta(\text{P.E.}) = \rho g (y_2 - y_1)(A_1 v_1 \Delta t). \qquad (24\text{-}16)$$

Combining Eqs. (24-14), (24-15), and (24-16), we obtain

$$p_1 - p_2 = \frac{1}{2}\rho(v_2^2 - v_1^2) + \rho g (y_2 - y_1) \qquad (24\text{-}17)$$

or

$$p + \frac{1}{2}\rho v^2 + \rho g y = \text{constant}.$$

Equation (24-17) is known as Bernoulli's equation for the streamline flow of an incompressible ideal fluid. From the form of our derivation, it is clear that it is just the statement of conservation of energy for a moving ideal fluid. It might also be noted that Eq. (24-17) reduces to Eq. (24-7) when $v_1 = v_2 = 0$, as it should for energy conservation in a static situation.

24-6 APPLICATIONS OF BERNOULLI'S EQUATION

There are a variety of devices for determining fluid speeds by means of pressure measurements and Eq. (24-17). One such device is the Venturi meter, which consists of a manometer tube (simply a U-tube) located so that it can measure the pressure difference between a point in the main body of a pipe and a point in a constricted portion of the pipe (called the throat) as shown in Figure 24-7. The

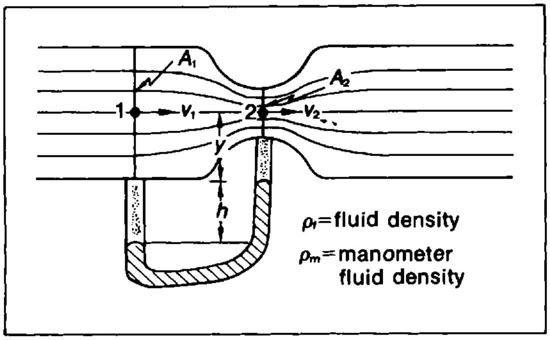


Figure 24-7 A Venturi meter.

value of v is to be found with the device. Since the pipe is horizontal, $y_2 = y_1$ in Eq. (24-17), measuring from the axis of the pipe. As a result, we can write

$$p_1 + \frac{1}{2}\rho_f v^2 = p_2 + \frac{1}{2}\rho_f v_2^2 \qquad (24\text{-}18)$$

Now

$$p_1 - p_2 = (\rho_m - \rho_f)gh. \qquad (24\text{-}19)$$

This is because

$$p_1 + \rho_f g(h + y) = p_2 + \rho_m gh + \rho_f gy$$

is the pressure at the level of the top of the left-hand column of manometer fluid. From Eq. (24-13), we find

$$v_2 = \frac{A_1}{A_2} v_1. \qquad (24\text{-}20)$$

Combining Eqs. (24-18), (24-19), and (24-20),

$$(\rho_m - \rho_f)gh = \frac{1}{2}\rho_f\left[\left(\frac{A_1}{A_2}\right)^2 - 1\right]v^2$$

and, therefore,

$$v = \left[\frac{2(\rho_m - \rho_f)gh}{\rho_f\left[\left(\frac{A_1}{A_2}\right)^2 - 1\right]}\right]^{1/2}. \qquad (24\text{-}21)$$

Example 3. Water flows in a pipe of cross-sectional areas $A_1 = 200\,\text{cm}^2$, $A_2 = 50\,\text{cm}^2$. The manometer fluid is mercury ($\rho_m = 13.6\,\text{gm/cm}^3$, $\rho_f =$ gm/cm^3), and h is found to be 30.4 cm. Find:
(a) v_1 and
(b) $Q = A_1 v$, the volume flux in the pipe.

SOLUTION

(a) Using Eq. (24-21),

$$v = \left[\frac{2(12.6\,\text{gm/cm}^3)(980\,\text{cm/sec}^2)(30.4\,\text{cm})}{(1\,\text{gm/cm}^3)\left[\left(\frac{200\,\text{cm}^2}{50\,\text{cm}^2}\right)^2 - 1\right]}\right]^{1/2}$$

$$= 224\,\text{cm/sec}.$$

(b)

$$Q = (200\,\text{cm}^2)(224\,\text{cm/sec}) = 4.48 \times 10^4\,\text{cm}^3/\text{sec}.$$

Notice that continuity requires that $Q = A_2 v_2 = A_1 v_1$.

Although our derivation of Bernoulli's equation was for an ideal liquid, it can be used to give a qualitative explanation for the curving of a pitched spinning baseball in air, a compressible fluid. The air flow past a stationary spinning ball, as indicated in Figure 24-8, will be the same as that for a spinning ball moving through stationary air. Due to the roughness of the ball cover, some air will be carried in rotation with the ball. As a result, the net velocity on either side of the ball will be the vector sum of the flow and rotational air velocities. In Figure 24-8, the velocity at the right is a difference, while the velocity at the left is a sum of the two velocities.

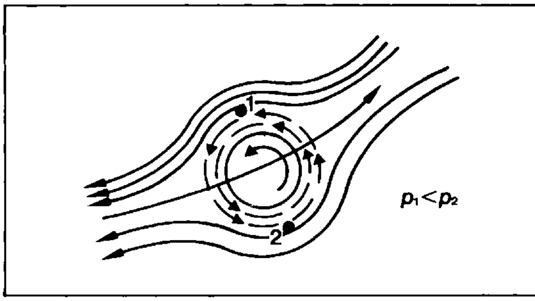


Figure 24-8 Top view—air flow around a spinning ball.

Therefore, $v_{1\,\text{net}} > v_{2\,\text{net}}$. Since $y_1 = y_2$, it follows that $p_1 < p_2$, and the pressure difference results in a side thrust to the left that creates the curved path.

In concluding this section, two warnings are appropriate. Many devices measure not the total or absolute pressure at a point but the gauge pressure, which is the difference between the absolute pressure and atmospheric pressure. Bernoulli's equation requires absolute pressures in all calculations. Furthermore, it is necessary to check carefully to be certain that all quantities involved in Bernoulli's equation are expressed in a single system of units. Otherwise all efforts will be futile.

PROBLEMS

1. What is the pressure at the bottom of a vessel 76 cm deep when filled with water?
2. Rain falls into a boat pulled up on the beach; can it be emptied by a siphon? If the boat were floating on the water, could it be emptied? A siphon is a pipe or tube bent to form two legs of unequal length by which a liquid can be transferred to a lower level over an intermediate elevation.
3. In a hydraulic press, the large piston has cross-section area $A_1 = 200\ \text{cm}^2$, and the small piston has cross-section area $A_2 = 5\ \text{cm}^2$. If a force $F_2 = 25$ nt is applied to the small piston, what is the force on the large piston?
4. A large tank contains water to a depth of 4 m, and a round hole $1\ \text{cm}^2$ in area is opened 1 m from the bottom. Find the rate of flow in liters/min.
5. If the velocity of water in a pipe $10\ \text{cm}^2$ in area is 2 m/sec, what is the velocity in a pipe $5\ \text{cm}^2$ in area that connects with it, both pipes flowing full?
6. If a baseball 4 inches in diameter is thrown through the air with a velocity of 20 ft/sec and is rotating once per second, as shown in Figure 24-9, compare the velocities of the air at the top side and bottom side of the ball.

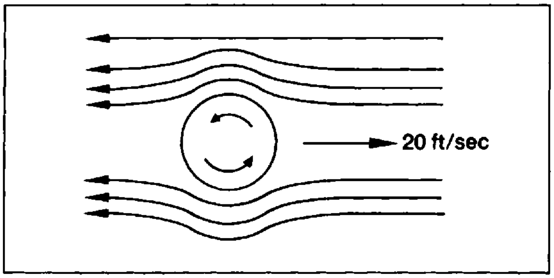


Figure 24-9

7. A glass ball of mass 25 gm has an apparent mass of 15 gm in water and 13 gm in brine. Find the density of the glass and the brine.
8. Because the density of air (ρ_a) is so small, the mass of an object ($\rho_x V_x$) will not differ greatly from the mass of the "weights" ($\rho_w V_w$). Show that, to a good approximation, Eq. (24-9) can be written

$$\rho_x V_x \simeq \rho_w V_w + \rho_a \left[\frac{\rho_w}{\rho_x} - 1\right] V_w.$$

Using

$$m_x = \rho_x V_x, \quad m_w = \rho_w V_w,$$

this can be written

$$m_x = m_w \left[1 + \frac{\rho_a}{\rho_w}\left(\frac{\rho_w}{\rho_x} - 1\right)\right].$$

9. A block of oak ($\rho_x = 0.720\ \text{g/cm}^3$) is balanced by brass weights ($\rho_w = 8.70\ \text{gm/cm}^3$) of mass $m_w = 10$ gm. Taking the density of air to be $\rho_a = 1.3 \times 10^{-3}\ \text{gm/cm}^3$, find:
   (a) the true mass m_x of the oak and
   (b) the percentage error involved if the buoyant effect of air is neglected.
10. What is the approximate volume of a floating log of specific gravity 0.60 which just supports a 160 lb man?

11. The pipe illustrated in Figure 24-10 carries water in streamline flow. The water discharges into the atmosphere with a velocity of 60 ft/sec at point B. The cross-sectional area of the pipe is 0.03 ft^2 at point B and 0.09 ft^2 at point A. The elevation of point B is 15 ft above the level of A. Determine the gauge pressure at point A.

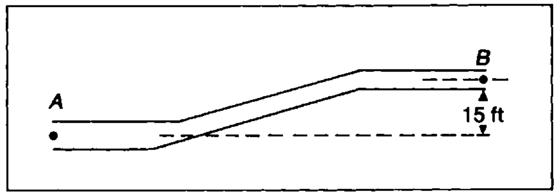


Figure 24-10

12. A frictionless fluid flows through the horizontal system shown in Figure 24-11. Points 1, 2, and 3 are on the axis of the system, and point 3 is at the wall. At point 1, the pipe has a cross-sectional area of 80 cm^2; at point 2, the area is 20 cm^2, and the two side tubes each have areas of 10 cm^2. The fluid flowing from the left discharges from the system at a rate of 0.02 m^3/sec.
    (a) Determine the speed of the fluid at points 1, 2, and 3.
    (b) Determine the pressure of the fluid at points 1, 2, and 3.

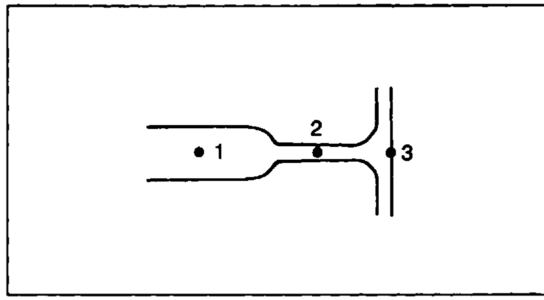


Figure 24-11

13. The volume flow rate of water in the pipe shown in Figure 24-12 is 12 ft^3/sec. The cross-sectional area at points 1, 2, and 3 are 0.4 ft^2, 0.3 ft^2, and 0.3 ft^2, respectively. The water discharges from the right end of the pipe into the atmosphere.
    (a) Find the velocities of flow at points 1 and 2.
    (b) Find the heights h_1 and h_2 to which the water will rise in the two vertical tubes.

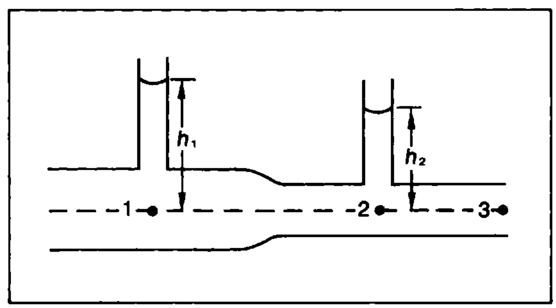


Figure 24-12

14. The section of pipe shown in Figure 24-13 carries water in streamline flow. At point A, the water discharges into the atmosphere at the rate of 10 ft^3/sec. The cross-sectional area of the pipe at point A is 0.25 ft^2 and 1 ft^2 at point B. Point B is 8 ft above point A. What is the gauge pressure at point B?

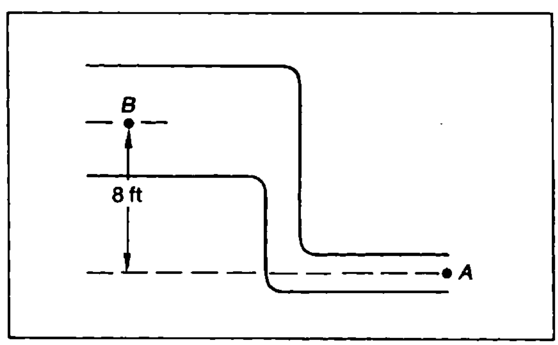


Figure 24-13

15. The section of pipe shown in Figure 24-14 carries water in streamline flow. At point A, the water discharges into the atmosphere at the rate of 9 ft^3/sec. The cross-sectional area of the pipe at point A is 0.75 ft^2 and 0.5 ft^2 at point B. Point B is 12 ft below point A. What is the gauge pressure at point B?

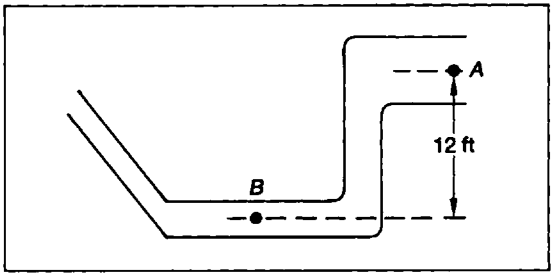


Figure 24-14

25 The First Law of Thermodynamics and Thermodynamic Processes

Thermodynamics is a subject that involves the performance of work, the flow of heat, and the change in energy of a system. The laws of thermodynamics are important to many branches of engineering, physics, and chemistry. In this chapter, we will discuss work, heat and internal energy, and the mechanical equivalent of heat. After a discussion of the first law of thermodynamics and thermodynamic processes, we will consider the specific heat of an ideal gas and the special features of an adiabatic process for an ideal gas.

25-1 WORK

Here we shall discuss work in terms of it being done on or by a system. This is called external work. Internal work, such as the work done by one part of a system on another or by atoms or molecules on one another, is not considered in thermodynamics.

We begin by showing that the work done on or by a system is equal to the pressure exerted on or by the system times the change in volume of the system. It will further be shown that this work depends upon how the process is carried out.

Suppose we consider a gas contained in a cylinder of cross-sectional area A fitted with a movable piston subjected to an external force F. If the piston is moved a small distance Δd by the force F, as illustrated in Figure 25-1, an amount of work W equal to

$$W = F\Delta d \tag{25-1}$$

is done on the system. Since force is equal to the product of pressure and area, Eq. (25-1) may be written as

$$W = pA\Delta d. \tag{25-2}$$

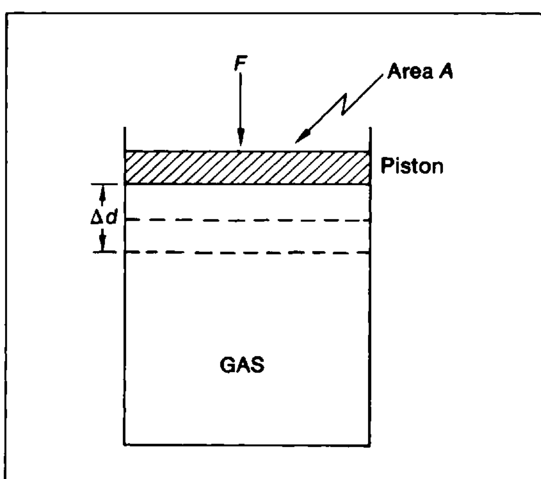


Figure 25-1 Work done on a gas by a force F.

But

$$A\Delta d = \Delta V, \qquad (25\text{-}3)$$

where ΔV is the change in the volume occupied by the gas. Therefore, the work done on the system is given by

$$W = p\Delta V. \qquad (25\text{-}4)$$

Example 1. One kilogram of water of volume $1\ m^3$ is all changed to steam at a volume $1671\ m^3$ when boiled at atmospheric pressure of $1.01 \times 10^6\ nt/m^2$. What is the external work done by the system?

SOLUTION

$$\text{External Work} = p\Delta V$$
$$= (1.01 \times 10^6\ nt/m^2)[(1671-1)m^3]$$
$$= 1.69 \times 10^9\ \text{joules}.$$

So far, we have considered work done under constant pressure. If, however, the pressure changes as the volume changes, as in the case of the compression or expansion of a gas, this knowledge can be obtained by plotting the pressure versus the volume of the gas as work is done on or by the gas. The graph obtained should be similar to Figure 25-2, and the work done will be equal to the area under the curve.

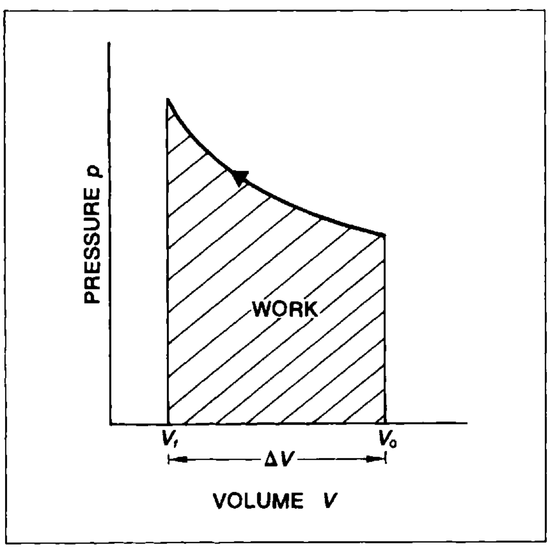


Figure 25-2 Work done when a gas changes its volume.

The area under the curve can be obtained as follows:

$$W = \int_{V_0}^{V_f} p\,dV, \qquad (25\text{-}5)$$

where V_0 is the initial volume and V_f is the final volume. Since the pressure changes as the volume changes, the integration cannot be performed analytically unless the pressure can be expressed as a function of the volume. If the gas can be treated as an ideal gas, the curve can be fitted by Boyle's law or, in other words, the pressure-volume relation obeys Boyle's law. Here the pressure at any point is given by

$$p = \frac{p_0 V_0}{V}$$

and

$$W = \int_{V_0}^{V_f} \frac{p_0 V_0}{V}\,dV = p_0 V_0 \ln \frac{V_f}{V_0}. \qquad (25\text{-}6)$$

It is obvious that the work depends on the path followed. If the curve of Figure 25-2 bowed upward rather than downward, the work done would be greater since the area under it would be greater. We can see, therefore, that the work done on or by a system depends not only on the initial and final states but also on the intermediate states, that is, on the path followed from the initial to the final state.

25-2 HEAT AND INTERNAL ENERGY

We discussed these two concepts briefly at the beginning of Chapter 1. However, it seems a good idea to repeat what we said in addition to making another point regarding both concepts.

Heat is energy transferred between two or more material substances or from one portion of the substance to another on a macroscopic scale. In other words, heat is energy in transit. Heat added to or removed from a system, like work done on or by a system, depends on the path followed from one state of the system to another. Again, suppose we have a gas contained in a cylinder fitted with a weighted movable piston, which can be clamped. Now, consider the two different processes. First we place the cylinder on a hot stove with the piston clamped. Then we place the same cylinder on the hot stove with the piston unclamped, so that it can rise as the gas expands. The amount of heat transfer required to increase the temperature of the gas over the same interval will be different for the two processes. In other words, the heat flow into the system depends on how we go from one state of the system to another.

If work is done on or by the system, and/or heat is added to or removed from the system, we have a change in what is called the *internal energy* of the system. From the molecular point of view, the internal energy of a system is the sum of the total of the kinetic and potential energies of its molecules.

Consider a system that undergoes a change from one state to another by means of several alternative paths. Each path involves a definite heat transfer and a definite amount of work done on or by the system. An important observation will be that for all the paths there will be the same change in internal energy of the system equal to the difference between the heat added and the work done by the system. It follows, therefore, that the change in internal energy of a system is independent of the path.

25-3 THE MECHANICAL EQUIVALENT OF HEAT

Heat energy can be converted into mechanical energy and vice versa. Rumford, in 1798, was the first to demonstrate that the amount of heat created in boring cannon was proportional to the amount of mechanical work expended. Joule was the first to prove experimentally the equivalence of heat and mechanical energy. He used an apparatus in which falling weights rotated a set of paddle wheels in a container of water (Figure 25-3). The mechanical energy was computed from the value of the weights, and their distance of descent and the heat generated was computed from the mass of water and its temperature rise.

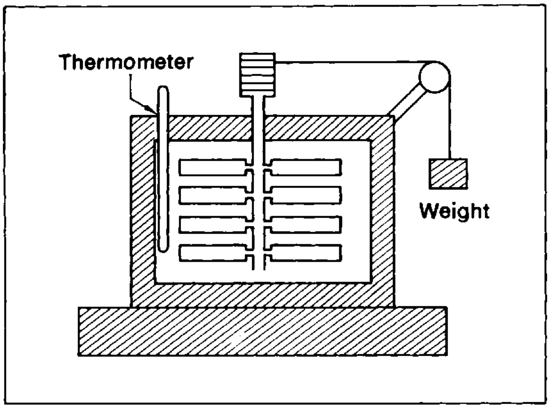


Figure 25-3 Joule's paddle-wheel experiment.

We may state the results of Joule's experiment in the form of an equation

$$W = JQ, \tag{25-7}$$

where W is the work expended in joules, Q is the heat generated in kilocalories, and J is a constant called the mechanical equivalent of heat. The best experiments to date give

$$\begin{aligned} J &= 4186 \text{ joules/kcal} \\ &= 4.186 \text{ joules/cal.} \end{aligned}$$

The conversion of work into heat energy is easily accomplished, and it is possible to approach 100% efficiency, as in the Joule experiment. However, the conversion of heat energy into work is much more difficult and, as will be shown in the following chapter, it is impossible to have 100% efficiency. It is necessary to have a working substance in order to convert heat energy into work. In the steam engine, the working substance is water vapor; in the gasoline engine, it is a mixture of gasoline vapor and air.

25-4 THE FIRST LAW OF THERMODYNAMICS

The first law of thermodynamics is the conservation of energy principle, which states that in the transformation of thermal energy into another form of energy, no energy is created or destroyed. If we have a system into which we put heat or on which work is done, its energy is increased by the heat put into it or the work done on it. Similarly, if we take heat out of it, or if we allow it to do work, its energy is decreased. This energy is that which we have called the internal energy of the system. In other words, the change in the internal energy of a system is equal to the sum of the heat put into or taken out of the system and the work done on or by the system. Symbolically, this statement, which is the first law of thermodynamics, can be written as follows:

$$\Delta U = Q + W. \tag{25-8}$$

ΔU, Q, and W may be either positive or negative. If Q units of heat are put into the system, Q is taken as positive (+) since it tends to increase the internal energy U of the system; if Q units of heat are taken out of the system, Q is taken as negative (−) since it tends to decrease the internal energy U of the system. The sign of ΔU will depend on the signs and relative magnitudes of Q and W. The first law of thermodynamics may be written in differential form if we assume that an infinitesimal amount of heat is transferred and an infinitesimal amount of work is done in the process. The differential form of the first law is

$$dU = dQ + dW. \qquad (25\text{-}9)$$

Example 2. What is the change in the internal energy of a system when 100 cal of heat are supplied to it at the same time that it does 200 joules of work?

SOLUTION

$$\begin{aligned}\Delta U &= Q + W\\ &= (100\text{ cal})(4.186\text{ joules/cal}) - (200\text{ joules})\\ &= 418.6\text{ joules} - 200\text{ joules}\\ &= 218.6\text{ joules.}\end{aligned}$$

25-5 THERMODYNAMIC PROCESSES

We shall consider four thermodynamic processes here—namely, adiabatic, isobaric, isovolumic, and isothermal processes. It should be pointed out that these four processes are not the only processes that can occur in thermodynamics, but they are among the more common ones.

An adiabatic process is one that occurs without transfer of heat to or from the system. Applying the first law of thermodynamics to an adiabatic process, we have

$$\Delta U = W. \qquad (25\text{-}10)$$

If a gas expands adiabatically, it does work on the surroundings and therefore loses internal energy. Conversely, if a gas is compressed adiabatically, its internal energy increases since work is done on the gas.

Example 3. During an adiabatic compression, 50 joules of work is done on the gas. What is the change in internal energy?

SOLUTION

$$\begin{aligned}\Delta U &= Q + W = 0\text{ joules} + 50\text{ joules}\\ &= 50\text{ joules.}\end{aligned}$$

An isobaric process is one that takes place at constant pressure. The work done on or by the system at constant pressure is

$$W = p\,\Delta V, \qquad (25\text{-}11)$$

where ΔV is the change in volume. The first law of thermodynamics for a constant pressure process takes the form

$$\Delta U = Q + p\,\Delta V. \qquad (25\text{-}12)$$

Many chemical processes take place at constant pressure.

Example 4. If 100 cal of heat are added to 10^{-3} m^3 of CO_2 in a cylinder with a movable piston, and the gas expands against an external pressure of 1 atmosphere (1.01×10^5 nt/m^2) to a volume of 1.5×10^{-3} m^3, what is the change in the internal energy of the gas?

SOLUTION

$$\begin{aligned}\Delta U &= Q - W = Q - p\,\Delta V\\ &= (100\text{ cal})(4.186\text{ joules/cal}) - (1.01 \times 10^5\text{ nt/m}^2)\\ &\quad \times (1.5 \times 10^{-3}\text{ m}^3 - 1.0 \times 10^{-3}\text{ m}^3)\\ &= (418.6\text{ joules} - 1.01 \times 0.5 \times 10^2\text{ joules})\\ &= 418.6\text{ joules} - 50.5\text{ joules}\\ &= 368.1\text{ joules.}\end{aligned}$$

An isovolumic process is one that takes place at constant volume. It is sometimes called an isochoric process. Since the volume does not change, no work is done on or by the system, and the first law takes the form

$$\Delta U = Q. \qquad (25\text{-}13)$$

All the heat that is added or removed from the system goes into a change in the internal energy.

Example 5. If 10 cal of heat are added to a gas in a rigid container so that its volume cannot change, what is the change in the internal energy of the gas?

SOLUTION

$$\begin{aligned}\Delta U &= Q + W = Q + p\Delta V = Q + 0\\ &= (10\text{ cal})(4.186\text{ joules/cal})\\ &= 41.9\text{ joules.}\end{aligned}$$

An isothermal process is one that takes place at constant temperature. The performance of an isothermal process requires that the system be in mechanical and thermal equilibrium with a heat reservoir. No real process is ever perfectly isothermal just as no real gas is ever a perfect or ideal gas;

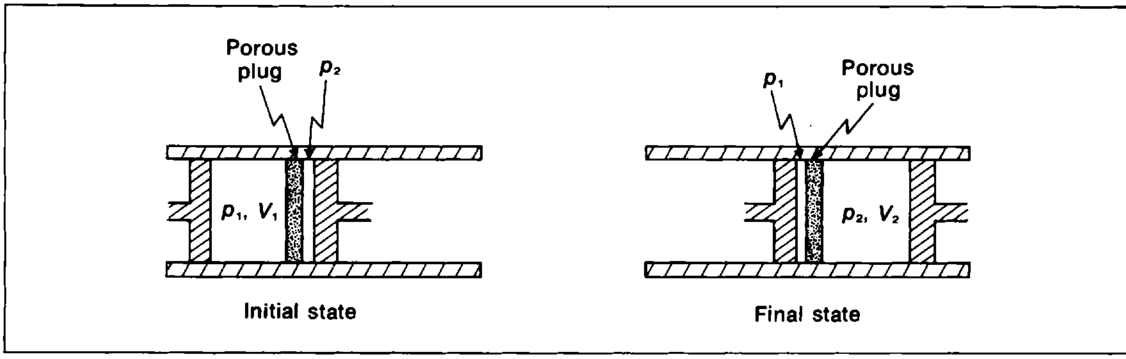


Figure 25-4 Throttling process.

but many processes are quite close to being isothermal. A simple and important isothermal process is the isothermal expansion or compression of an ideal gas. The internal energy of an ideal gas depends only on the temperature; its change is, therefore, zero in an isothermal process. Hence, for an isothermal process involving an ideal gas, $\Delta U = 0$ in the first law and

$$Q = W, \qquad (25\text{-}14)$$

where both Q and W have the same units and can be + or − according to the sign rule adopted for the first law.

Example 6. An ideal gas does 100 joules of work in an isothermal expansion. How many calories of heat must be added to the gas during the process?

SOLUTION

Since the gas is ideal, $\Delta U = 0$ in the expansion. Therefore,

$$0 = W + Q$$

and

$$0 = -100 \text{ joules} + Q$$

$$Q = 100 \text{ joules} = \frac{100 \text{ joules}}{4.186 \text{ joules/cal}} = 23.9 \text{ cal.}$$

Two other thermodynamic processes deserve brief mention. One of these is the *free expansion*, which is an internal energy conservation process, in which a gas is permitted to expand freely into a vacuum without any heat flow into or out of the gas.

In a free expansion,

$$U_i = U_f, \qquad (25\text{-}15)$$

where U_i is the initial internal energy and U_f is the final internal energy. The free expansion process has some theoretical importance but no practical importance.

The other thermodynamic process is known as a *throttling process.* A throttling process is performed by forcing a gas through a porous plug to a region of lower pressure in an insulated cylinder, as illustrated in Figure 25-4. A continuous throttling process can be performed by a pump that maintains a constant high pressure on one side of a porous wall and a constant lower pressure on the other side, as illustrated in Figure 25-5. It can be shown that in a throttling process

$$U_1 + p_1 V_1 = U_2 + p_2 V_2, \qquad (25\text{-}16)$$

where U_1, p_1, and V_1 are the internal energy, pressure, and volume of the gas on one side of the porous wall and U_2, p_2, and V_2 are the internal energy, pressure, and volume on the opposite side of the wall. The quantity $U + pV$ is called the *enthalpy*. The throttling process and Eq. (25-16) are very important in steam engines and refrigerators.

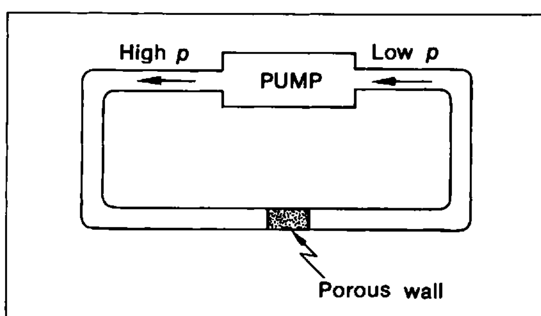


Figure 25-5 Continuous throttling process.

25-6 SPECIFIC HEAT OF AN IDEAL GAS

The specific heat of a gas, unlike a liquid or solid, depends on whether it is measured at constant pressure or constant volume. At constant pressure, the volume of the gas increases when heated, and external work equal to $p\,dV$ is done by the gas. At constant volume, the pressure increases but no external work is done and all of the heat goes into increasing the internal energy of the gas. Therefore, the specific heat at constant pressure is greater than the specific heat at constant volume.

By the use of the first law of thermodynamics, we can obtain an expression that gives the difference between these specific heats for an ideal gas. Suppose Q units of heat are added to n moles of an ideal gas at constant pressure, and the gas undergoes a change in temperature dT. Then,

$$dQ = nC_p\,dT, \qquad (25\text{-}17)$$

where C_p is the molar specific heat at constant pressure. The first law of thermodynamics, applied to the case when heat is added to the gas and the gas expands at constant pressure or does work, is

$$dU = dQ - dW. \qquad (25\text{-}18)$$

Since $dU = nC_v\,dT$, where C_v is the molar specific heat at constant volume, and $dW = p\,dV$, we have

$$nC_v\,dT = nC_p\,dT - p\,dV$$

or

$$n(C_p - C_v)dT = p\,dV. \qquad (25\text{-}19)$$

Differentiating the general gas law, $pV = nRT$ gives

$$p\,dV + V\,dp = nR\,dT.$$

At constant pressure $dp = 0$ and

$$p\,dV = nR\,dT. \qquad (25\text{-}20)$$

Substituting Eq. (25-20) into Eq. (25-19), gives

$$n(C_p - C_v)dT = nR\,dT$$

or

$$C_p - C_v = R. \qquad (25\text{-}21)$$

Equation (25-21) shows that the molar specific heat at constant pressure exceeds the molar specific heat at constant volume for an ideal gas by the universal gas constant R. The relation agrees well with experimental results for simple molecules but not so well for complex ones. The ratio C_p/C_v is important since it is easier to measure experimentally than either C_p or C_v, independently. It is denoted by the Greek letter γ. The value of γ decreases with the complexity of the gas. For a monatomic gas, $\gamma = 1.67$; for a diatomic gas, $\gamma = 1.40$. Approximate values of C_p, C_v, $C_p - C_v$, and C_p/C_v are given in Table 25-1 for a number of real gases.

25-7 ADIABATIC PROCESS FOR AN IDEAL GAS

In an adiabatic process, the pressure changes more rapidly with volume than it does in an isothermal process. In addition to the volume change, there is also a temperature change in an adiabatic process. This leads to a greater change in pressure than is the case for an isothermal process. To see this, consider an ideal gas contained in a perfectly non-conducting cylinder fitted with a movable piston. Let the gas undergo a small adiabatic expan-

Table 25-1 Approximate Molar Specific Heats of Gases.

Type of gas	Gas	$C_p\left(\frac{\text{cal}}{\text{mole°C}}\right)$	$C_v\left(\frac{\text{cal}}{\text{mole°C}}\right)$	$C_p - C_v$	$\gamma = \frac{C_p}{C_v}$
Monatomic	A	5.0	3.0	2.0	1.67
	He	5.0	3.0	2.0	1.67
Diatomic	Cl_2	7.2	6.0	2.1	1.35
	CO	7.0	5.0	2.0	1.40
	H_2	6.9	4.9	2.0	1.41
	N_2	7.0	5.0	2.0	1.40
	O_2	7.0	5.0	2.0	1.40
Polyatomic	CO_2	8.8	6.8	2.0	1.30
	H_2S	8.3	6.2	2.1	1.34
	SO_2	9.7	7.5	2.2	1.30
	Ethane (C_2H_6)	12.4	10.3	2.1	1.20
	Ether [$(C_2H_5)_2O$]	33.3	30.8	2.5	1.08

sion so that there is a change in volume dV. The work done by the gas will equal $p\,dV$. Since $dQ = 0$ in an adiabatic process, and $dU = nC_v dT$ for an ideal gas, we have from the first law

$$nC_v dT = -p\,dV. \qquad (25\text{-}22)$$

In Section 25-6, we saw that

$$p\,dV + V dp = nR\,dT.$$

Eliminating dT between these equations, we obtain

$$nC_v(p\,dV + Vdp) + nRp\,dV = 0.$$

Since $R = C_p - C_v$,

$$C_v Vdp + C_p p dV = 0.$$

Dividing by $C_v V$

$$\frac{dp}{p} + \frac{C_p dV}{C_v V} = 0$$

or

$$\frac{dp}{p} + \gamma\frac{dV}{V} = 0,$$

where

$$\gamma = \frac{C_p}{C_v}.$$

Integration of this equation gives

$$\ln p + \gamma \ln V = \text{constant}$$

or

$$pV^\gamma = \text{constant}. \qquad (25\text{-}23)$$

Equation (25-23) may also be written

$$p_1V_1^\gamma = p_2V_2^\gamma \qquad (25\text{-}24)$$

for any two points of the process. By applying the general gas law to Eq. (25-24), relations between T and V and T and p may be obtained. These relations are

$$T_1V_1^{\gamma-1} = T_2V_2^{\gamma-1} \qquad (25\text{-}25)$$

and

$$p_1^{((1/\gamma)-1)}T_1 = p_2^{((1/\gamma)-1)}T_2. \qquad (25\text{-}26)$$

Figure 25-6 is a $p-V$ diagram for isothermal and adiabatic processes. It can be observed that the adiabatic curves have steeper slopes than the isothermal curves at any one point, and that an adiabatic curve cuts a set of isothermal curves.

Example 7. A volume of CO_2 gas at a temperature of 27°C is compressed adiabatically to $\frac{1}{10}$ of its original volume. Find the final temperature.

SOLUTION

$$T_1V_1^{\gamma-1} = T_2V_2^{\gamma-1}$$

$$300°\text{K}\; V_1^{1.30-1} = T_2\left(\frac{V_1}{10}\right)^{1.30-1}$$

$$T_2 = (300°\text{K})(10^{0.3}) = 597°\text{K} = 324°\text{C}.$$

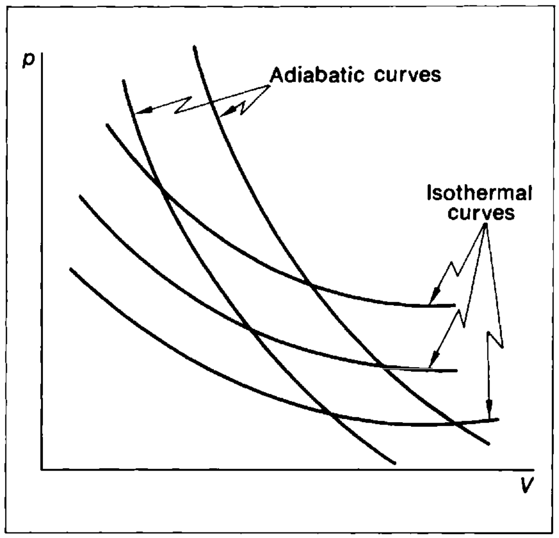


Figure 25-6 *p-V* diagram for adiabatic and isothermal processes.

PROBLEMS

1. A sample of CO_2 in a cylinder with a movable piston is heated by a burner. If 60 cal of heat are added to the gas and the gas expands against an external pressure of 1 atmosphere (1.01×10^5 nt/m^2) from an initial volume of 1000 cm^3 to a final volume of 1600 cm^3, how much work does the gas do on the piston? What is the change in internal energy of the gas?

2. The temperature of 1 mole of nitrogen gas is raised from 10°C to 110°C. Find the heat required to accomplish this at constant volume. If the process is carried out at constant pressure, find:
   (a) the heat added,
   (b) the increase in internal energy, and
   (c) the external work done by the gas.

3. Derive the following formulae for an adiabatic process of an ideal gas, assuming γ to be a constant.

$$TV^{\gamma-1} = \text{constant}$$

and

$$p^{((1/\gamma)-1)}T = \text{constant}.$$

4. In a Wilson cloud chamber at a temperature of 20°C, particle tracks are made visible by causing condensation on ions by an adiabatic expansion of the gas to 1.40 of its initial volume. If $\gamma = 1.40$ for the gas, what is the gas temperature after expansion?

5. Find the work done by a gas initially at a pressure $p_0 = 10^5$ nt/m^2, temperature $t_0 = 27$°C, and volume $V_0 = 1$ m^3 when it:
   (a) expands to a volume of 2 m^3 at constant pressure,
   (b) expands to a volume of 2 m^3 at constant temperature, and
   (c) has its pressure increased to 2×10^5 nt/m^2 at constant volume.

6. Niagara Falls is about 50 m high. What approximate temperature increase would you expect for the water at the bottom of the Falls?

7. Starting with the definition of work W done by an expanding gas, show that for an adiabatic expansion of a gas from a state (p_1, V_1) to a state (p_2, V_2) the work done is given by

$$W = \frac{p_1V_1 - p_2V_2}{\gamma - 1}.$$

8. When a system is taken from state a to state b along the path acb shown in Figure 25-7, 20 cal of heat flow into the system and 30 joules of work are done by the system.
   (a) How much heat flows into the system along the path adb if the work done by the system is 10 joules.
   (b) When the system is returned to a along the curved path, the work done by the system is 20 joules. Does the system absorb or liberate heat, and how much?

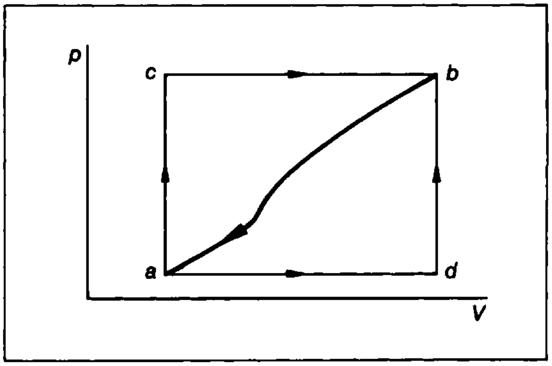


Figure 25-7

9. How much heat must be added to helium gas initially at a pressure of 10^5 nt/m^2, temperature 27°C, and of volume 1 m^3 when the gas:
   (a) expands to a volume of 2 m^3 at constant pressure,
   (b) expands to a volume of 2 m^3 at constant temperature, and
   (c) has its pressure increased to 2×10^5 nt/m^2 at constant volume.

10. When a system is taken from state i to state f along the path iaf in Figure 25-8, it is found that $Q_{fi} = 50$ cal and $W_{fi} = 20$ cal. Along the ibf path $Q_{fi} = 36$ cal.
   (a) What is W_{fi} along the path ibf?
   (b) If $W_{if} = -13$ cal for the curved return path fi, what is Q_{if} for this path?
   (c) Take $U_i = 10$ cal, what is U_f?
   (d) If $U_b = 22$ cal, what is Q_{bi}? What is Q_{fb}?

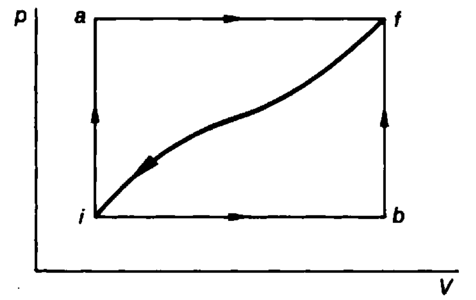


Figure 25-8

11. Helium at 27°C and 1 atmosphere pressure is compressed adiabatically to a pressure of 5 atmospheres. What is the final temperature?
12. A boy pumps up his bicycle tires on a day when the temperature is 27°C and the atmospheric pressure is 15 lb/in^2. Find the maximum temperature of the air in the bicycle pump if the tire pressures are to be 30 lb/in^2 and the air in the pump is assumed to be compressed adiabatically.
13. A block of copper weighing 0.50 kg initially at 500° A is placed in a tank containing 3.78 kg of water at 300° A. The tank is made of copper weighing 1 kg. If these items do not exchange heat with the external surroundings, calculate:
   (a) the change of internal energy of the copper block,
   (b) the change of internal energy of the water, and
   (c) the change of internal energy if all the components are considered a single system.
14. The volume of 2 moles of an ideal gas is increased isothermally from 1 to 10 liters at 27°C. How many joules of work are done?
15. The density of air at 0°C and 1 atmosphere pressure is 0.001293 gm/cm^3. If 10 gms of air are heated from 0°C to 100°C at constant pressure:
   (a) How much will the volume change?
   (b) How many joules of work will be done by the gas on expanding?
   (c) The specific heat or air at constant pressure (c_p) and constant volume (c_v) is 0.24 and 0.17 cal/gm°C, respectively. From the work done in expanding the gas, determined the value of γ.
16. Compute the specific heats c_p and c_v in cal/gm°C for helium, hydrogen, nitrogen, and oxygen.
17. An atomic bomb explosion causes a ball of fire of about 20 m in diameter at a temperature of 200,000°K. Estimate the diameter of the fireball at 2000°C, and state the assumptions you made.
18. Two insulated containers are connected by a pipe with a valve. One container contains 1 mole of helium and the other contains 1 mole of nitrogen. If the valve is opened, what is the final temperature of the mixture?
19. During a process in a closed system, the internal energy of a fluid changes from an initial value of 500 kcal/kg to a final value of 400 kcal/kg. If 90,000 joules/kg of work are performed by the fluid, compute the quantity of heat added to or removed from the fluid during the process.
20. If 22.4 liters of air at 0°C and 1 atmosphere pressure has its pressure suddenly doubled, what is the new temperature and new volume?

26 The Second Law of Thermodynamics and Entropy

26-1 THE SECOND LAW OF THERMODYNAMICS

The first law of thermodynamics, which is a conservation of energy law, places no limitations on the possibility of transforming heat into work, or vice versa. In the transformation of work into heat, a complete conversion is possible. However, there are definite limitations in the transformation of heat into work. The second law of thermodynamics rules out the possibility of converting heat entirely into work.

There are two statements of the second law, one postulated by Kelvin and the other by Clausius. The Kelvin statement is:

> It is impossible to construct an engine that will continuously operate by extracting heat from a source at a higher temperature and converting it entirely into mechanical work.

In other words, it is impossible to construct an engine that is 100% efficient, and that some of the heat taken in by the engine must be ejected. We shall see from the next statement of the second law (the Clausius statement) that the ejected heat must be into a region of a lower temperature than that of the source. The Clausius statement is:

> It is impossible to construct a self-acting machine that will continuously operate by extracting heat from a region at a lower temperature and ejecting it into a region at a higher temperature.

In other words, for a machine to transfer heat from a region of lower temperature to a region of higher temperature, there must be energy supplied to the machine from an outside energy source. An everyday example of this is a home electric refrigerator. In order for heat to be removed from the inside of the refrigerator and ejected into the room, electrical energy must be supplied by an outside source. The Kelvin statement and the Clausius statement of the second law can be shown to be equivalent. The first and second laws of thermodynamics, like all physical laws, are not subject to direct proof. However, all attempts to disprove them have failed.

26-2 THE CARNOT CYCLE

If a substance is taken from an initial state back to the same state, the substance is said to have undergone a cycle of operations. Any cycle that can be made to occur in the reverse direction by very small changes in the conditions is called a reversible cycle. A Carnot cycle is a reversible cycle consisting of two isothermal processes and two adiabatic processes, the details of which are discussed below.

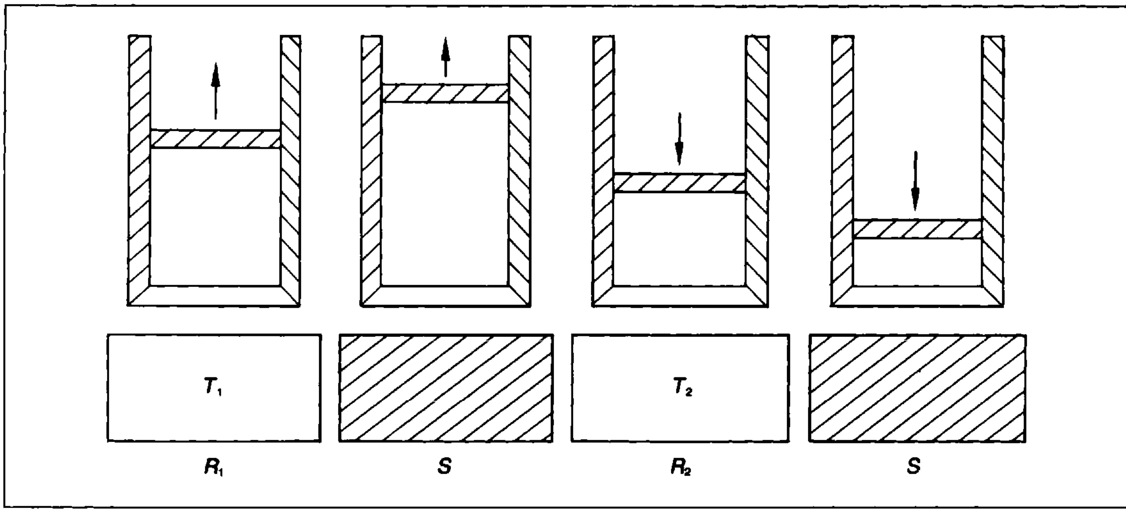


Figure 26-1 The Carnot engine (schematic).

According to the Kelvin statement of the second law of thermodynamics, in order for an engine to convert heat extracted from a source at a higher temperature into work, it is necessary to eject some of this heat into a source of lower temperature. If we have two such sources at temperatures T_1 and T_2, with T_1 greater than T_2, we can transform heat into work by the following process, which is called a Carnot cycle, or the cycle of a Carnot engine.

Consider a fluid contained in a cylinder with insulating walls and a conducting bottom fitted with an insulating piston. Let R_1 and R_2 be two reservoirs of heat at temperatures T_1 and T_2 ($T_1 > T_2$), and S—an insulating stand (Figure 26-1). We shall assume that the heat reservoirs are so large that their temperatures remain unchanged by the addition or subtraction of finite amounts of heat.

With the cylinder on R_1, we raise the piston very slowly so that the fluid undergoes a reversible isothermal expansion at temperature T_1, represented by *AB* in Figure 26-2. The fluid absorbs an amount of heat Q_1 from R_1.

We now place the cylinder on the insulating stand S and again raise the piston very slowly. Because the system is thermally insulated during the process, the fluid undergoes a reversible adiabatic expansion with a decrease in temperature from T_1 to T_2. The process is represented by *BC* in Figure 26-2.

The cylinder is now transferred to R_2, and the piston is pushed down very slowly so that the fluid undergoes a reversible isothermal compression at temperature T_2 represented by *CD* in Figure 26-2. The fluid gives up an amount of heat Q_2 to R_2.

Finally, the cylinder is again placed on the insulating stand S, and the piston is pushed down very slowly until the temperature of the fluid is again T_1. The fluid undergoes a reversible adiabatic compression and is now in its initial state again. The process is represented by *DA* in Figure 26-2.

The net amount of heat absorbed by the system during the cycle is $Q_1 - Q_2$. From the first law of thermodynamics, the net work done (W) by the

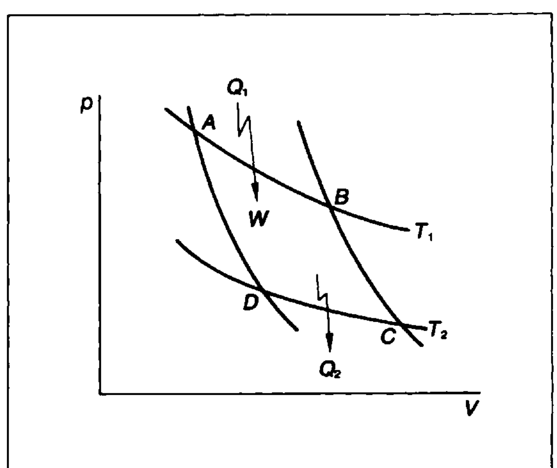


Figure 26-2 The Carnot cycle.

system is

$$W = Q_1 - Q_2. \qquad (26\text{-}1)$$

The efficiency (η) of the Carnot cycle is defined as the ratio of the net or useful work done by the cycle to the heat absorbed from the high temperature source.

$$\eta = \frac{W}{Q_1} = \frac{Q_1 - Q_2}{Q_1}. \qquad (26\text{-}2)$$

Since the Carnot cycle is reversible, the above operations can be carried out in the reverse directions. When this is done, work is done on the system; it absorbs heat Q_2 at a temperature T_2, and rejects heat Q_1 at a temperature T_1 greater than T_2. This is the process performed in the operation of a refrigerator.

Figure 26-3 is a schematic diagram of a Carnot refrigerator. External work W is done on the refrigerator, heat Q_2 enters the refrigerator at low

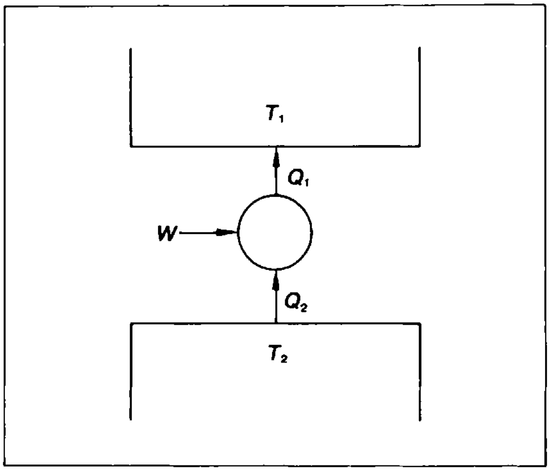


Figure 26-3 Schematic diagram of a refrigerator.

temperature T_2, and heat Q_1 leaves at a high temperature T_1. From the first law of thermodynamics,

$$W = Q_1 - Q_2 = \left(\frac{Q_1 - Q_2}{Q_2}\right)Q_2. \qquad (26\text{-}3)$$

The best refrigerator is the one that removes the greatest amount of heat (Q_2) from the cold reservoir for the least amount of external work (W). We, therefore, define the "coefficient of performance" (ω) of a refrigerator as

$$\omega = \frac{Q_2}{W} = \frac{Q_2}{Q_1 - Q_2}. \qquad (26\text{-}4)$$

26-3 CARNOT'S THEOREM

Carnot's theorem is stated as follows:

> No other engine working between the same temperatures T_1 and T_2 can be more efficient than a reversible engine.

We shall now proceed to prove this theorem. Suppose an irreversible engine I has efficiency η_I greater than the efficiency η_R of a reversible Carnot engine, R, working between the same temperatures T_1 and T_2, and that both engines are adjusted to do the same amount of work. Let Q_1 be the heat absorbed and Q_2 the heat rejected by the Carnot reversible engine (R), and Q_1' and Q_2' the heat absorbed and rejected by the irreversible engine (I).

Because both engines are adjusted to do the same amount of work and η_I is assumed greater than η_R,

$$W = Q_1' - Q_2' = Q_1 - Q_2$$

and

$$\eta_I = \frac{Q_1' - Q_2'}{Q_1'} > \eta_R = \frac{Q_1 - Q_2}{Q_1}. \qquad (26\text{-}5)$$

Since the numerators of the above equation are equal,

$$Q_1' < Q_1 \quad \text{and} \quad Q_2' < Q_2. \qquad (26\text{-}6)$$

Let engine I drive engine R backwards, as illustrated in Figure 26-4. The combined system does nothing but transfer $Q_2 - Q_2' = Q_1 - Q_1'$ from T_2 to T_1. Since this is a violation of the Clausius statement of the second law of thermodynamics, the as-

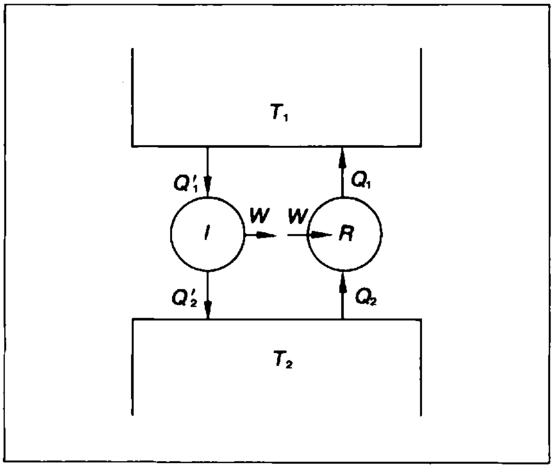


Figure 26-4 Schematic diagram of an irreversible engine driving a reversible Carnot engine backwards.

sumption that $\eta_I > \eta_R$ is false. The conclusion is that $\eta_I \leq \eta_R$, and that Carnot's theorem is true.

We leave it to the student to prove an important corollary to Carnot's theorem, namely:

> All Carnot engines operating between the same two temperatures have the same efficiency.

From this statement, it is obvious that the efficiency of a Carnot cycle is independent of the working substance.

26-4 THE THERMODYNAMIC OR KELVIN TEMPERATURE SCALE

The important corollary to Carnot's theorem stated above shows that the ratio Q_2/Q_1 has the same value for all reversible engines that operate between the same temperatures t_1 and t_2. Therefore, Q_2/Q_1 depends only on the temperatures t_1 and t_2. We may, therefore, write

$$\frac{Q_2}{Q_1} = f(t_1, t_2), \tag{26-7}$$

where f is an unknown function of t_1 and t_2. Consider a sequence of engines, each receiving heat rejected by the previous one (Figure 26-5).

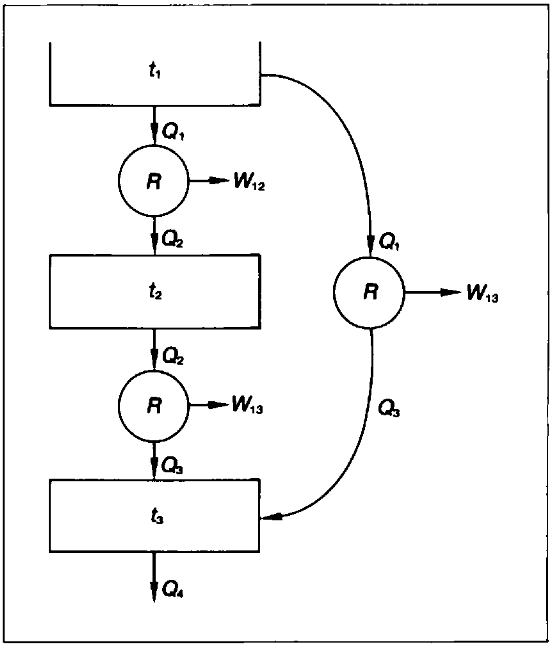


Figure 26-5 Schematic diagram of a sequence of Carnot engines.

The efficiencies of the first two engines are given by

$$\eta_{12} = \frac{Q_1 - Q_2}{Q_1} = f(t_1, t_2) \tag{26-8}$$

and

$$\eta_{23} = \frac{Q_2 - Q_3}{Q_2} = f(t_2, t_3). \tag{26-9}$$

From Eq. (26-8),

$$\frac{Q_1}{Q_2} = \frac{1}{1 - \eta_{12}} = F(t_1, t_2), \tag{26-10}$$

where F is also an unknown function. The first two engines working together constitute a third engine and

$$\frac{Q_1}{Q_3} = F(t_1, t_3).$$

Therefore,

$$\frac{Q_1}{Q_2} = \frac{Q_1}{Q_3}\frac{Q_3}{Q_2} = \frac{F(t_1, t_3)}{F(t_2, t_3)}. \tag{26-11}$$

Hence, from Eqs. (26-10) and (26-11),

$$F(t_1, t_2) = \frac{F(t_1, t_3)}{F(t_2, t_3)},$$

for any t_3. Since t_3 does not appear on the left-hand side of the above equation, it may be cancelled, and we may write

$$\frac{Q_1}{Q_2} = \frac{\theta(t_1)}{\theta(t_2)}, \tag{26-12}$$

where θ is another unknown function. The ratio on the right side of Eq. (26-12) is defined as the ratio of the two Kelvin temperatures, and is denoted by T_1/T_2. We may, therefore, write Eq. (26-12)

$$\frac{Q_1}{Q_2} = \frac{T_1}{T_2}. \tag{26-13}$$

It should be pointed out that the temperatures in Eqs. (26-7) through (26-12) refer to arbitrary temperature scales. Equation (26-13) defines the Kelvin temperature scale. Equation (26-13) states that the ratio of two Kelvin temperatures is the ratio of the heat absorbed to the heat rejected by a reversible engine operating between these two temperatures, and is independent of the working substances.

The definition of the Kelvin scale is completed by assigning to T_1 the value of 273.16°K, the temperature of the triple point of water as in Chapter 1. For a Carnot engine operating between temperatures T and T_1, we have

$$T = 273.16°\text{K}\frac{Q}{Q_1}. \qquad (26\text{-}14)$$

The smaller the value of Q, the lower the corresponding temperature T. The smallest value of Q is zero, and the corresponding T is absolute zero. It was pointed out that in a Carnot cycle the heat is transferred during the isothermal processes. Therefore, if a system undergoes a reversible isothermal process without heat transfer, the temperature at which the process takes place is absolute zero. This is a fundamental definition of absolute zero, since it is independent of the properties of the substance.

Making use of Eq. (26-13), we may now write the efficiency of a Carnot engine and the coefficient of performance of a Carnot refrigerator in terms of the Kelvin temperatures as follows:

$$\eta = \frac{T_1 - T_2}{T_1} \qquad (26\text{-}15)$$

and

$$\omega = \frac{T_2}{T_1 - T_2}. \qquad (26\text{-}16)$$

For a Carnot engine, it is only possible to have 100% efficiency if $T_2 = 0$. Since absolute zero temperature is impossible to attain, a 100% efficient heat engine is impossible. In a real engine, the maximum efficiency is attained by making the intake temperature as high as possible and the exhaust temperature as low as possible. For a Carnot refrigerator, Eq. (26-3) may be written

$$W = \left(\frac{T_1 - T_2}{T_2}\right)Q_2 = \left(\frac{T_1}{T_2} - 1\right)Q_2. \qquad (26\text{-}17)$$

From this equation, it is clear that the greater the temperature ratio T_1/T_2, the more work is required to remove a given quantity of heat from the cold reservoir. This is true for real refrigerators as well as Carnot refrigerators.

Example 1. Is it possible to have an engine that takes in 25×10^6 cal of heat from its fuel, rejects 5×10^6 cal of heat in the exhaust, and delivers 10^7 joules of mechanical work?

SOLUTION

$$Q_1 - Q_2 = (25 \times 10^6 \text{ cal}) - (5 \times 10^6 \text{ cal}) = 20 \times 10^6 \text{ cal}$$

$$= (4.186 \text{ joules/cal})(20 \times 10^6 \text{ cal})$$

$$= 8.372 \times 10^6 \text{ joules}.$$

It is stated that

$$W = 10^7 \text{ joules}.$$

This is impossible, since the engine would deliver more energy than it received from the fuel.

26-5 ENTROPY

If a substance undergoes a reversible process, and takes in an amount of heat dQ at a temperature T, its entropy (S) is said to increase by an amount dQ/T. We therefore define the change in entropy of a substance by the relation

$$dS = \frac{dQ}{T}. \qquad (26\text{-}18)$$

Let us first consider the change in entropy of a substance that takes place when it is carried through a Carnot cycle. For the adiabatic portions of the cycle, no heat enters or leaves the substance, so here the change in entropy is zero. During the isothermal expansion part of the cycle, the substance takes in heat Q_1 at a temperature T_1, and its entropy increases by an amount Q_1/T_1. During the isothermal compression part of the cycle, the substance gives out an amount of heat Q_2 at a temperature T_2, and its entropy decreases by an amount Q_2/T_2. The total change in entropy during the whole process is

$$\left(\frac{Q_1}{T_1} - \frac{Q_2}{T_2}\right).$$

From Eq. (26-13), we have

$$\frac{Q_1}{T_1} = \frac{Q_2}{T_2}.$$

Therefore, the total change in entropy is zero.

Any reversible cycle may be replaced by a set of isothermal and adiabatic processes or divided into a series of Carnot cycles as illustrated in Figure 26-6. The total change in entropy for all the Carnot cycles will be

$$\sum \frac{\Delta Q_c}{T_c} = 0, \qquad (26\text{-}19)$$

where $\Delta Q_c/T_c$ is the entropy change for each Carnot cycle. By enclosing a large number of Carnot cycles in the reversible cycle, we may approximate the reversible cycle more closely. In the limit Eq. (26-19) transforms to an integral around the reversible cycle and

$$\oint_R \frac{dQ}{T} = 0. \qquad (26\text{-}20)$$

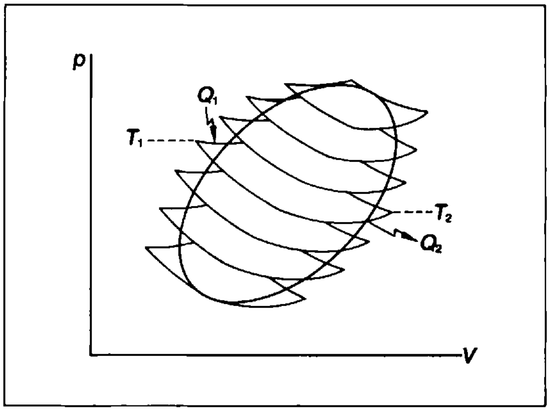


Figure 26-6 A reversible cycle divided into a series of Carnot cycles.

It follows that the change in entropy is the same for any reversible path leading from one state to another, otherwise Eq. (26-20) would not be true. We may then write that the entropy change of a system taken from state 1 to state 2 along a reversible path is given by

$$S_2 - S_1 = \int_1^2 \frac{dQ}{T}. \qquad (26\text{-}21)$$

So far we have discussed entropy in terms of a reversible process, which is only a hypothetical process and can never be realized in practice no matter how closely actual changes may approximate it. Some books on thermodynamics discuss entropy almost solely on reversible processes and only discuss irreversible changes in a meager way. This seems to us a serious error as it gives the reader a confused notion of the subject.

It can be shown that any process in which the total entropy change of the system and its surroundings is negative violates the second law of thermodynamics. Therefore, the entropy change must be equal to or greater than zero, and the second law of thermodynamics may be written mathematically as

$$S_f - S_i \geq 0 \qquad (26\text{-}22)$$

when S_i is the initial entropy of the universe and S_f is its final entropy. If only reversible cyclic processes occur, the entropy change is zero, otherwise it is positive. Therefore, since all natural processes are irreversible, and hence increase the entropy of the universe, we may make the statement that the entropy of the universe is continually increasing.

The term entropy is used as a measure of energy unavailable for work. If we have two separate reservoirs of water, one hot and the other cold, a heat engine could be devised to perform mechanical work by removing heat from the hot reservoir and giving heat to the cold reservoir. However, if the hot and cold water are mixed, and come to a uniform temperature, this energy is no longer available, since there is no temperature difference. In this example, the energy of the system has remained constant, but the entropy has increased. The warm water will never unmix itself, and we have an irreversible process. Since irreversible processes are continually going on in nature, energy is continually becoming unavailable for work. This is known as the *principle of the degradation of energy*.

Entropy is also a measure of the disorder of a system. If heat is added to a system, there is a greater random motion of the molecules, and the system becomes more disordered. Adding a given quantity of heat to a system at a low temperature increases the disorder more than adding the same quantity of heat to the same system at a high temperature. It is reasonable, therefore, to define the change in disorder of the system when dQ units of heat are added to it at a temperature T by dQ/T which is equal to the change in entropy of the system.

We showed that the change in entropy of a system for any reversible path is given by Eq. (26-21). We shall conclude the discussion on entropy by saying that the change in entropy is the same for an irreversible path as for a reversible one between the same two states, since change in entropy is a function of the initial and final states only and not on the path followed between the two states.

We can, therefore, calculate the change in entropy for an irreversible process by replacing it with a reversible one and using Eq. (26-21). The change in entropy can also be calculated if S_2 and S_1 are known for the two states as a function of, say, V and T or from tables. Here the change in entropy will be just $S_2 - S_1$.

Example 2. What is the change in entropy of 1 mole of an ideal gas that expands to twice its original volume into an evacuated vessel?

SOLUTION

This is an irreversible process, and must be replaced by a reversible one. Since there is no temperature change in the free expansion of an ideal gas, it can be replaced by a reversible isothermal expansion if we imagine that the gas is in a cylinder with a piston which is allowed to be pushed out slowly by gradually reducing the exterior pressure on it. The internal energy does not change, but heat enters the gas and work is done by it. From the first law of thermodynamics,

$$dQ = dW = p\,dV = \frac{RT}{V}dV.$$

From Eq. (26-21),

$$S_2 - S_1 = \int_V^{2V} \frac{dQ}{T} = R\int_V^{2V} \frac{dV}{V} = R \ln 2.$$

PROBLEMS

1. A Carnot engine, whose high temperature reservoir is at 400°K, takes in 100 cal of heat at this temperature in each cycle and gives up 80 cal of heat to the low temperature reservoir.
   (a) What is the temperature of the low temperature reservoir?
   (b) What is the efficiency of the engine?
2. A Carnot engine absorbs 400 cal of heat from a source and rejects 300 cal to a reservoir.
   (a) Find the efficiency.
   (b) Find the ratio of the temperature of the source to the temperature of the reservoir.
   (c) Reverse the Carnot engine given above so that it becomes a refrigerator. Find the coefficient of performance.
3. A Carnot engine during each cycle accepts 14 cal from the hot reservoir and rejects 6 cal to the cold reservoir.
   (a) How much mechanical work is done per cycle?
   (b) What is the thermal efficiency?
   (c) What is the ratio of the Kelvin temperature of the hot reservoir to the Kelvin temperature of the cold reservoir?
4. The motor in a refrigerator has a power output of 200 watts. If the freezing compartment is at 270°K and the outside air is at 300°K, assuming ideal efficiency, what is the maximum amount of heat that can be extracted from the freezing compartment in 10 minutes?
5. The entropy of saturated water at 100°C is 0.31 cal/gm°C and that of saturated steam at the same temperature is 1.76 cal/gm°C. What is the heat of vaporization at this temperature?
6. When 1 kg of water at 0°C is mixed with 0.5 kg of water at 50°C, what is the change in entropy?
7. A refrigerator has a coefficient of performance equal to one-quarter that of a Carnot refrigerator. The refrigerator is operated between two reservoirs at temperatures of 200°K and 300°K. It absorbs 400 joules from the low temperature reservoir. How much heat is rejected to the high temperature reservoir?
8. What is the thermal efficiency of an engine which operates by taking n moles of an ideal gas through the following cycle?
   (a) Start at p_0, V_0, T_0;
   (b) change to $3p_0$, V_0;
   (c) change to $3p_0$, $3V_0$;
   (d) change to p_0, $3V_0$;
   (e) change to p_0, V_0.
9. A Carnot refrigerator takes heat from water at 0°C and rejects 1000 kcal of heat to the room at 27°C.
   (a) How much water is converted to ice?
   (b) How much energy must be supplied to the refrigerator?
10. How much work must be done by a refrigerator to remove 1 kcal of heat from a reservoir at − 73°C and reject it to a reservoir at 27°C if the refrigerator has a coefficient of performance equal to 80% of a Carnot refrigerator?

11. An engine of 20% thermal efficiency is used to drive a refrigerator having a coefficient of performance of 4. What is the ratio of the heat input to the engine to the heat removed from the cold reservoir by the refrigerator?
12. Calculate the change in entropy when a 2 kg block of aluminum at 727°C is brought in contact with 1 kg of water at 27°C.
13. Calculate the change in entropy of 2 kg of water when it is raised from the freezing point to the boiling point temperature.
14. Show that, for any cycle,

$$\oint \frac{dQ}{T} \leq 0.$$

15. Show, by applying the result of Problem 14, that the entropy of an irreversible process is

$$dS > \frac{dQ}{T},$$

and in the limiting case of the process in an isolated system,

$$dS > 0.$$

27 Kinetic Theory of Matter

In this chapter we will discuss pressure, temperature, and specific heat in terms of molecular motion.

27-1 MATTER

The material of which all bodies are made is called matter. There are three different states of matter—solid, liquid, and gas. So far we have discussed the macroscopic properties of matter. However, all matter is made up of very small particles called molecules. There are spaces between the molecules, which are called intermolecular spaces; all macroscopic properties of matter are really based on the properties of these molecules. Forces called intermolecular forces exist between the molecules. Although the intermolecular forces are very strong, the distance through which these forces act is very small. The molecules of matter are in more or less rapid motion, and the temperature of a given substance is determined by the average velocity of its molecules. Differences exist in the kind of motion possessed by the molecules in three states. In gases, the intermolecular forces are very small and there is complete random motion. In solids, where the intermolecular forces are relatively large, the molecules oscillate about certain fixed points, being held by the attractive forces of their neighbors. In liquids, the molecules have no fixed positions and slip about with ease over one another.

27-2 KINETIC THEORY OF GASES

We have previously discussed the macroscopic behavior of gases and the equation of state (the general gas law) for an ideal gas. We are now in a position to develop a microscopic theory of gases and to derive an equation of state for an ideal gas from a molecular model. We shall first make a few assumptions:

1. All molecules are in motion in all directions.
2. Molecules move in straight lines between collisions.
3. All collisions are perfectly elastic.
4. Diameters of molecules are neglected in comparison to the distance traveled between collisions.
5. Intermolecular forces are very small.
6. Time spent during collision is much less than time spent between collisions.

Now let us consider n ideal gas molecules,

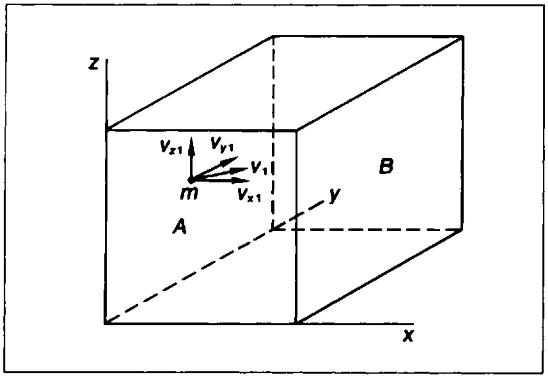


Figure 27-1 Molecule with a velocity v_1 in a cubical container.

each of mass m, contained in a cubical box of unit edge length as illustrated in Figure 27-1. Consider a molecule moving from face A to face B with a velocity v_1. Let the component of the velocity in the x-direction be v_{x1}. The change in momentum on impact with face B will equal $2mv_{x1}$. Since the time for the molecule to travel across the cube and back will be $2/v_{x1}$, the number of impacts per second with face B will equal $v_{x1}/2$. From the above information, we now know that the rate of change of momentum produced at face B in the x-direction, which equals the average force F_{x1}, is

$$F_{x1} = (2mv_{x1})\frac{v_{x1}}{2} = mv_{x1}^2. \qquad (27\text{-}1)$$

Since pressure is force divided by area and the area of each face of the cube is unity, the average pressure due to this molecule is

$$p_1 = mv_{x1}^2. \qquad (27\text{-}2)$$

Let us now consider other molecules with x-components of velocities v_{x2}, v_{x3}, etc. Since pressures add the total pressure, p, with all molecules of the same mass, is given by

$$p = p_1 + p_2 + \cdots = m(v_{x1}^2 + v_{x2}^2 + \cdots). \qquad (27\text{-}3)$$

Let n be the total number of molecules in the unit cube. The average value of the square of the x-component of velocities of all the molecules is

$$\overline{v_x^2} = \frac{v_{x1}^2 + v_{x2}^2 + \cdots}{n}, \qquad (27\text{-}4)$$

and Eq. (27-3) can be written

$$p = nm\overline{v_x^2}. \qquad (27\text{-}5)$$

Similar expressions can be obtained for the y- and z-directions, namely,

$$p = nm\overline{v_y^2} \quad \text{and} \quad p = nm\overline{v_z^2}.$$

Since

$$\overline{v^2} = \overline{v_x^2} + \overline{v_y^2} + \overline{v_z^2} \qquad (27\text{-}6)$$

and because the molecules do not move in any preferred direction,

$$\overline{v_x^2} = \overline{v_y^2} = \overline{v_z^2}, \qquad (27\text{-}7)$$

$$\overline{v_x^2} = \overline{v_y^2} = \overline{v_z^2} = \frac{\overline{v^2}}{3}. \qquad (27\text{-}8)$$

Hence, Eq. (27-5) becomes

$$p = \frac{1}{3}nm\overline{v^2}. \qquad (27\text{-}9)$$

If the gas is contained in a volume V,

$$p = \frac{1}{3}\frac{N}{V}m\overline{v^2} \qquad (27\text{-}10)$$

or

$$pV = \frac{1}{3}Nm\overline{v^2}, \qquad (27\text{-}11)$$

where N is the total number of molecules in the volume V.

Since density equals mass per unit volume, we can write Eq. (27-10) as

$$p = \frac{1}{3}\rho\overline{v^2}$$

or

$$V_{rms} = \sqrt{\frac{3p}{\rho}}. \qquad (27\text{-}12)$$

Therefore, the value of $\bar{v}$ can be calculated if we know the pressure and density of the gas at a given temperature. We call $\overline{v^2}$ the *mean-square velocity* of the gas molecules and $\sqrt{\overline{v^2}} = V_{rms}$—the *root-mean-square* velocity. Equation (27-12) states that the *root-mean-square* velocity of the molecules in a gas varies inversely with the square root of the density of the gas at constant pressure and temperature.

Example 1. Calculate the root-mean-square velocity of the molecules in air at 0°C and 1 atmosphere pressure.

SOLUTION

For air at 0°C and 1 atmosphere pressure,

$\rho = 1.296 \text{ kg/m}^3$

$$p = 1 \text{ atm} = 1.013 \times 10^5 \text{ nt/m}^2$$

$$V_{rms} = \sqrt{\frac{3p}{\rho}} = \sqrt{\frac{3 \times 1.013 \times 10^5 \text{ nt/m}^2}{1.296 \text{ kg/m}^3}}$$

$$= 485 \text{ m/sec.}$$

27-3 KINETIC THEORY INTERPRETATION OF TEMPERATURE AND SPECIFIC HEAT

For one mole of an ideal gas, Eq. (27-11) may be written

$$pV = \frac{1}{3} N_0 m\overline{v^2} = \frac{1}{3} M\overline{v^2} = \frac{2}{3}\left(\frac{1}{2} M\overline{v^2}\right), \tag{27-13}$$

where N_0 is the number of molecules in a mole of the gas (Avogadro's number) and M is the molecular mass of the gas. The quantity $\frac{1}{2}M\overline{v^2}$ is the total translational kinetic energy of the gas molecules; for an ideal gas, it equals the internal energy (U) of a mole of the gas.

The equation of state for one mole of an ideal gas is

$$pV = RT. \tag{27-14}$$

Combining Eqs. (27-13) and (27-14),

$$\frac{2}{3}\left(\frac{1}{2} M\overline{v^2}\right) = \frac{2}{3} U = RT. \tag{27-15}$$

The important deduction from this equation is that the internal energy of an ideal gas varies directly with the absolute temperature, and depends only on the temperature. It defines temperature on the basis of the kinetic theory, and gives us an understanding of the temperature of a gas in terms of the motion of its molecules.

Since R and N_0 are both universal constants, their ratio is also a universal constant, called the *Boltzmann constant k*.

$$k = \frac{R}{N_0} = \frac{8.31 \text{ joules/mole-°K}}{6.02 \times 10^{23} \text{ molecules/mole}}$$

$$= 1.38 \times 10^{-23} \text{ joules/molecule-°K.}$$

Equation (27-15) may now be written as

$$\frac{1}{2} m\overline{v^2} = \frac{3}{2} kT, \tag{27-16}$$

which gives the average kinetic energy per molecule of a gas in terms of its temperature.

Since the translational kinetic energy of a molecule depends on the values of the three rectangular components of its velocity, we say that the molecule has three translational degrees of freedom. On the average, it is to be expected that the energy associated with each degree of freedom is the same. Since the total average kinetic energy per molecule is $\frac{3}{2}kT$, each molecule possesses an average kinetic energy of $kT/2$ per degree of freedom. This is a special example of the *Principle of equipartition of energy.*

For a monatomic gas, the molecular energy is all kinetic energy, and its internal energy per mole can be represented by Eq. (27-15)—namely,

$$U = \frac{3}{2} RT. \tag{27-17}$$

If the temperature of the gas is increased by dT, the kinetic energy increases by

$$dU = \frac{3}{2} R\, dT.$$

Since the specific heat at constant volume (C_v) is equal to dU/dT,

$$C_v = \frac{3}{2} R. \tag{27-18}$$

Also, since $C_p = C_v + R$, the ratio $C_p/C_v = \gamma$ is equal to

$$\gamma = \frac{\frac{3}{2}R + R}{\frac{3}{2}R} = \frac{5}{3} = 1.67.$$

The above results for C_v and γ agree well with experimental results.

Boltzmann pointed out that the heat required to raise the temperature of a gas depends on the number of degrees of freedom—that is, the number of independent coordinates necessary to describe its motion. For a diatomic gas, there are 5 degrees of freedom (3 translational and 2 rotational), and for a polyatomic gas—6 degrees of freedom (3 translational and 3 rotational). Therefore, for a diatomic gas,

$$C_v = \frac{5}{2} R$$

and

$$\gamma = \frac{\frac{5}{2}R + R}{\frac{5}{2}R} = \frac{7}{5} = 1.40.$$

These values are in good agreement with experiment.

Example 2. What is the average kinetic energy of a molecule of a gas at a temperature of 27°C?

If the gas is CO_2, what are the values of C_v, C_p, and γ?

SOLUTION

$$\frac{1}{2}mv^2 = \frac{3}{2}KT = \left(\frac{3}{2}\right)(1.38\times 10^{-23}\ \text{joules/°K})(300°\text{K})$$

$$= 6.20\times 10^{-21}\ \text{joules}.$$

Carbon dioxide (CO_2) is a triatomic gas with 3 atoms per molecule and 6 degrees of freedom.

$$C_v = \frac{6}{2}R = 3R = 6\ \text{cal/mole-°C},$$

$$C_p = C_v + R = 4R = 8\ \text{cal/mole-°C},$$

$$\gamma = \frac{C_p}{C_v} = \frac{8\ \text{cal/mole-°C}}{6\ \text{cal/mole-°C}} = 1.33.$$

27-4 AVOGADRO'S LAW AND DALTON'S LAW OF PARTIAL PRESSURES

Avogadro's law is stated as follows:

> Equal volumes of different or the same gases under the same conditions of pressure and temperature contain the same number of molecules.

The law is obtained from the kinetic theory applied to two different gases at the same pressure and temperature. Let us apply Eqs. (27-9) and (27-16) to a unit volume of each of the gases.

$$p = \frac{1}{3}n_1m_1\overline{v_1^2},\quad p = \frac{1}{3}n_2m_2\overline{v_2^2}$$

and

$$\frac{1}{2}m_1\overline{v_1^2} = \frac{1}{2}m_2\overline{v_2^2}.$$

Therefore,

$$n_1 = n_2, \tag{27-19}$$

where n_1 and n_2 are the number of molecules in a unit volume of each of the gases.

Dalton's law of partial pressures is stated as follows:

> For a mixture of gases in a given volume at a given temperature, the pressure exerted by each gas is the same as if it alone occupied the volume; and the total pressure is equal to the sum of the partial pressures exerted by each gas.

This law can also be obtained by applying the kinetic theory to two different gases at the same temperature. Apply Eq. (27-9) to each of the gases confined separately in a unit volume.

$$p_1 = \frac{1}{3}n_1m_1\overline{v_1^2},\quad p_2 = \frac{1}{3}n_2m_2\overline{v_2^2}. \tag{27-20}$$

The pressure exerted by a mixture of the two gases in a unit volume is

$$p = \frac{1}{3}n_1m_1\overline{v_1^2} + \frac{1}{3}n_2m_2\overline{v_2^3}. \tag{27-21}$$

Therefore,

$$p = p_1 + p_2,$$

which is Dalton's law.

Example 3. A volume of $100\ \text{cm}^3$ of hydrogen is collected over water at a temperature of 20°C and a pressure of 75 cm of Hg. Find the volume of dry hydrogen at a temperature of 0°C and a pressure of 76 cm of Hg.

SOLUTION

Since the hydrogen is collected over water, it is saturated with water vapor. The pressure of saturated water vapor at 20°C, from Table 5-1, is 1.75 cm of Hg. The partial pressure of hydrogen (p_H), from Dalton's law, is

$$p_H = 75 - 1.75 = 73.25\ \text{cm of Hg}.$$

Applying the ideal gas law in the form of Eq. (3-17),

$$\frac{(76\ \text{cm Hg})(V)}{(273°\text{K})} = \frac{(73.25\ \text{cm Hg})(100\ \text{cm}^3)}{(293°\text{K})}.$$

$$V = 90\ \text{cm}^3.$$

27-5 REAL GASES IN CONTRAST TO IDEAL GASES

The assumptions made in the kinetic theory derivations of Section 27-2 are never quite realized in the case of real gases. Experiment shows that the equation of state for the ideal gas only holds for a real gas at low pressures. For example, Figure 27-2 is a plot of pV/nRT versus p in atm for hydrogen and oxygen. The horizontal line indicates the behavior expected for an ideal gas. The differences exhibited are not difficult to explain. As the pressure is increased, the volume is decreased, and the volume of the molecules becomes a more significant part of the total volume. The product pV is no longer constant, but increases with pressure for both gases. Also, the cohesive forces play a dominant role when the gas is at an intermediate pressure

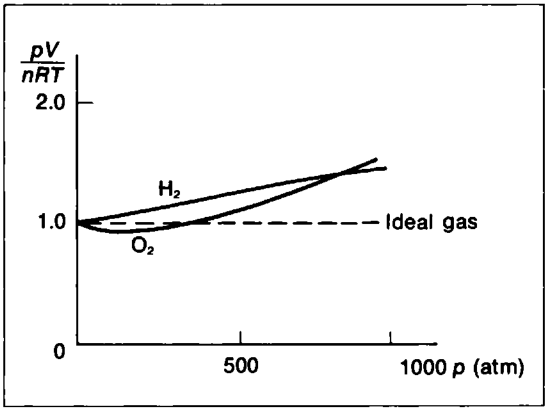


Figure 27-2 pV/nRT versus p for oxygen and hydrogen.

and the molecules are closer together. These forces account for the dip in the O_2 curve of Figure 27-2. At higher pressures, the repulsion of the molecules overcomes the cohesive forces. The lack of a dip in the hydrogen curve can be attributed to the small size of the hydrogen molecule.

At sufficiently high pressure and low temperature, the cohesive forces are able to prevent the elastic collisions of the molecules to a large extent, and we no longer have the kinetic properties of a gas. At this point, the gas condenses into a liquid.

27-6 LIQUIDS

In terms of kinetic theory, a liquid may be considered as a continuation of the gas phase into a region of small volumes and high molecular attractions. The molecular attractions must be high enough to keep the molecules confined to a definite volume. The molecules of a liquid have less freedom of motion than the molecules of a gas. This can be observed directly by *Brownian motion*, discovered by Robert Brown in 1827. Using a microscope, he observed small pollen grains floating in water, and found that they underwent random motion due to the bombardment of the water molecules. The same type of observations can be performed with small particles suspended in a gas, such as smoke particles in air. On comparison of the two observations, it can be concluded that the gas molecules have much greater freedom of motion than liquid molecules.

At temperatures near the critical temperature, it is possible to regard a liquid as essentially a gas whose molecules have a finite volume and attract one another. The equation of state here resembles van der Waal's equation [Eq. (5-1)]. The theory of liquids is in a much less satisfactory state than the theories of gases and solids, but important progress is being made in our understanding of the structure of liquids.

27-7 SOLIDS

The attractive forces between the atoms of solids are stronger than in liquids. In fact, the forces between the atoms or molecules are so strong that they are not free to move about as in a gas or liquid. They are held together as if by springs, and oscillate back and forth continuously. Thus, a solid is elastic because of the elastic behavior of the atoms making up the solid.

As noted in Section 5-1, most solids have a periodic arrangement of atoms, and are called *crystalline solids.* Most metals are crystalline solids. We may ask why there is a periodic arrangement of the atoms in a crystalline solid. The answer is that the potential energy of such a system is a minimum if the atoms occupy the smallest possible volume. It is observed that the most stable configuration of a system is the one with a minimum potential energy. A system of balls (or atoms) can occupy the smallest volume only by the formation of a periodic structure. A few solids, however, do not have a periodic structure, and thus do not have crystalline forms. These solids are called *amorphous* solids, two common examples of which are glass and rubber.

Since the atomic forces in a solid are very strong, the following question may be asked. What is the source of the forces that hold the atoms together in a solid? The answer is that the source is electromagnetic since each of the atoms is made up of a positively charged nucleus and negatively charged electrons. We may distinguish between three types of interatomic forces in solids.

(a) The Metallic-Bonded Solid

In this type of solid, the force may be simply expressed as electrostatic forces between the positive nucleus and the electron cloud, which not only surrounds the parent nucleus but also the nucleii of neighboring atoms. A material in which the atoms are held together by this type of bond has the high electrical and thermal conductivity characteristic of metals.

(b) The Covalent-Bonded Solid

In this type of solid, the neighboring atoms share valence electrons. It is a very strong bond. Covalent crystals have poor electrical and thermal conductivity. Diamond and carborundum are examples of such solids.

(c) The Ionic-Bonded Solid

The ionic bond is characteristic of an exchange of electrons between atoms and the formation of ions. It is a strong bond, and the solid has ionic conductivity at high temperatures. Salts such as sodium chloride and potassium chloride are examples of ionic-bonded solids or ionic crystals.

It should be pointed out that there must be strong repulsive forces as well as attractive forces in a solid, since large forces are required to compress a solid. Figure 27-3 is a graph of interatomic forces versus atomic separation. At a separation r_0, a stable state exists in which the resultant force is zero. The source of the repulsive force that balances the attractive force is not so easily explained. It is made up of several effects, the most important of which occurs when the outer filled electron shells of both atoms begin to overlap one another. When two completely filled shells overlap each other, the number of electrons in the single group formed by these shells is too large for them all to be contained in states of low energy, and many are forced to occupy higher energy levels. The increase of energy of these electrons is enough to give strong repulsive forces which oppose the attractive forces that tend to bring the ions together.

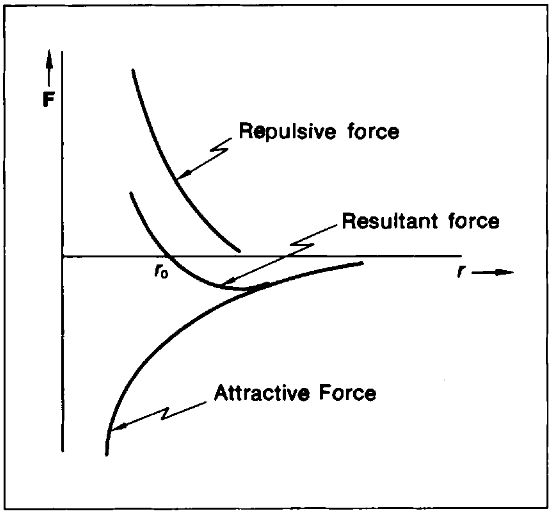


Figure 27-3 Interatomic forces versus atomic separation.

PROBLEMS

1. Are the following statements true or false?
   (a) The lighter molecules in the air move faster, on the average, than the heavy ones.
   (b) The pressure of air in a closed container might be doubled by heating the air, so as to make the molecules go twice as fast.
2. Consider 1 mole of an ideal gas at 0°C and 1 atmosphere pressure. Imagine each molecule to be, on the average, at the center of a small cube.
   (a) What is the length of an edge of the cube?
   (b) If the diameter of an average molecule is about 3×10^{-8} cm, what is the ratio of the length of an edge of the cube to the average diameter of the molecule?
3. Using the kinetic theory of matter, explain how the vapor forms in a closed space above a liquid, and how this vapor exerts a pressure on the enclosing walls. Why does this pressure have a certain fixed value at a given temperature?
4. (a) If we wish to double the root-mean-square velocity of a molecule, by what factor do we change the temperature?
   (b) If, at a given temperature, a molecule A has a mass double that of a molecule B, what is the ratio of their root-mean-square velocities?
5. (a) What is the translational kinetic energy of the molecules in 10 gms of oxygen gas at 27°C?
   (b) What is the root-mean-square velocity of a molecule of hydrogen at 273°C and 1 atmosphere pressure?

6. Calculate the mass, root-mean-square velocity, and kinetic energy of helium contained at 1 atmosphere and 20°C in a balloon of volume 10 m^3.
7. How much heat must be added to 3 kg of O_2 gas to raise its temperature from 0°C to 25°C at constant pressure? What is the increase in internal energy? How much external work is done? How much heat would need to be added in raising the temperature by the same amount at constant volume?
8. Two grams of oxygen and 6 gms of carbon dioxide are put into an evacuated 5 liter flask. Find the pressure inside the flask if the temperature is 27°C.
9. Show that

$$C_v = \frac{R/M}{\gamma - 1} \quad \text{and} \quad C_p = \frac{\gamma R/M}{\gamma - 1},$$

where $\gamma = C_p/C_v$.

10. The density of oxygen is 1.43×10^{-3} gm/cm^3 at 0°C and 1 atmosphere pressure. At what temperature would the root-mean-square velocity of oxygen equal the escape velocity from the Earth, which is approximately 11.2 km/sec.
11. The density of hydrogen is 9×10^{-5} gm/cm^3 at 0°C and 1 atmosphere pressure. Calculate the root-mean-square velocity of hydrogen at 1000°C. If temperatures approaching 1000°C are present in the upper atmosphere, and the escape velocity from the surface of the Earth is 11.2 km/sec, can there be much hydrogen in the Earth's atmosphere?
12. The velocity of a gas forced through a small hole in a wall of a container is proportional to the velocity of the molecules. Compare the time required to force out equal volumes of hydrogen and oxygen under the same pressure.

28 Sound

All sounds have one thing in common. Each is caused by the vibration of a material medium and is transmitted by waves. The speed of sound depends on the density, elasticity, and the temperature of the medium. Because of the above properties of sound, we have left a discussion of it until after we discussed wave motion, properties of matter, and temperature.

28-1 PRODUCTION AND SPEED OF SOUND

Sound is a vibratory motion that produces the sensation of hearing. Sound waves are longitudinal vibrations of material media. If the frequency of the vibrations is greater than that which can be detected by the ear, it is called an *ultrasonic* vibration, if less—an *infrasonic* vibration. The audible range for a human is from 20 to 20,000 cycles per second (cps).

In order for sound waves to be produced, there must be a source that initiates a mechanical disturbance and a material medium through which the disturbance can be transmitted. This material medium requirement is not difficult to demonstrate. Suppose we suspend an electric bell by means of a fine string in a container which can be evacuated by means of a pump. We start the bell ringing, and at the same time pump the air from the container. The sound will become fainter and fainter as the air is exhausted until the sound cannot be heard.

The speed of sound depends upon the elastic properties of the medium. In general, the speed of a wave in a material medium is given by

$$v = \sqrt{\frac{\text{elastic force factor}}{\text{inertia factor}}}.$$

The quantities used for both the elastic force factor and the inertia factor depend on the kind of wave motion and the medium. For sound waves in solids, the elastic force factor is the bulk modulus plus $\frac{4}{3}$ the shear modulus of the solid (see Chapter 23), while for fluids it is only the bulk modulus since the shear modulus is zero for a fluid. In both solids and liquids, the inertia factor is the density of the medium. For sound waves in a solid rod, the elastic force factor is the Young's modulus of the medium, which is also discussed in Chapter 23, and the inertia factor is the density of the medium.

In the case of gases, the adiabatic bulk modulus is used for the elastic force factor since they are highly compressible and the heat of compression cannot escape at ordinary frequencies. Therefore, the speed of sound in gases is expressed by the relation

$$v = \sqrt{\frac{B_{ad}}{\rho}}. \qquad (28\text{-}1)$$

However,

$$B_{ad} = -V\left(\frac{dp}{dV}\right)_{ad},$$

and the equation of state of an ideal gas for an adiabatic process is

$$pV^{\gamma} = K.$$

Taking logarithms of both sides of this equation, we get

$$\ln p + \gamma \ln V = \ln K,$$

and taking the differential of this equation gives

$$\frac{dp}{p} + \gamma \ln \frac{dV}{V} = 0$$

from which we obtain

$$\left(\frac{dp}{dV}\right)_{ad} = -\frac{\gamma p}{V}.$$

Therefore,

$$v = \sqrt{\frac{\gamma p}{\rho}}. \qquad (28\text{-}2)$$

Since the density of a gas varies inversely with the absolute temperature at constant pressure, the speed of sound in a gas varies directly with the absolute temperature. For an ideal gas,

$$\frac{p}{\rho} = \frac{RT}{M},$$

where R is the universal gas constant, M is the molecular weight, and T is the absolute temperature. Therefore,

$$v = \sqrt{\frac{\gamma RT}{M}}, \qquad (28\text{-}3)$$

which gives good agreement with the measured speed of sound in gases.

Example 1. Find the speed of sound in oxygen at 300°K.

SOLUTION

For oxygen, $\gamma = 1.40$ and $M = 0.032$ kg/mole. $R = 8.31$ joules/mole deg.

$$v = \sqrt{\frac{(1.40)(8.31 \text{ joules/mole deg})(300°\text{K})}{(0.032 \text{ kg/mole})}}$$

$$= 330 \text{ m/sec}.$$

The speed of sound also varies directly with the square root of the pressure, but since it also varies inversely with the square root of the density and since an increase in pressure increases the density, the effect of pressure depends on the state of the medium. The speed of sound in a gas is independent of changes in pressure within wide limits since the ratio of pressure to density is a constant. However, the effect of pressure is much greater in liquids and solids. In the case of liquids, it is an even more important factor than temperature.

Since sound is a wave motion, its wave length is related to the speed v, and is given by the relation

$$\lambda = \frac{v}{f}, \qquad (28\text{-}4)$$

where f is the frequency of the sound. Table 28-1 gives the speed of sound in air at various temperatures, and Table 28-2 gives the speed of sound in thin solid rods and in liquids.

Table 28-1 Measured Speed of Sound in Air.

Temperature (°C)	Speed (m/sec)
0	331.4
20	344.0
100	366.0
500	553.0
1000	700.0

Table 28-2 Measured Speed of Sound in Solid Rods and Liquids.

Solid rods	Speed (m/sec)	Liquids	Speed (m/sec)
Aluminum	5100	Alcohol	1143
Copper	3560	Ether	1032
Iron	1322	Turpentine	1326
Nickel	4973	Water	1461
Glass	5550	Sea water	1500

28-2 INTENSITY OF SOUND

A sound wave is a series of compressions and rarefactions. We therefore have a pressure amplitude as well as a displacement amplitude. The pressure amplitude (P) is the maximum amount by which pressure differs from atmospheric pressure. The intensity (I) of a sound wave is the energy transported per unit area per unit time perpendicular to the direction of propagation of the sound wave. A loud sound is one of high intensity, even though no simple mathematical relation exists between them. This is so because the loudness of a sound depends upon the particular sensitivity of the ear, while the intensity is a physical measurement.

It is frequently useful to compare sound intensities. Two sounds having an energy ratio of 1 to 10 are said to differ by 1 *bel*. Since the ear is sensitive to a very great range of intensities, a logarithmic scale is used which is implicit in the definition of the bel. It has been found that the decibel (db) is a more convenient unit, and the intensity level (β) measured in decibels is defined by the equation

$$\beta = 10 \log \frac{I}{I_0}, \qquad (28\text{-}5)$$

where I_0 is the intensity corresponding to the threshold of hearing or, in other words, the faintest sound that can be heard by the average human. The maximum that the human ear can tolerate is about 120 decibels. Referring to Eq. (28-5), this means that the ear has a hearing range from the lowest audible intensity to one which is 10^{12} times greater. Actually, this hearing range depends on the sound frequency. Figure 28-1 is an audibility diagram for a person with good hearing. Equation (28-5) is ordinarily used to compare the intensities of two sounds. From Eq. (28-5), it is readily seen that a sound 10 times as intense as another is 10 db higher, one 100 times as intense is 20 db higher, and one 1000 times as intense is 30 db higher.

Example 2. A sonar operator on board a ship pings on a submarine. The return signal is one-half the intensity of the ping. What is the attenuation of the ping by the water and the hull of the submarine in decibels?

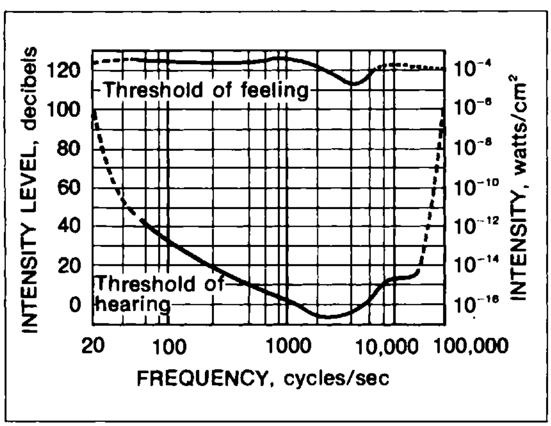


Figure 28-1 Audibility diagram for a person with good hearing.

SOLUTION

$$\beta = 10 \log \frac{(I \text{ watts/m}^2)}{(I_0 \text{ watts/m}^2)} = 10 \log 0.5 = -3 \text{ db}.$$

Thus, the return signal was 3 db below the ping in intensity. Therefore, the attenuation of the ping by the water and the hull was 3 db.

28-3 PSYCHOLOGICAL EFFECTS OF SOUND AND ORIGINS OF MUSICAL SOUNDS

A sound wave causes the eardrum (a fibrous membrane) to vibrate in synchronization with the sound vibrations. These vibrations of the eardrum are transmitted to the inner ear. The inner ear is filled with a fluid in which are located nerve endings that transmit the sound vibrations to the brain.

The quality of a sound refers to the psychological effect it produces for a listener. A musical note usually consists of a number of vibrations of different frequencies, consisting of the lowest frequency, called the *fundamental*, and other vibrations having frequencies of two, three, or more times that of the fundamental, called *harmonics*. By contrast, a noise is a mixture of many unrelated vibrations. It is the number of harmonics present and their respective intensities that determine the quality of a sound. Two notes with the same fundamental frequency and intensity will differ in quality when sounded on a guitar and a violin because of the spectrum of harmonics produced. The pitch of a musical note refers to the dominant frequency. When several musical notes produce a pleasing sensation to the listener, they are said to be in harmony. If the effect produced is disturbing, the notes are said to be discordant. The explanation is roughly as follows: the listener is aware of not only the dominant frequencies of the notes but also the difference frequencies or beat tones. Low frequency, unrelated beat tones produce a discordant sensation. Higher frequency beat tones, in cases where the same beat frequency results from more than one pair of dominant frequencies in the set of musical notes, result in a sound that has a pleasing fullness or richness. Musical instruments, whether classified as strings (piano, violin), winds (clarinet, flute), or percussion (xylophone, glockenspiel) are all designed to minimize discordant effects.

The means of initiating vibrations is different for the various classes of instruments. This is in part necessary because the medium in which the vibrations originate differs. Thus, transverse standing waves are excited in stretched strings in string instruments. For wind instruments, the mode of vibration is standing longitudinal waves in a column of air. Percussion instruments involve transverse standing waves in solid plates or stretched membranes. In Section 10-3, we discussed standing transverse waves of a stretched string. Equations (10-22) and (10-23) provide the relations between the wavelength, frequency, and length of the string for all harmonics, including the fundamental.

Another reason for the various methods of initiating vibrations is to provide musical notes of higher quality. For example, a violinist enhances the quality of notes by using his fingers to create points of zero amplitude along the vibrating string. This prevents the existence of any harmonics that would have non-zero amplitudes at those points. The same result is accomplished in a flute by covering various holes along the length of the instrument.

The reader should not infer that the physics of musical sounds has been fully treated in this section. The discussion here is in fact highly idealized, and no consideration has been given to such questions as musical scales and the human voice. However, the ideas presented do represent a satisfactory basis for further study.†

28-4 THE DOPPLER EFFECT

When a listener and/or source of sound are moving relative to each other, the frequency of the sound as heard by the listener is not the same as the true frequency of the source. We shall illustrate this by discussing two specific cases.

(a) Source Stationary and Listener Moving

Suppose a stationary source emits a note of frequency f, and the listener approaches the source with a velocity v_L, where v_L is less than the velocity of the sound v. This is illustrated in Figure 28-2, where we take the initial distance between the listener L and the source S equal to the distance the sound travels in 1 second or, in other words, numerically equal to the velocity of the sound.

†The interested reader is referred to the text, *Horns, Strings and Harmony*, by Arthur H. Benade, Doubleday & Co., 1960.

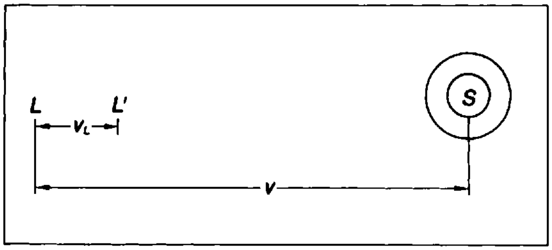


Figure 28-2 Listener moving toward source.

If the listener was stationary at L, f waves would hit his ear per second. However, during 1 second he travels a distance v_L to L'. He therefore hears an extra vibration for each wavelength in the path length LL'. The number of extra vibrations per second he hears is equal to LL'/λ or v_L/λ. Therefore, the frequency he hears (f') is given by

$$f' = f + \frac{v_L}{\lambda} = f + \frac{fv_L}{v},$$

or

$$f' = f\left(\frac{v + v_L}{v}\right). \qquad (28\text{-}6)$$

The wavelength of the sound is unchanged, but the listener goes to meet the oncoming waves and, consequently, more waves hit his ear per second than the source emits in 1 second.

If the listener had been going away from the source, he would have heard v_L/λ fewer vibrations per second. Therefore, the frequency he would have heard is given by

$$f' = f - \frac{v_L}{\lambda} = f - \frac{fv_L}{v}$$

or

$$f' = f\left(\frac{v - v_L}{v}\right). \qquad (28\text{-}7)$$

(b) Listener Stationary and Source Moving

Suppose the source S is producing a note of frequency f, and is approaching the listener L with a velocity $v_S < v$. As before, we shall let S be a distance v from L, as illustrated in Figure 28-3 where v is equivalent to the distance traveled in 1 second by the sound. The first of the f vibrations is given out by the source when it is at S, and the last of the f vibrations—when it has advanced to S', a distance v_S from S. Since $v_S < v$, there will be f vibrations between S' and L. Hence, the new wavelength λ' in front of the source will be less

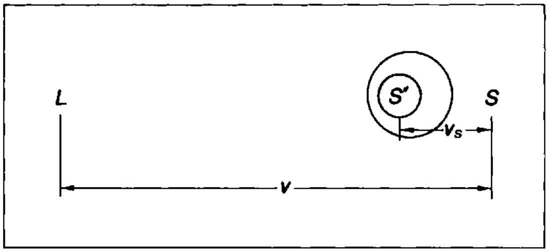


Figure 28-3 Source moving toward listener.

than λ for the source at rest. The new wavelength is

$$\lambda' = \frac{S'L}{f}.$$

But

$$f'\lambda' = v,$$

where f' is the frequency the listener hears. Therefore,

$$\lambda' = \frac{v}{f'} = \frac{S'L}{f} = \frac{v - v_S}{f}$$

or

$$f' = f\left(\frac{v}{v - v_S}\right). \tag{28-8}$$

If the source had been going away from the listener, by the same argument as presented above, the frequency the listener would have heard is given by

$$f' = f\left(\frac{v}{v + v_S}\right). \tag{28-9}$$

With the source and listener both in motion, we can combine Eqs. (28-6), (28-7), (28-8), and (28-9), and obtain

$$f' = f\left(\frac{v \pm v_L}{v \pm v_S}\right). \tag{28-10}$$

The signs adopted are such as to increase the ratio f'/f when the source and listener are approaching one another, and to decrease the ratio f'/f when they are receding from one another. For example, to a listener, the frequency of a locomotive whistle will appear to be increased when there is relative motion toward the listener, and decreased when there is relative motion away from the listener. If the medium is moving in the same direction as the sound wave, its velocity is added to the velocity of the sound, and subtracted if moving in the opposite direction. This alteration of frequency due to the motion of the listener and/or source is known as the Doppler effect.

We have derived the above equations for the case where the source and listener are moving along a line joining them. In general, it can be shown that the equations hold when v_S and v_L are the components of the velocities along the line joining the source and the listener. This is also true for the velocity of the medium.

When the velocity of the source is greater than the velocity of the sound wave, no regular wave train results. Instead, a bow and stern wave is set up similar to that produced by a boat moving over water. Another example is the shock wave associated with a supersonic boom.

Example 3. A bat flies straight toward a wall at 30 m/sec on a day when the speed of sound is 330 m/sec, while emitting a note of 15,000 cps. What frequency does it hear?

SOLUTION

The bat is both the source of the sound and the listener.

$$f' = f\left(\frac{v + v_L}{v - v_S}\right) = 15{,}000 \text{ cps}\left[\frac{330 + 30) \text{ m/sec}}{(330 - 30) \text{ m/sec}}\right]$$
$$= 18{,}000 \text{ cps}.$$

Example 4. A locomotive goes past a station at 20 m/sec, blowing its whistle at a frequency of 1000 cps on a day when the velocity of sound is 330 m/sec. What change in frequency is heard by a boy standing on the station platform?

SOLUTION

The frequency heard by the boy as the train approaches is

$$f' = 1000 \text{ cps}\left(\frac{330 \text{ m/sec}}{(330 - 20) \text{ m/sec}}\right) = 1065 \text{ cps}.$$

The frequency heard by the boy as the train recedes is

$$f' = 1000 \text{ cps}\left(\frac{330 \text{ m/sec}}{(330 + 20) \text{ m/sec}}\right) = 943 \text{ cps}.$$

$$\text{Change in frequency} = (1065 - 943) \text{ cps}$$
$$= 122 \text{ cps}.$$

PROBLEMS

1. The speed of sound in steel is 5000 m/sec and the density of steel is 7.8×10^3 kg/m^3. What is the elastic force factor of steel?

2. What is the adiabatic bulk modulus of a gas if the speed of sound in the gas is 300 m/sec and the density is 2 kg/m^3?
3. A sound of frequency 1000 cycles/sec is radiated uniformly in all directions with an intensity of 16 $watts/m^2$ at a radius of 5 m on a day when the speed of sound is 300 m/sec.
   (a) What is the total acoustic power radiated by the source?
   (b) What is the apparent frequency of the sound to an observer approaching the source at a rate of 20 m/sec?
4. A cowardly man, running away from a dentist at a speed of 10 m/sec, emits a shriek of terror having a frequency of 5000 cycles/sec. What is the frequency heard by the dentist who is pursuing the man at 4 m/sec, if the temperature is 20°C?
5. In Problem 3, if the intensity of the sound 50 m away is 16×10^{-2} $watts/m^2$, what is the intensity level in decibels at a radius of 5 m compared to the intensity at a radius of 50 m?
6. A locomotive, traveling with a speed of 25 m/sec, sounds its whistle as it approaches an observer at a grade crossing when the temperature is 20°C. The frequency of the whistle is 250 cycles/sec. What is the wavelength of the sound waves in the region between the locomotive and the observer?
7. If the sound source in Problem 3 is mounted on a vehicle and is carried at a uniform speed of 50 m/sec past a stationary observer, what is the difference between the apparent frequency as it approaches and the apparent frequency as it goes away?
8. An automobile moving at 30 m/sec is approaching a factory whistle having a frequency of 500 cycles/sec. If the speed of sound in air is 340 m/sec, what is the apparent frequency of the whistle as heard by the driver?
9. A ship, dead in the water, emits a sound pulse from its sonar gear, and 2 seconds later receives an echo identifiable as a submarine. A short while later, the hydrophone picks up the sound of a torpedo being fired, and the alarm is immediately given. Assuming the torpedo travels with a constant speed of 50 m/sec, how much time does the crew have to abandon ship from the time the alarm was sounded to the time the torpedo strikes? Take the speed of sound in sea water as 1500 m/sec and neglect the motion of the submarine.
10. A man runs at 5 m/sec on a calm day when the temperature is 20°C between two fire stations. The fire whistles sound simultaneously in the two stations at a frequency of 600 cycles/sec. Find the apparent frequency of each whistle.
11. What sound intensity is 3 db louder than a sound of 0.10 $watts/m^2$?
12. A locomotive travels 30 m/sec in still air, blowing its whistle. The frequency of the whistle is 600 cycles/sec and the speed of sound in the air is 340 m/sec.
    (a) What frequency would be heard by a passenger on a second train traveling at 20 m/sec away from the first?
    (b) What frequency would be heard by the passenger if a wind of velocity 10 m/sec is blowing in the direction that the locomotives are traveling?
13. A train passes by a station, blowing a whistle, on a day when the speed of sound is 330 m/sec. A stationary observer on the platform notices that the frequency of the note he hears changes in the ratio 9/8 as the engine passes him. Find the speed of the train.
14. How rapidly must an automobile be moving away from an observer so that the frequency of its horn will drop from 1100 to 900 cps on a day when the temperature is 20°C?

29 Charge, Field, and Potential

29-1 INTRODUCTION

The earliest observations of electrical phenomena were made by the Greeks, but no careful experiments were carried out until the seventeenth century. The original observations indicated that certain substances, when rubbed with fur or cloth, attracted small objects in the neighborhood. Such substances are said to be electrically charged.

Following the observations of the force exerted by a charged body on all uncharged bodies, the next step was the study of the forces exerted by different charged bodies on each other. It was found, for example, that if two pieces of glass were each rubbed with a separate piece of silk so that all four acquired a charge, the two pieces of glass repelled one another, but each piece of glass attracted each piece of silk. Similarly, glass that had been rubbed with silk attracted hard rubber that had been rubbed with fur, but repelled the fur. This led to the realization that there must be two kinds of electric charge.

To distinguish between the two kinds of charge, that produced on glass by rubbing with silk was called *positive* (+), and that produced on hard rubber by rubbing with fur was called *negative* (−). Then, by application of the rule that *unlike charges attract and like charges repel*, it was possible to find the sign of the charge on any body.

It was established early in this century that all electrical charges are integer multiples of a fundamental or smallest unit of charge. The electron is a negatively charged particle, and has a charge magnitude that corresponds to this smallest unit of charge. Matter is made up of atoms consisting of a central core or nucleus surrounded by lighter particles called electrons. The nucleus contains two kinds of particles: protons and neutrons, which are roughly 1840 times as heavy as the electrons. The protons are positively charged, while the neutrons are uncharged. The number of electrons surrounding the nucleus is equal to the number of protons, and each electron has a negative charge equal in magnitude to the positive charge of the proton, so that the atom as a whole is electrically neutral. The number of neutrons is roughly equal to the number of protons in light elements, but is greater in the heaviest elements. A given element may have several stable forms of different nuclear mass, that is, it may have a different number of neutrons but the same number of protons. This means that the element can exist in forms with different mass but with the same charge. Such forms of an element are called isotopes.

For the purpose of electrostatic theory, substances can be divided into two classes: conductors and insulators. In a conductor, the negative electrical charge (electrons) can flow easily throughout the substance under the influence of an electric field, while in an insulator, a large electric field is required to cause the electrons to flow.

29-2 ELECTRICAL FORCES

Electrical forces, like gravitational forces, vary inversely as the square of the distance, but they are about 10^{36} times stronger. If you were standing at arm's length from someone, and each of you had about 1% more electrons than protons, or vice versa, the repulsive force would be enough to lift a weight equal to that of the Earth. Since you feel no force at all at this distance from someone else, this must mean there is essentially perfect balance between the protons and electrons of each person.

We have said that the electrical force varies inversely as the square of the distance. This means that if we have two point electrical charges, the force of attraction or repulsion of one charge on the other is inversely proportional to the square of the distance between them. The force is also directly proportional to the size or magnitude of these charges.

This law was discovered by Coulomb in 1785. In his apparatus, the charges were carried on pith balls and the force between them was measured with a torsion balance as shown in Figure 29-1. Two small spheres A and B are attached to the ends of a rod,

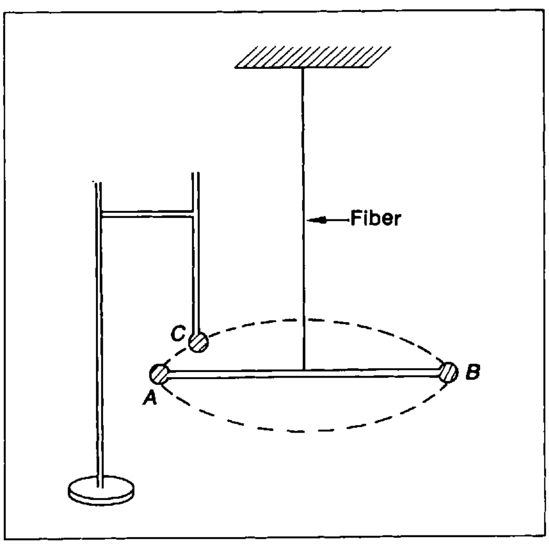


Figure 29-1 Torsion balance to demonstrate Coulomb's law.

which is suspended by a torsion fiber attached to the center of mass of the rod. If sphere A is charged and a sphere C charged with a like charge is placed next to it, they will repel one another and the fiber will be twisted through a measurable angle. The force of repulsion is determined by the amount of twist put in the fiber when the charged spheres are at a given distance. Coulomb's law states that:

> The force between two point charges is directly proportional to the product of the charges and inversely proportional to the square of the distance between them.

$$F = K\frac{qq_1}{r^2}, \tag{29-1}$$

where K is a constant of proportionality.

The vector equation for the force is

$$\mathbf{F} = K\frac{qq_1}{r^2}\hat{r}_0, \tag{29-2}$$

where $\hat{r}_0$ is a unit vector pointing along the line joining the charges.

If other charges are in the vicinity of q, each will exert a force on q, and the total electrostatic force on q will be

$$\mathbf{F} = Kq\sum_{i=1}^{n}\frac{q_i}{r_i^2}\hat{r}_{0i}. \tag{29-3}$$

The $\hat{r}_{0i}$'s all have unit magnitude, but they may point in different directions. The mutual interaction of a pair of charges is unaffected by other charges. The algebraic sum of the charges within a closed system is constant.

$$\left(\sum_{i=1}^{n} q_i = \text{constant}\right).$$

However, Coulomb's law is not perfectly true when the charges are moving. The electrical forces also depend on the motion of the charges.

29-3 UNITS OF CHARGE

In the MKS units, the force is expressed in newtons, the distance in meters, and q in coulombs. The constant K is not defined directly but through another constant ϵ_0 by the equation

$$K = \frac{1}{4\pi\epsilon_0}.$$

The new constant, called the electric permittivity of free space, is defined to be

$$\epsilon_0 = \frac{10^7}{4\pi c^2}\frac{\text{coul}^2}{\text{nt}-\text{m}^2},$$

where c is the velocity of light in free space; $c = 3 \times 10^8$ m/sec. Now we see that in free space

$$\epsilon_0 = 8.85 \times 10^{-12}\frac{\text{coul}^2}{\text{nt}-\text{m}^2}$$

and

$$K = 9 \times 10^9 \frac{\text{nt-m}^2}{\text{coul}^2}.$$

In MKS units, Coulomb's law is

$$\mathbf{F} = K\frac{qq_1}{r^2}\hat{r}_0 = \frac{1}{4\pi\epsilon_0}\frac{qq_1}{r^2}\hat{r}_0 = 9 \times 10^9 \frac{qq_1}{r^2}\hat{r}_0, \qquad (29\text{-}4)$$

where
$\mathbf{F}$ = force (nt)
q = charge (coul)
q_1 = charge 1 (coul)
r = distance between point charges (m)
$\hat{r}_0$ = unit vector pointing in direction of line joining the charges.

The total force exerted on a given charge q by n other charges is given by Eq. (29-3). In a normal laboratory situation, the number of electrons and/or protons distributed in and around the laboratory that can affect a given test charge q is usually very large. As a result, the calculation required in the summation in Eq. (29-3) is prohibitively difficult. This difficulty is easily by-passed, however. One simply "maps" the electrical influences in the laboratory by carrying the charge q throughout the volume of the laboratory and measuring the magnitude and direction of the total force experienced by q at each point. If q is replaced by a charge q', the force it would experience is simply (q'/q) times the force measured for q, and a second measurement is unnecessary.

Any physical quantity that takes on different values at different points in space is said to define a field. If the values change in magnitude only, the field is a scalar field. If the values change in magnitude and direction, the field is a vector field.

29-4 FIELDS

A field is any physical quantity that takes on different values at different points in space. There are two kinds of fields: namely, scalar fields and vector fields. Temperature is an example of a scalar field, while the velocity of a flowing fluid is an example of a vector field. Also electric and magnetic fields are vector fields.

A vector field may be represented by drawing a set of arrows whose magnitudes and directions indicate the values of the vector field at the points from which the arrows are drawn (Figure 29-2).

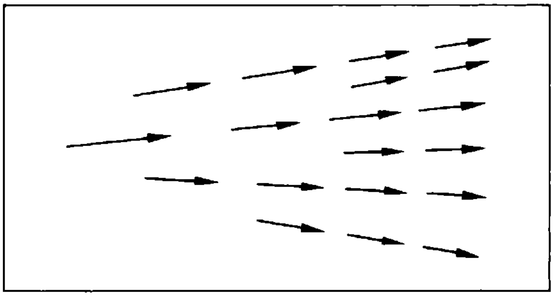


Figure 29-2 Representation of an electric field by arrows.

We can go further and draw lines that are tangent to the vectors everywhere, with the density of the lines proportional to the magnitude of the vector field (Figure 29-3).

The lines are sometimes called lines of force. In electrostatics, they originate on positive charges and end on negative charges. Lines of force do not intersect one another since the direction of the field cannot have two values at one point.

We adopt the convention that the number of lines per unit area over a surface perpendicular to the direction of the lines is proportional to the field strength.

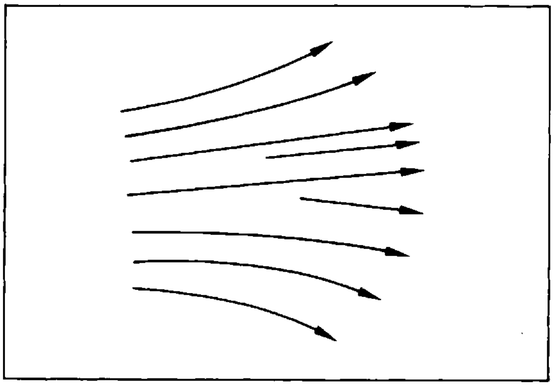


Figure 29-3 Representation of an electric field by lines.

An electric field is a vector point function, that is, a vector function of position in space in terms of which the force on a charge at rest at that point in space may be determined. An electric field exists at a point in space if a force is exerted on an electric charge placed at that point.

The electric field intensity or the electric field strength at a point is a vector having the direction of the force that would be exerted on a positive point test charge placed at the point, and a magnitude equal to the magnitude of this force divided by the magnitude of the test charge,

$$\mathbf{E} = \frac{\mathbf{F}}{q}. \tag{29-5}$$

If q is positive, $\mathbf{E}$ is in the direction of $\mathbf{F}$ and equal in magnitude and direction to $\mathbf{F}/q$. If q is negative, $\mathbf{E}$ and $\mathbf{F}/q$ are opposite in direction to $\mathbf{F}$. The unit of $\mathbf{E}$ in MKS units is newtons/coulomb.

If the charges causing the field are free to move (for example, metallic bodies), the presence of the test charge q may change their positions, and hence change the field we wish to measure. The change will be small if q is small, so an accurate definition of $\mathbf{E}$ in the form of an equation is

$$E = \lim_{q\to 0} \frac{\mathbf{F}}{q} \quad \text{or} \quad \mathbf{E} = \frac{d\mathbf{F}}{dq}. \tag{29-6}$$

Example 1. Find the field strength in magnitude and direction at a point on the bisector and 6 m from the line joining the charges of 5×10^{-9} coulombs each for both unlike and like charges separated by a distance of 16 m.

SOLUTION

Put a test charge q at point P in Figure 29-4.

$$E_1 = E_2 = \left(9 \times 10^9 \frac{\text{nt-m}^2}{\text{coul}^2}\right)\left[\frac{5 \times 10^{-9}\ \text{coul})q}{(10^2\ \text{m}^2)q}\right]$$
$$= 0.45\ \text{nt/coul}$$

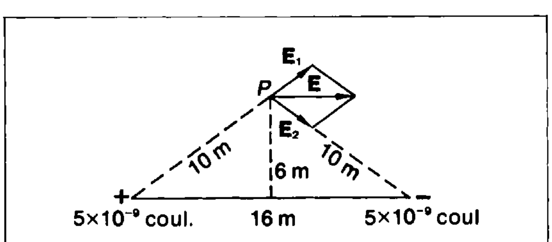


Figure 29-4

$$E = (0.45\ \text{nt/coul})\frac{8}{10} + (0.45\ \text{nt/coul})\frac{8}{10}$$
$$= 0.72\ \text{nt/coul}.$$

Therefore, $\mathbf{E} = 0.72$ nt/coul parallel to line of charges.

Put a test charge q at point P in Figure 29-5.

$$E_1 = E_2 = \left(9 \times 10^9 \frac{\text{nt-m}^2}{\text{coul}^2}\right)\left[\frac{(5 \times 10^{-9}\ \text{coul})q}{(10^2\ \text{m}^2)q}\right]$$
$$= 0.45\ \text{nt/coul}$$
$$E = (0.45\ \text{nt/coul})\frac{6}{10} + (0.45\ \text{nt/coul})\frac{6}{10}$$
$$= 0.54\ \text{nt/coul}.$$

Therefore, $\mathbf{E} = 0.54$ nt/coul along bisector.

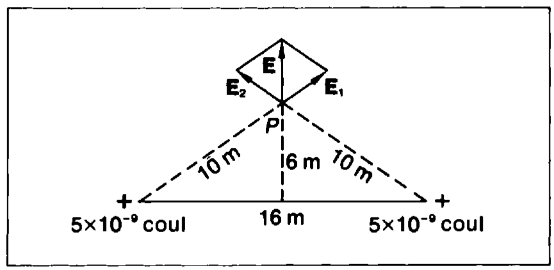


Figure 29-5

29-5 ELECTRIC POTENTIAL

When a charge moves in an electric field, work is done on the charge by the electric forces. There is a change in the electric potential energy of the charge. Just as it was convenient to define $\mathbf{E}$ as the force per unit charge, it is similarly convenient to define the electrical potential V as the electrical potential energy per unit charge.

Electric potential is best understood by considering first the difference of potential between two points in an electric field.

The difference of potential between two points in an electric field is defined as the work per unit charge necessary to carry a test charge from one point to another.

The work done by applying an external force, $\mathbf{F}_A$, in moving a charge q an infinitesimal distance ds is

$$\mathbf{F}_A \cdot d\mathbf{s} = dW = -\mathbf{F} \cdot d\mathbf{s},$$

where $\mathbf{F}$ is the force of the field.

The negative sign is inserted because the component of the motion anti-parallel to $\mathbf{E}$ is opposed by the field, or work is done on the charge by the external force. Since

$$\mathbf{F} = q\mathbf{E}$$

$$dW = -q\mathbf{E}\cdot d\mathbf{s}$$

or

$$\frac{dW}{q} = -\mathbf{E}\cdot d\mathbf{s}.$$

Hence

$$dV = -\mathbf{E}\cdot d\mathbf{s}$$

or

$$dV = -E\cos\theta\, ds, \tag{29-7}$$

where dV is the infinitesimal increase in potential, and θ is the angle between the positive directions of the vector **E** and the vector $d\mathbf{s}$.

The difference of potential between a point A and a point B is given by the equation

$$V_B - V_A = -\int_A^B \mathbf{E}\cdot d\mathbf{s}, \tag{29-8}$$

and is independent of the path because the potential energy difference is independent of the path.

The unit of potential in MKS units is the volt that is equal to a joule per coulomb. The electric field intensity can now be expressed in volts per meter, which is equivalent to newtons per coulomb.

We shall now consider the potential at a point. Potential is usually defined as being zero at a point infinitely distant from the charges producing the field. Therefore, the potential at a point s is the work that must be expended in carrying a test charge from infinity to the point s.

$$W = -\int_\infty^s \mathbf{F}\cdot d\mathbf{s} = -q\int_\infty^s \mathbf{E}\cdot d\mathbf{s}$$

or

$$V(\mathbf{s}) = -\int_\infty^s \mathbf{E}\cdot d\mathbf{s}. \tag{29-9}$$

As an example, let us consider the potential in the vicinity of a point charge. Since the field from the charge is radial, we can write,

$$\mathbf{E}\cdot d\mathbf{s} = \mathbf{E}\cdot d\mathbf{r} = E\,dr, \quad \text{and} \quad V(r) = -\int_\infty^r E\,dr.$$

if we take the charge as the origin of the coordinate system. But

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

Therefore,

$$V(\mathbf{r}) = -\frac{1}{4\pi\epsilon_0} q\int_\infty^r \frac{dr}{r^2}$$

or

$$V(\mathbf{r}) = \frac{1}{4\pi\epsilon_0}\frac{q}{r}. \tag{29-10}$$

The potential due to several charges is given by

$$V(\mathbf{r}) = \frac{1}{4\pi\epsilon_0}\sum_{i=1}^{n}\frac{q_i}{r_i}, \tag{29-11}$$

where $\mathbf{r}_i$ is the position vector of the point **r** with respect to the ith charge. If the potential is due to a continuous distribution of charge,

$$V(\mathbf{r}) = \frac{1}{4\pi\epsilon_0}\int\frac{dq}{r}. \tag{29-12}$$

29-6 CALCULATION OF E FROM V

The electric field **E** can be obtained easily from the potential V. We have established that

$$\frac{dW}{q} = dV = -\mathbf{E}\cdot d\mathbf{s}.$$

In rectangular coordinates,

$$\mathbf{E} = \mathbf{i}E_x + \mathbf{j}E_y + \mathbf{k}E_z$$

and

$$d\mathbf{s} = \mathbf{i}dx + \mathbf{j}dy + \mathbf{k}dz.$$

Therefore,

$$dV = -(E_x dx + E_y dy + E_z dz)$$

also

$$dV = \frac{\partial V}{\partial x}dx + \frac{\partial V}{\partial y}dy + \frac{\partial V}{\partial z}dz.$$

Therefore

$$E_x = -\frac{\partial V}{\partial x},\quad E_y = -\frac{\partial V}{\partial y},\quad E_z = -\frac{\partial V}{\partial z}.$$

Hence

$$\mathbf{E} = -\mathbf{i}\frac{\partial V}{\partial x} - \mathbf{j}\frac{\partial V}{\partial y} - \mathbf{k}\frac{\partial V}{\partial z} = -\nabla V \tag{29-13}$$

where

$$\nabla = \mathbf{i}\frac{\partial}{\partial x} + \mathbf{j}\frac{\partial}{\partial y} + \mathbf{k}\frac{\partial}{\partial z}$$

and is called the gradient operator (see Appendix I-8, p. 332).

Thus, if V is known for all points in space, that is, if the function $V(x, y, z)$ is known, the components of **E** and hence **E** itself can be obtained.

Example 2. The potential V of a particular electric field **E** is given by $V = a(2x^2 - 3y^2 - z^2)$, where a is a constant. Find the components of **E**.

SOLUTION

$$\mathbf{E} = a\left(-\mathbf{i}\frac{\partial V}{\partial x} - \mathbf{j}\frac{\partial V}{\partial y} - \mathbf{k}\frac{\partial V}{\partial z}\right) = a\left[-\mathbf{i}\frac{\partial}{\partial x}(2x^2) + \mathbf{j}\frac{\partial}{\partial y}(3y^2) + \mathbf{k}\frac{\partial}{\partial z}(z^2)\right] = -4ax\mathbf{i} + 6ay\mathbf{j} + 2az\mathbf{k}.$$

Therefore,

$$E_x = -4ax, \quad E_y = 6ay \quad \text{and} \quad E_z = 2az.$$

PROBLEMS

1. Two point charges are placed as $A = -30 \times 10^{-10}$ coul at $(-9\text{ m}, 0)$ and $B = +40 \times 10^{-10}$ coul at $(+16\text{ m}, 0)$.
   (a) What is the magnitude and direction of the force which charge A exerts on charge B?
   (b) What is the magnitude of the electric field intensity at the point $P(0, 12\text{ m})$?
   (c) What is the electric potential at the point P?
2. Two point charges are placed as $A = +30 \times 10^{-12}$ coul at $(-9\text{ m}, 0)$ and $B = -40$ coul at $(+16\text{ m}, 0)$.
   (a) What is the electric field intensity at the point $P(0, 12\text{ m})$? Calculate the magnitude, and show the direction in a diagram roughly to scale.
   (b) What is the electric potential at the point P?
3. A negative point charge of 20 microcoul is located at the point $A(x = 0,\ y = 30\text{ cm})$, and an equal positive charge is located at the origin.
   (a) What is the electric potential at point $B(x = 40\text{ cm},\ y = 0)$?
   (b) What is the electric field intensity at point B?
   (c) What is the electric potential energy of a positive point charge of 5 microcoul placed at the point B?
4. An electron ($q_e = 1.60 \times 10^{-19}$ coul, $m = 9.11 \times 10^{-31}$ kg) is in the electric field between two charged plates.
   (a) What field is required so that the electron has no vertical motion?
   (b) If the field is doubled, determine the time it takes for the electron to move 5 cm vertically upward.
   (c) If the direction of the field is changed by 180° in part (b) determine the time for the electron to move 5 cm vertically downward.
5. An electron (mass $= 9.11 \times 10^{-31}$ kg; charge $= 1.60 \times 10^{-19}$ coul) is projected horizontally with a velocity of 10^6 m/sec between two parallel plates at right angles to a uniform electric field of intensity 2×10^3 nt/coul, upward, as in Figure 29-6.
   (a) Indicate in the diagram, by pluses or minuses, the charge distribution on the plates.
   (b) Calculate the voltage between the plates required to produce this field.
   (c) What is the net force on the electron in the field region?
   (d) What is the acceleration of the electron?

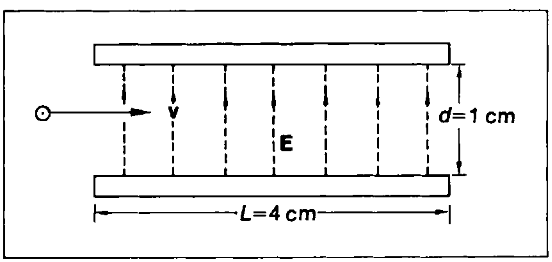


Figure 29-6

6. Two small 1 gm pith balls are attached to fine threads 1 m long hung from a common point. When the balls are given equal quantities of positive charge, they repel one another so that the threads are at an angle of 30° with one another.
   (a) Draw a diagram showing all the forces on each ball.
   (b) Find the charge on each ball.

7. An electron (mass = 9.11×10^{-31} kg; charge = 1.60×10^{-19} coul) is projected horizontally with a velocity of 10^6 m/sec between two parallel plates at right angles to a uniform electric field of intensity 5×10^3 nt/coul, upward, as in Figure 29-7.
   (a) Indicate on the diagram, by pluses or minuses, the charge distribution on the plates.
   (b) Calculate the voltage between the plates required to produce this field.
   (c) What is the net force on the electron in the field region?
   (d) What is the acceleration of the electron?

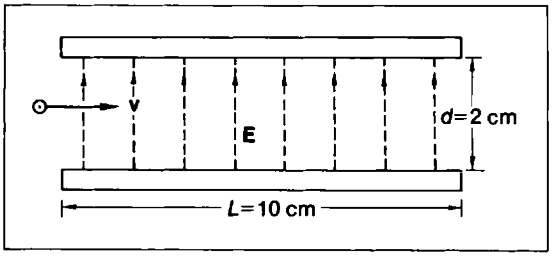


Figure 29-7

8. (a) What is the electric field intensity halfway between two isolated positive point charges of 4×10^{-10} coul, 60 cm apart?
   (b) What is the electric potential (relative to a point at infinity) at the point halfway between the two isolated charges of part (a)?
9. An electron (mass = 9.11×10^{-31} kg, charge = 1.60×10^{-19} coul) released from rest in an electric field of 3×10^3 volts/m travels 0.2 cm.
   (a) What is its final speed?
   (b) How long a time elapses before it reaches its final speed?
   (c) What is its final energy?
10. Charges A, B, C, and D are located at the corners of a square 20 cm on a side as illustrated in Figure 29-8. Charge $A = +15 \times 10^{-9}$ coul, $B = +15 \times 10^{-9}$ coul, $C = -15 \times 10^{-9}$ coul, and $D = -15 \times 10^{-9}$ coul. At the intersection of the diagonals (E) find:
   (a) The magnitude and direction of the electric field intensity and
   (b) the potential.

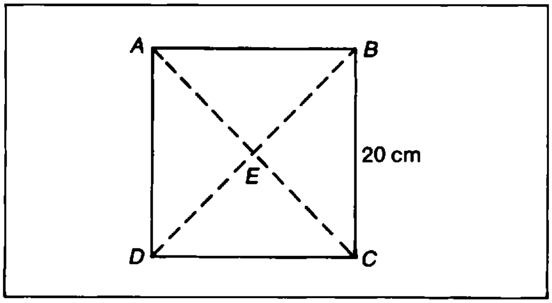


Figure 29-8

11. Two electric charges of opposite sign, but of the same magnitude, 4 microcoul are separated by a distance of 6 m in a vacuum. What is the electric field intensity and potential at a point midway between them on a straight line joining them?
12. An electron (mass = 9.11×10^{-31} kg, charge = 1.60×10^{-19} coul) revolves about a proton which has the same charge as the electron in an atom. The radius of the circular orbit is 10^{-10} m.
   (a) What is the radial acceleration of the electron?
   (b) What is the angular velocity of the electron?
13. If the electric potential along the x-axis is given by $V = (8 + 4x - 2x^2)$ volts, where x is in meters, what is the electric field along the x-axis at $x = 0$ and at $x = 2$ m?
14. A vacuum tube consists of a cylindrical cathode of 0.10 cm in radius inside a hollow cylindrical anode of inside radius 0.60 cm. The potential of the anode is 400 volts above that of the cathode. An electron leaves the surface of the cathode with zero initial velocity. Find its velocity when it strikes the anode.

15. In a certain electric field, the potential is found to be given by the formula $V = cd$, where c is a positive constant and d is a coordinate.
    (a) Find the value of **E** in the direction of increasing d.
    (b) What formula would give the work done in moving a charge from $+d_1$ to $+d_2$ (d_1 greater than d_2)?
    (c) Give the units of the constant c.
16. A positively charged ring of radius R lies in the y-z plane. Show that the formula for the electric field intensity at a point on the axis of the ring (x-axis) a distance r from the center of the ring is

$$E = \frac{Kqr}{R^2 + r^2},$$

where q is the total charge on the ring.

30 Gauss's Law, Capacitance, and Dielectrics

30-1 GAUSS'S LAW

In Chapter 29, we replaced calculations of forces on and potential energy of charges with calculations of field intensity and potential. This made it possible to avoid the serious complications related to the large numbers of charges involved in a typical situation. However, it will not have escaped the perceptive reader that the simplifications are more apparent than real for large numbers of charges. For example, it is no easier to calculate the total field intensity due to many charges than it is to calculate the total force exerted on a test charge due to the same charges. What is needed is some further concept that will make possible calculations such as the determination of the field intensity at all points in space due to an array of charges uniformly distributed on the surface of a hollow sphere.

Such a distribution has a degree of symmetry that affords the needed additional simplification. That is, because of the spherical symmetry of the charge distribution, the field intensity should be the same at all points that are the same distance radially from the surface of the sphere. This assertion is easily verified experimentally. We therefore turn to a presentation of Gauss's law, the method by which the field intensity is calculated for symmetric charge distributions.

To derive this law, let us take an arbitrary surface A, called a Gaussian surface, which encloses a positive point charge q, as illustrated in Figure 30-1. Let q be a distance r from an element of area dA on the surface A, and $\mathbf{n}$—a unit vector perpendicular to dA in an outward direction. The magnitude of the field intensity $\mathbf{E}$ at dA has the value

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

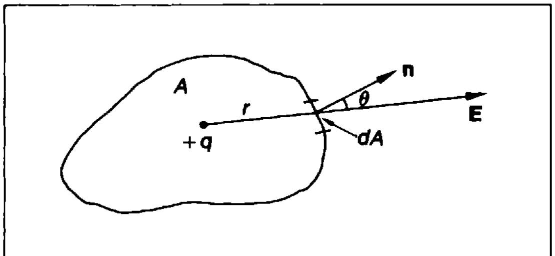


Figure 30-1 Gaussian surface surrounding a point charge.

The product of E normal to the surface, $\mathbf{E}\cdot\mathbf{n}$, and the area dA is

$$\mathbf{E}\cdot\mathbf{n}\,dA = E\cos\theta\,dA = \frac{1}{4\pi\epsilon_0}\frac{q\cos\theta\,dA}{r^2},$$

where θ is the angle between $\mathbf{E}$ and $\mathbf{n}$.

The solid angle $d\omega$ in steradians subtended by dA at q, is, by definition,

$$d\omega = \frac{\cos\theta\,dA}{r^2}.$$

Therefore,

$$\mathbf{E}\cdot\mathbf{n}\,dA = \frac{1}{4\pi\epsilon_0}q\,d\omega.$$

We now integrate both sides of the equation over the entire closed surface,

$$\oint_A \mathbf{E}\cdot\mathbf{n}\,dA = \frac{q}{4\pi\epsilon_0}\oint d\omega.$$

Since a closed surface subtends a total solid angle of 4π steradians, at any point within the volume enclosed by the surface

$$\oint_A \mathbf{E}\cdot\mathbf{n}\,dA = \frac{q}{\epsilon_0}. \tag{30-1}$$

Equation (30-1) is Gauss's law for the special case where the enclosed charge is a single point charge. Suppose the enclosed charge is some arbitrary distribution of charge. Since any distribution of charge can be considered the sum of a number of point charges, we can write Eq. (30-1) as follows:

$$\oint_A \mathbf{E}\cdot\mathbf{n}\,dA = \frac{\Sigma q}{\epsilon_0}, \tag{30-2}$$

where the summation is taken only over the charges lying within the closed surface. The result is independent of the way in which the charge is distributed. It also does not depend on whether the charge is positive or negative. Why?

Equation (30-2) is a mathematical expression of Gauss's law, namely:

> The surface integral of the normal component of **E** over any closed surface in an electrostatic field equals $\Sigma q/\epsilon_0$, where the Σq is the net charge inside the surface.

In dealing with a continuous distribution of charge, we can write Gauss's law as

$$\oint_A \mathbf{E}\cdot\mathbf{n}\,dA = \frac{1}{\epsilon_0}\oint_v \rho\,dv, \tag{30-3}$$

where ρ is the volume charge density inside the volume v enclosed by the Gaussian surface A. In the general case, ρ is a function of position.

30-2 APPLICATIONS OF GAUSS'S LAW

Gauss's law can be used to solve a number of electrostatic field problems involving a special symmetry. We shall now apply it to a few such problems.

(a) Field of a Charged Spherical Conductor

Let us find the magnitude of E at a point P which is a distance r from the center of a spherical conductor of radius a bearing a total charge $+a$ uniformly distributed over the surface. We first choose a spherical surface (Gaussian surface) A of radius $r > a$ around the charge distribution and concentric with it, as shown in Figure 30-2. We now apply Gauss's law. E has the same magnitude at all points on A since E can only be a function of r. Therefore,

$$\oint_A \mathbf{E}\cdot\mathbf{n}\,dA = E\oint_A dA = 4\pi r^2 E = \frac{Q}{\epsilon_0}.$$

Thus, the electric field has the magnitude

$$E = \frac{Q}{4\pi\epsilon_0 r^2}, \tag{30-4}$$

which is the same as if all the charge were concentrated at the center of the spherical shell. If the Gaussian surface had a radius $r < a$, the charge enclosed would be zero (why?) so that $\mathbf{E} = 0$ inside the sphere. The above results are the same for a hollow charged spherical conductor, since for a solid conductor all of the charges are on the outside.

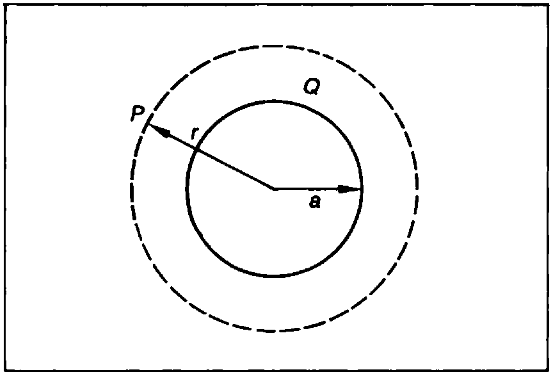


Figure 30-2 Spherical Gaussian surface surrounding a charged sphere.

(b) Field of a Charged Conducting Plate

Suppose that a charged plate of finite thickness is infinite in extent and that the charge per unit area is σ. Considerations of symmetry lead us to believe that the field direction is everywhere perpendicular to the plate and must be the same on each side. This time we choose for our Gaussian surface a cylinder with end areas A, one within and one outside the plate, and side walls perpendicular to the plate, as shown in Figure 30-3.

No lines of force cut the side walls of the cylinder, since they are parallel to the side walls, so that the component of **E** perpendicular to these walls is zero. Since **E** equals zero everywhere inside the plate, the surface integral of **E** calculated over the entire surface of the cylinder is EA, where A is the area of the end of the cylinder outside the plate. The total charge enclosed by the cylinder is σA. Hence, from Gauss's law,

$$EA = \frac{1}{\epsilon_0}\sigma A,$$

from which

$$E = \frac{\sigma}{\epsilon_0}. \qquad (30\text{-}5)$$

Example 1. A soap bubble 10 cm in radius with a wall thickness of 3.33×10^{-6} cm is charged to a potential of 100 volts. Show that if it breaks and falls as a spherical drop the potential of the drop is 10,000 volts.

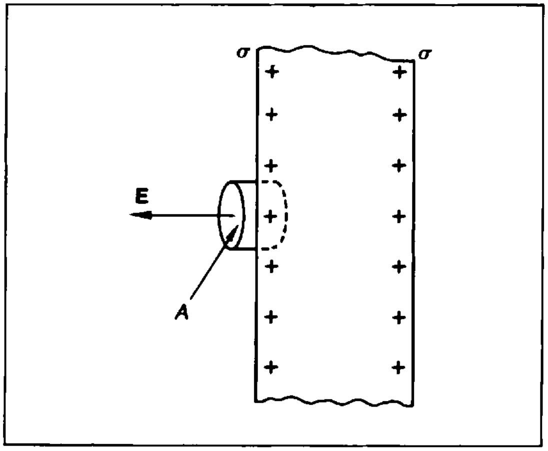


Figure 30-3 Cylindrical Gaussian surface in a plane conducting plate.

SOLUTION

$$\text{Volume} = 4\pi r_1^2 \Delta r_1 = (4\pi)((0.1)^2\,\text{m}^2)(3.33 \times 10^{-8}\,\text{m})$$

$$\frac{4}{3}\pi r_2^3 = (4\pi)(0.01\,\text{m}^2)(3.33 \times 10^{-8}\,\text{m})$$

$$r_2^3 = (3 \times 0.01)(3.33 \times 10^{-8})\,\text{m}^3$$

$$= 1 \times 10^{-9}\,\text{m}^3.$$

Therefore,

$$r_2 = 1 \times 10^{-3}\,\text{m}$$

$$\frac{V_2}{V_1} = \frac{\dfrac{Q}{4\pi\epsilon_0 r_2}}{\dfrac{Q}{4\pi\epsilon_0 r_1}} = \frac{r_1}{r_2}$$

$$\frac{V_2}{100\,\text{volts}} = \frac{0.1\,\text{m}}{0.001\,\text{m}}$$

$$V_2 = \frac{10\,\text{m-volts}}{0.001\,\text{m}} = 10{,}000\,\text{volts}.$$

Example 2. A hollow metallic spherical shell 0.50 m radius is charged uniformly with a negative charge of 2 microcoulombs.

(a) What is the electric field intensity at the center of the sphere?

(b) What is the electric potential at the center of the sphere?

(c) What is the magnitude of the electric field intensity at a point 1 m from the center of the sphere?

(d) What is the electric potential at a point 1 m from the center of the sphere?

SOLUTION

(a) $\mathbf{E} = 0$ at the center of the sphere, since there is no charge in the space inside the sphere.

(b) The electric potential at the center of the sphere is the same as at its surface. It is constant inside the sphere.

$$V = \mathbf{E} \cdot \mathbf{r} = Er = \frac{1}{4\pi\epsilon_0}\frac{Q}{r}$$

$$= \left(9 \times 10^9 \frac{\text{nt-m}^2}{\text{coul}^2}\right)\left(\frac{2 \times 10^{-6}\,\text{coul}}{0.5\,\text{m}}\right)$$

$$= 36 \times 10^3\,\text{volts}.$$

(c) $$E = \frac{1}{4\pi\epsilon_0}\frac{Q}{r^2} = \left(9 \times 10^9 \frac{\text{nt-m}^2}{\text{coul}^2}\right)\left(\frac{2 \times 10^{-6}\,\text{coul}}{(1)^2\,\text{m}^2}\right)$$

$$= 18 \times 10^3\,\text{volts/m}.$$

(d) $$V = \frac{1}{4\pi\epsilon_0}\frac{Q}{r} = \left(9 \times 10^9 \frac{\text{nt-m}^2}{\text{coul}^2}\right)\left(\frac{2 \times 10^{-6}\,\text{coul}}{1\,\text{m}}\right)$$

$$= 18 \times 10^3\,\text{volts}.$$

Example 3. What is the magnitude of the electric field intensity just outside the surface of a plate 2 m long and 1 m wide, which carries a charge of 4 microcoulombs?

SOLUTION

$$\sigma = \left(\frac{4 \times 10^{-6}\,\text{coul}}{(2\,\text{m})(1\,\text{m})}\right) = 2 \times 10^{-6}\,\text{coul/m}^2$$

$$E = \frac{\sigma}{\epsilon_0} = \frac{2 \times 10^{-6}\,\text{coul/m}^2}{8.85 \times 10^{-12}\,\dfrac{\text{coul}^2}{\text{nt-m}^2}}.$$

$$= 2.26 \times 10^7\,\text{volts/m}.$$

30-3 CAPACITANCE

The capacitance of a conductor is defined as the ratio of the charge carried by the conductor to the potential of the conductor relative to zero. That is,

$$C = \frac{Q}{V}. \qquad (30\text{-}6)$$

For the case of two conductors, the charge that must be transferred from one to the other per unit potential difference is defined as the capacitance between the two conductors,

$$C = \frac{Q}{\Delta V}. \qquad (30\text{-}7)$$

In MKS units, Q is in coulombs and V is in volts. Therefore, from Eq. (30-6), C is in coul/volt and is given the name *farad.*

$$1\ \text{farad} = \frac{1\ \text{coul}}{1\ \text{volt}}.$$

The farad is too large a unit of capacitance for practical work, so smaller units are usually chosen, such as the microfarad (10^{-6} farads) and the pico- or micromicrofarad (10^{-12} farads).

The combination of two nearby conductors having equal and opposite charge is called a capacitor.

30-4 THE PARALLEL-PLATE CAPACITOR

Let us calculate as an example the capacitance of the parallel plate capacitor of Figure 30-4. The electric field will be uniform and perpendicular to the plates everywhere except near the edges, a region that we neglect here. If σ is the surface charge density on one plate, Eq. (30-5) tells us that

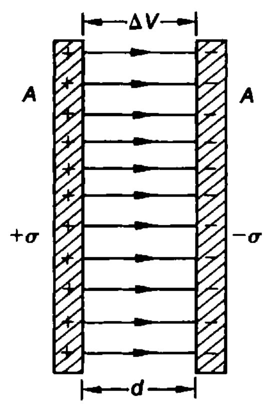


Figure 30-4 Parallel plate capacitor.

$$E = \frac{\sigma}{\epsilon_0}. \qquad (30\text{-}8)$$

Since the field is uniform, the charge density on the other plate must be the same. By definition,

$$\sigma = \frac{Q}{A}, \qquad (30\text{-}9)$$

where Q is the total charge on a plate. Therefore, the electric field is equal to

$$E = \frac{Q}{\epsilon_0 A}. \qquad (30\text{-}10)$$

The potential difference between the plates is

$$\Delta V = Ed = \frac{\sigma}{\epsilon_0} d. \qquad (30\text{-}11)$$

Hence, the capacitance is

$$C = \frac{Q}{\Delta V} = \frac{\sigma A}{\dfrac{\sigma}{\epsilon_0} d} = \frac{\epsilon_0 A}{d}. \qquad (30\text{-}12)$$

30-5 THE ENERGY OF A CHARGED CAPACITOR

In charging a capacitor, it is necessary to do work to carry the charge from one plate to the other. The amount of work required increases as more charge is transferred. Suppose at the beginning of the charging both plates are uncharged, and at the end they have a charge Q. The element of work at any time to take an element of charge dQ from one plate and deposit on the other through a potential difference ΔV is

$$dW = \Delta V dQ = \frac{Q}{C} dQ.$$

The total work to increase the charge from zero to Q is

$$W = \frac{1}{C}\int_0^Q Q dQ = \frac{1}{2}\frac{Q^2}{C},$$

which may be written

$$W = \frac{1}{2}\frac{Q^2}{C} = \frac{1}{2}C\Delta V^2 = \frac{1}{2}Q\Delta V. \qquad (30\text{-}13)$$

The energy supplied to a capacitor in charging it is stored by the capacitor and released when it discharges. It is reasonable to assume that the energy stored by a capacitor is stored in the electric field, since the electric field increases as Q or ΔV increase.

Let us choose a parallel plate capacitor of area A and plate separation d. Its capacitance, we have seen, is $C = \epsilon_0 A/d$. Also, $Q = \sigma A$. Substitution of these two expressions for C and Q into Eq. (30-13) gives

$$W = \frac{\sigma^2 A d}{2\epsilon_0}.$$

Since the volume v containing the field is just Ad and $E = \sigma/\epsilon_0$, we can write

$$\frac{W}{v} = \frac{1}{2}\epsilon_0 E^2, \qquad (30\text{-}14)$$

where W/v is the energy per unit volume stored in the field. We will later show this to be a more generally valid statement concerning energy storage in an electric field.

30-6 CAPACITORS IN SERIES AND PARALLEL

Figure 30-5 shows two capacitors connected in series. What is the equivalent capacitance of this combination if a potential difference ΔV is maintained across it? The total work done per unit charge taken from one plate to the other on both capacitors is

$$\Delta V = \Delta V_1 + \Delta V_2,$$

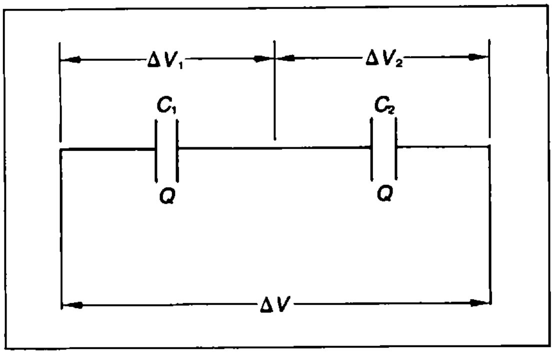


Figure 30-5 Two capacitors in series.

and, since $Q = C\Delta V$, we have

$$\frac{Q}{C} = \frac{Q}{C_1} + \frac{Q}{C_2}$$

or

$$\frac{1}{C} = \frac{1}{C_1} + \frac{1}{C_2}. \qquad (30\text{-}15)$$

Thus, the reciprocal of the equivalent capacitance of a set of capacitors connected in series equals the sum of the reciprocals of the individual capacitances.

Figure 30-6 shows two capacitors connected in parallel. What is the equivalent capacitance of this combination if a potential difference ΔV is maintained across it? If Q_1 is the charge put on one capacitor and Q_2 the charge put on the other, the total charge supplied by the source is

$$Q = Q_1 + Q_2.$$

Each capacitor, however, has the same potential difference so that

$$\Delta V = \Delta V_1 = \Delta V_2.$$

Since

$$Q = C\Delta V,$$

$$C\Delta V = C_1\Delta V + C_2\Delta V$$

or

$$C = C_1 + C_2, \qquad (30\text{-}16)$$

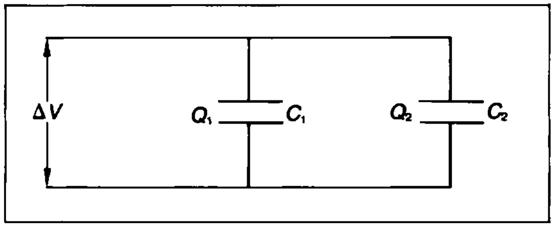


Figure 30-6 Two capacitors in parallel.

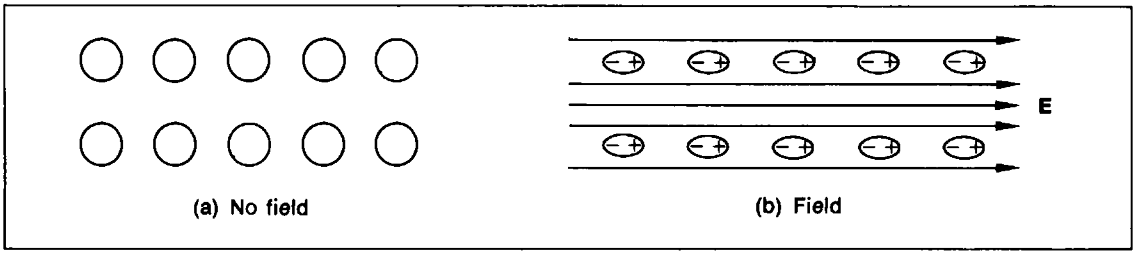


Figure 30-7 Array of non-polar molecules.

Thus, the equivalent capacitance of a set of capacitors connected in parallel is equal to the sum of the individual capacitances.

30-7 A DIELECTRIC AND POLARIZATION OF A DIELECTRIC

A dielectric, or insulator, is a material containing no free charge. However, a charge appears on the surface of a dielectric when it is placed between the plates of a charged capacitor due to the rotation of or production of dipoles. A dipole consists of two separated charges of equal magnitude and opposite sign. There are two kinds of dielectrics. They are called polar and non-polar dielectrics. A polar dielectric consists of permanent dipoles, while a non-polar dielectric only possesses dipoles when it is in an electric field. These dipoles are called induced dipoles.

We shall first discuss non-polar dielectrics. In the absence of an external electric field, the electrons of a given atom of a non-polar dielectric are distributed symmetrically around the nucleus, as illustrated in Figure 30-7(a). When a field is applied, the electrons are displaced in the direction opposite to that of the field [Figure 30-7(b)]. The center of gravity remains fixed since there is no translational force on the atom as a whole. Each atom thus acquires an electric dipole moment (**p**), which is parallel and in the same direction as the applied field. The dipole moment **p** of a pair of charges is defined as

$$\mathbf{p} = q\mathbf{d}, \qquad (30\text{-}17)$$

where q is the magnitude of one of the charges and **d** is the charge separation. The action of the electric field in giving each atom of the dielectric an induced dipole moment is called polarization.

The polarization (**P**) of a substance is defined as the electric dipole moment per unit volume, and it is proportional in magnitude to the applied electric field.

In a polar dielectric with permanent dipoles, the dipoles are randomly oriented when no external field is present, as shown in Figure 30-8(a). When an external field is applied, an orientation of these dipoles in the direction of the field takes place, as illustrated in Figure 30-8(b). The degree of orientation is increased as the electric field strength is increased.

Consider a dielectric slab with a field **E** applied perpendicular to it as illustrated in Figure 30-9. If the slab is a non-polar dielectric, the field polarizes it; that is, it induces dipole moments throughout the slab. If the slab is a polar dielectric, the field orients the dipoles in the direction of the field. We can readily see that the net effect of the field is the same for both non-polar and polar dielectrics. The interior of the dielectric remains neutral since the

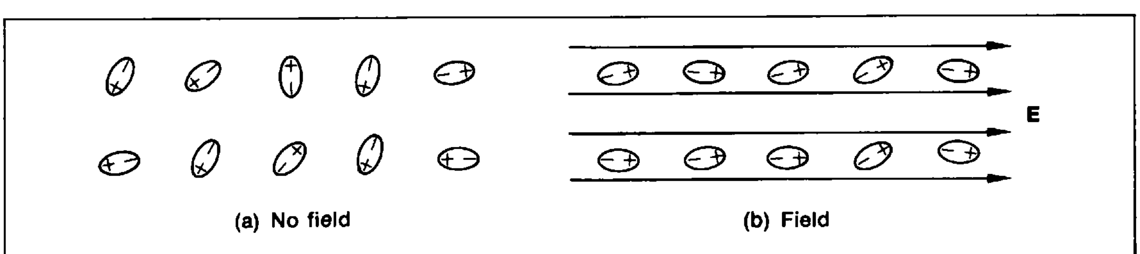


Figure 30-8 Array of polar molecules.

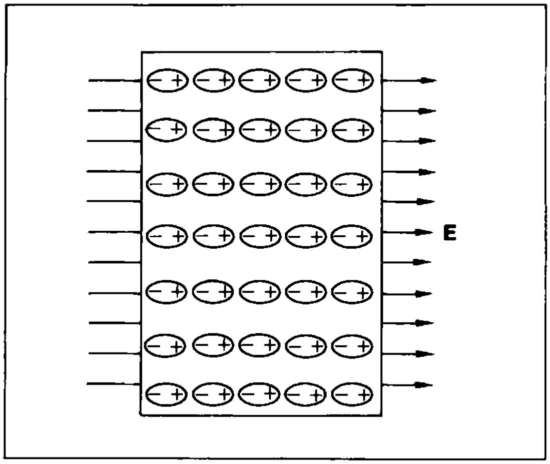


Figure 30-9 Dielectric slab in an electric field.

positive and negative charges of adjacent dipoles cancel out each other's effects. The net result is the production of a layer of bound negative charges on one surface of the slab and a layer of bound positive charges on the other, as depicted in Figure 30-9. From the definition of polarization and dipole moment, it can be seen that $P = \sigma_b$, where σ_b is the bound charge per unit area.

30-8 SUSCEPTIBILITY, DIELECTRIC CONSTANT, AND DISPLACEMENT

It is possible to discuss the consequences of polarization without a consideration of the atomic processes involved. The polarization **P** depends on the nature of the substance and, for most dielectrics, is proportional and parallel to the electric field in the dielectric $\mathbf{E}_D$. We define a property of the dielectric, called its susceptibility χ_e, by the equation

$$\chi_e = \frac{\mathbf{P}}{\epsilon_0 \mathbf{E}_D} \quad \text{or} \quad \mathbf{P} = \chi_e \epsilon_0 \mathbf{E}_D. \tag{30-18}$$

Since only a material substance can be polarized, the susceptibility of a vacuum is zero. One may question why χ_e is not just defined as $\chi_e = \mathbf{P}/\mathbf{E}_D$ since **P** is proportional to $\mathbf{E}_D$. The answer to this question is that with χ_e defined by Eq. (30-18) we may write χ_e as a pure dimensionless number, which is the same regardless of the system of units used for measuring **P** and $\mathbf{E}_D$.

The surface density of bound charge at any point on the surface of a dielectric is equal to the normal component of **P** at the surface. For the special case of a surface perpendicular to **E**,

$$\sigma_b = P = \chi_e \epsilon_0 E_D. \tag{30-19}$$

Consider a dielectric between the plates of a capacitor, as illustrated in Figure 30-10. In the region between the plates and the dielectric, we encounter two different kinds of charge. On the surface of the dielectric, there is a concentration of bound charge per unit area σ_b equal to P. On the surface of the conductor, there is a concentration of free charge per unit area σ_f from an external source equal to $\epsilon_0 E_f$. Since these charge concentrations are opposite in sign, we can write

$$\epsilon_0 E_D = \epsilon_0(E_f - E_b) = \sigma_f - \sigma_b. \tag{30-20}$$

Since $\sigma_b = P$, we see that

$$\sigma_f = (\sigma_f - \sigma_b) + \sigma_b = \epsilon_0 E_D + P$$

or

$$\epsilon_0 E_f = \epsilon_0 E_D + P. \tag{30-21}$$

Since

$$P = \chi_e \epsilon_0 E_D,$$

$$\epsilon_0 E_f = \epsilon_0(1 + \chi_e) E_D. \tag{30-22}$$

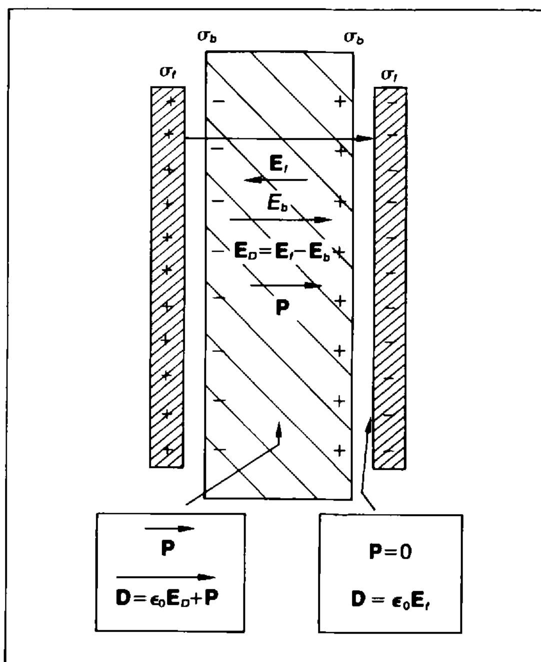


Figure 30-10 Showing $\mathbf{E}_f$, $\mathbf{E}_b$, $\mathbf{E}_D$, and **P** in a parallel plate capacitor.

The quantity $\epsilon_0(1+\chi_e)$ is called the *permittivity* of the dielectric ϵ. Therefore,

$$\epsilon = \epsilon_0(1+\chi_e). \qquad (30\text{-}23)$$

Since the permittivity of a dielectric is always greater than the permittivity of vacuum, it is convenient to have the ratio of the permittivity of the dielectric to that of vacuum. This ratio is called the relative permittivity or the dielectric constant (K) of the dielectric. Thus,

$$K = \frac{\epsilon}{\epsilon_0}. \qquad (30\text{-}24)$$

The dielectric constant is a pure dimensionless number. It is equal to unity for vacuum and is greater than unity for a material substance. Its numerical value can be obtained for a particular dielectric by measuring the capacitance (C) of a capacitor with the dielectric between the plates and the capacitance (C_0) with free space between them. The ratio C/C_0 then gives the dielectric constant of the dielectric directly.

It is convenient to set the quantity $(\epsilon_0 E_D + P)$ of Eq. (30-21) equal to D. D is given the name *electric displacement* and, in general, is a vector quantity defined by the vector equation

$$\mathbf{D} = \epsilon_0\mathbf{E}_D + \mathbf{P} = \epsilon_0\mathbf{E}_f = \sigma_f. \qquad (30\text{-}25)$$

Equation (30-22) can now be written

$$\mathbf{D} = \epsilon_0(1+\chi_e)\mathbf{E}_D = \epsilon\mathbf{E}_D = \sigma_f. \qquad (30\text{-}26)$$

The electric displacement $\mathbf{D}$ has some similarity to $\mathbf{E}_D$, but depends only on the free charge, while $\mathbf{E}_D$ depends on both the free and bound charges.

From Eq. (30-25), as illustrated in Figure 30-10, we see that while the magnitude of $\mathbf{E}_D$ in the dielectric differs from that in the gap between the dielectric and the plate, the magnitude of $\mathbf{D}$ does not differ from one location to the other. If we know σ_f on the plates of a capacitor and ϵ for the dielectric, we may find $\mathbf{E}_D$ from Eq. (30-26).

We have seen that in free space the energy per unit volume stored in the field equals $\frac{1}{2}\epsilon_0 E^2$. For a dielectric material, we replace ϵ_0 by ϵ, which leads us to the expression for the energy stored per unit volume in a dielectric—namely,

$$\frac{W}{v} = \frac{1}{2}\epsilon E^2 = \frac{DE}{2} = \frac{D^2}{2\epsilon}. \qquad (30\text{-}27)$$

We can readily see that the stored energy per unit volume with a material dielectric is greater than that for free space. This is due to the extra work that must be done to polarize the dielectric.

Example 4. A parallel plate capacitor in air with plates separated by 0.4 mm when charged with 1 microcoulomb is found to have a potential difference of 40 volts between the plates. What is the area of each plate?

SOLUTION

$$C = \frac{\epsilon_0 A}{d}$$

and

$$\Delta V = \frac{Q}{C} = \frac{Qd}{\epsilon_0 A}$$

or

$$A = \frac{Qd}{\epsilon_0 \Delta V} = \frac{(10^{-6}\text{ coul})(4\times 10^{-4}\text{ m})}{\left(8.85\times 10^{-12}\dfrac{\text{coul}^2}{\text{nt-m}^2}\right)(40\text{ volts})}$$

$$= \frac{10^{-11}\text{ coul-m}}{8.85\times 10^{-12}\dfrac{\text{coul}^2\text{ volts}}{\text{nt-m}^2}}$$

$$= 1.13\text{ m}^2.$$

Example 5. A 2 microfarad capacitor (1), charged to a potential difference of 100 volts, is suddenly connected to and shares its charge with an uncharged 1 microfarad capacitor (2). What is the loss of energy in this process?

SOLUTION

In order for capacitor (1) to share its charge with capacitor (2), they must be connected in parallel.

$$Q = C_1\Delta V_1 = (2\times 10^{-6}\text{ farads})\,(100\text{ volts}) = 2\times 10^{-4}\text{ coul}$$

$$Q_1 + Q_2 = 2\times 10^{-4}\text{ coul} = C_1\Delta V + C_2\Delta V = (2\times 10^{-6}\text{ farads} + 10^{-6}\text{ farads})\ \Delta V$$

$$\Delta V = \frac{2\times 10^{-4}\text{ coul}}{3\times 10^{-6}\text{ farads}} = 66.7\text{ volts}$$

$$W_{\text{before}} = \frac{1}{2}C_1\Delta V^2 = \frac{1}{2}(2\times 10^{-6}\text{ farads}) \times [(100)^2\text{ volts}^2] = 10^{-2}\text{ joules}$$

$$W_{\text{after}} = \frac{1}{2}C\Delta V^2 = \frac{1}{2}[(2+1)\times 10^{-6}\text{ farads}] \times [(66.7)^2\text{ volts}^2] = 0.67\times 10^{-2}\text{ joules}$$

$$\Delta W = (10^{-2}\text{ joules}) - (0.67\times 10^{-2}\text{ joules}) = 3.3\times 10^{-3}\text{ joules}.$$

Example 6. Two parallel conducting plates have equal and opposite charges. The space between the plates contains a dielectric with a dielectric constant of 4. The resultant electric field intensity in the dielectric is 10^7 volts/m. Compute:

(a) the free charge density (σ_f) on the plates,
(b) the bound charge density (σ_b) on the surface of the dielectric,
(c) the susceptibility (χ_e) of the dielectric,
(d) the permittivity (ϵ) of the dielectric,
(e) the polarization (**P**) in the dielectric,
(f) the displacement (**D**) in the dielectric,
(g) the energy density in the dielectric.

SOLUTION

(a) $\sigma_f = \epsilon E_D = \epsilon_0 K E_D$

$$= \left(8.85\times 10^{-12}\,\frac{\text{coul}^2}{\text{nt-m}^2}\right)(4)(10^7\text{ volts/m})$$
$$= 354\times 10^{-6}\text{ coul/m}^2.$$

(b) $\sigma_b = \sigma_f - \epsilon_0 E_D$

$$= (354\times 10^{-6}\text{ coul/m}^2) - \left[\left(8.85\times 10^{-12}\,\frac{\text{coul}^2}{\text{nt-m}^2}\right)(10^7\text{ volts/m})\right]$$
$$= 265.5\times 10^{-6}\text{ coul/m}^2.$$

(c) $\chi_e = \dfrac{\epsilon}{\epsilon_0} - 1 = K - 1 = 4 - 1 = 3.$

(d) $\epsilon = K\epsilon_0 = (4)\left(8.85\times 10^{-12}\,\dfrac{\text{coul}^2}{\text{nt-m}^2}\right)$

$$= 3.54\times 10^{-11}\text{ coul}^2/\text{nt-m}^2.$$

(e) $P = \sigma_f - \epsilon_0 E_D = \sigma_b = 265.5\times 10^{-6}\text{ coul/m}^2.$

(f) $D = \epsilon_0 E_D = \sigma_f = 354\times 10^{-6}\text{ coul/m}^2.$

(g) $$\frac{W}{v} = \frac{D^2}{2\epsilon} = \frac{\left[(354\times 10^{-6})^2\,\dfrac{\text{coul}^2}{\text{m}^4}\right]}{(2)\left(3.54\times 10^{-11}\,\dfrac{\text{coul}^2}{\text{nt-m}^2}\right)}$$
$$= 1770\text{ joules/m}^3.$$

PROBLEMS

1. A long solid cylinder of radius a has a uniform volume charge density ρ. Find expressions for the field and for the potential at all points outside the cylinder if the zero potential is defined to be at $r = r_0$.
2. Find expressions for the field and for the potential for all points inside the cylinder of Problem 1.
3. Prove that the potential at a point P surrounded by charges $q_1, q_2, q_3, \ldots$, at distances $r_1, r_2, r_3, \ldots$, from the point P is given by

$$V_p = K\left(\frac{q_1}{r_1} + \frac{q_2}{r_2} + \frac{q_3}{r_3} + \cdots\right),$$

where the zero potential is taken at infinity.

4. Figure 30-11 represents two long coaxial conducting cylinders. The inner conductor is grounded and the outer cylinder carries a positive charge per unit length of λ. Sketch **E** and V as a function of r for all values of r from 0 to ∞.

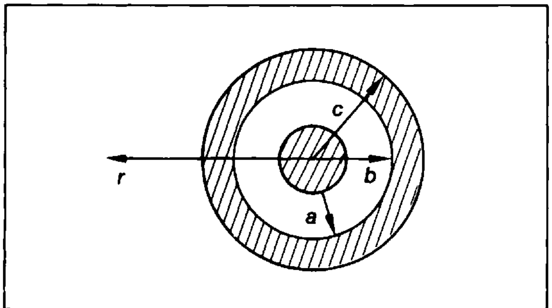


Figure 30-11

5. A solid sphere of radius R has uniform charge density ρ. Find expressions for the field and for the potential at all points inside the sphere if the zero of potential is defined to be at $r = r_0$.
6. Find expressions for the field and potential for all points outside the sphere of Problem 5.

7. Two concentric conducting spherical shells of radii R_1 and R_2 ($R_1 < R_2$) bear charges Q_1 and Q_2, respectively. Find expressions for the electric field at:
   (a) $r < R_1$,
   (b) $R_1 < r < R_2$, and
   (c) $r > R_2$.
8. Find expressions for the potential, relative to zero potential at infinity, in Problem 7 at
   (a) $r < R_1$,
   (b) $R_1 < r < R_2$, and
   (c) $r > R_2$.
9. A capacitor has a capacitance of 2 microfarads when its plates are separated by a layer of air. It is charged to 400 volts by means of a battery.
   (a) Find the charge on the plates.
   (b) Find the energy stored in the capacitor.
10. Compute the energy stored in a 3 microfarad capacitor with air between the plates
   (a) when charged to a potential of 100 volts and
   (b) when the charge on each plate is 2×10^{-4} coulombs.
11. The charged capacitor in Problem 9 is first disconnected from the battery and then immersed in oil having a dielectric constant of 3.
   (a) What is the difference of potential between the plates?
   (b) What is the energy of the capacitor?
   (c) What is the source of energy change?
12. The capacitor in Problem 10, after being charged to a potential of 100 volts, is immersed in a liquid of dielectric constant 2.
   (a) What is the potential difference between the plates?
   (b) What is the energy of the capacitor?
13. Show that the energy of a dipole is given by the relation

$$W = -\mathbf{p} \cdot \mathbf{E}$$

   and the torque on the dipole is given by

$$\mathbf{T} = \mathbf{p} \times \mathbf{E},$$

   where $\mathbf{p}$ is the dipole moment.
14. A 2 microfarad capacitor charged to 100 volts and a microfarad capacitor charged to 200 volts are connected as in Figure 30-12. Find the difference of potential and charge on each capacitor and the loss of energy that has taken place.

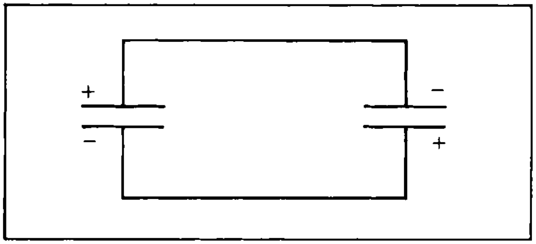


Figure 30-12

15. Two capacitors, of capacitances $C_1 = 3$ micromicrofarads and $C_2 = 6$ micromicrofarads, are connected in series, and the resulting combination is connected across 1000 volts. Compute:
   (a) the equivalent capacitance of the combination,
   (b) the total charge on the combination and the charge on each capacitor,
   (c) the potential difference across each capacitor, and
   (d) the energy W stored in the capacitors.
16. (a) Calculate the capacitance of a capacitor consisting of 2 parallel plates 100 cm^2 in area separated by a layer of paraffin 1 mm thick and dielectric constant 2.
   (b) If the capacitor is connected to a 200 volt source, calculate the charge on the capacitor and the energy stored in the capacitor.

17. Two oppositely charged conducting plates, having numerically equal quantities of charge per unit area, are separated by a dielectric 5 mm thick of dielectric constant 1.5. The value of **E** in the dielectric is 10^6 volts/m. Compute:
    (a) the free charge per unit area on the conducting plates,
    (b) the bound charge per unit area on the surface of the dielectric,
    (c) the polarization **P** in the dielectric,
    (d) the displacement **D** in the dielectric, and
    (e) the energy density in the dielectric.
18. Calculate the capacitance of a spherical capacitor, formed from two thin metal spheres of radii R_1 and R_2, the space between which is filled with a dielectric of dielectric constant K.
19. For the capacitor of Problem 18, find **E**, **P**, and **D** for a point in the dielectric medium when the capacitor bears a charge Q.
20. In Figure 30-13, find the capacity between the points A and B. A 40 volt battery is connected between A and B. What is the energy stored in the 4 μf capacitor?

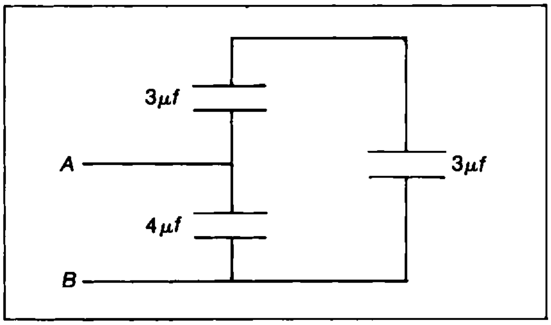


Figure 30-13

31 Direct Electric Currents

In this chapter we make the transition from static electricity to current electricity. We will discuss the flow of a charge in conductors and introduce the concepts of current resistance and electromotive force. Following this, we will give a detailed discussion of direct current electric circuits and some measuring instruments.

31-1 CURRENT AND CURRENT DENSITY

Electric current is defined as the time rate of flow of charge across any cross section of a conductor.

$$I = \frac{dQ}{dt}. \tag{31-1}$$

The MKS unit of current is the ampere. One ampere is equal to one coulomb per second. Direct or steady currents are constant in time; that is, equal quantities of charge pass a given cross section of a conductor in equal intervals of time. The direction of the current is taken conventionally as being that in which a positive charge would move in an electric field. In a majority of cases, it is actually a negative charge flow in the opposite direction that produces the current.

If we are dealing with a current distributed uniformly over a cross-sectional area A of a conductor perpendicular to the direction of the current, we can express the current in terms of the current density j, which is the current per unit area by the equation

$$I = jA. \tag{31-2}$$

The general relationship for the total current flowing through a surface area at any orientation, with the current direction as shown in Figure 31-1, is

$$I = \int_A \mathbf{j} \cdot \mathbf{n}\, dA, \tag{31-3}$$

where $\mathbf{n}$ is the unit vector normal to the surface and dA is an element of the area A. The integral is taken over the surface through which we are calculating the current, and is independent of the shape of the surface.

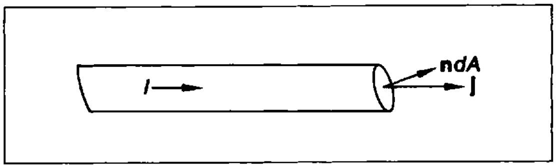


Figure 31-1 Relationship between current and current density.

Since current is due to charged particles of charge e (electron charge), with a mean drift velocity v and density N per unit volume, the current density is

$$\mathbf{j} = Ne\mathbf{v}. \qquad (31\text{-}4)$$

Thus, j is a vector whose direction is that of the velocity v of the particles.

31-2 OHM'S LAW, RESISTANCE, RESISTIVITY, AND CONDUCTIVITY

In many conductors (for example, metals), the current I is proportional to the difference of potential ΔV,

$$I \alpha \Delta V$$

or

$$\Delta V = IR, \qquad (31\text{-}5)$$

where R is a constant called the resistance. It depends upon the dimensions, material, and temperature of the conductor. The above equation is the usual expression of Ohm's law. The unit of R is the ohm (Ω) if ΔV is in volts and I is in amperes.

Another form of Ohm's law is obtained from the fact that the current through a conductor is proportional to the electric field. This fact may be expressed mathematically in terms of the current density as

$$\mathbf{j} \alpha \mathbf{E}$$

or

$$\mathbf{E} = \rho \mathbf{j}, \qquad (31\text{-}6)$$

where ρ is a constant called the resistivity of the material. This relation contains the same information as the familiar form of Ohm's law [Eq. (31-5)], but it has a more general application. It is a vector equation that applies to any point in the material. It can be applied to a conductor of any shape in which neither the field nor the current density are uniform and integrated to give Eq. (31-5).

The reciprocal of the resistivity $1/\rho$ is called the conductivity of the material σ. Equation (31-6), written in terms of the conductivity, is

$$\mathbf{j} = \sigma \mathbf{E}. \qquad (31\text{-}7)$$

It should be pointed out that, in many materials, ρ and consequently σ are not constant at a particular temperature but vary with the current in the material. These materials are said to be non-ohmic or non-linear.

For a uniform isotropic conductor of length l and cross-sectional area A, with a potential difference ΔV between its ends, the electric field and the current density will be constant throughout the conductor and will have the values

$$E = \frac{\Delta V}{l} \quad \text{and} \quad j = \frac{I}{A}.$$

The resistivity will be given by

$$\rho = \frac{E}{j} = \frac{\Delta V/l}{I/A}.$$

Since $R = \frac{\Delta V}{I}$, we can write

$$R = \rho\frac{l}{A}, \qquad (31\text{-}8)$$

which shows that the resistance of a uniform conductor is directly proportional to its length and inversely proportional to its cross-sectional area. We also see from the previous equations that the units of resistivity and conductivity are ohm-m and $(\text{ohm-m})^{-1}$, respectively.

Example 1. A wire 100 m long and 2 mm in diameter has a potential difference of 1.5 volts between the ends of the wire, and carries a current of 1 ampere.

(a) What is its resistance?

(b) What is the resistivity of the material from which the wire is made?

SOLUTION

(a) $R = \frac{\Delta V}{I} = \frac{1.5\text{ volts}}{1\text{ amp}} = 1.5\text{ ohms}$

(b) $\rho = \frac{RA}{l} = \frac{(1.5\text{ ohms})(\pi)[(0.001)^2\text{ m}^2]}{(100\text{ m})}$

$= 4.7 \times 10^{-8}\text{ ohm-m}.$

The resistivities of various materials at room temperature are given in Table 31-1.

Most metals increase in resistivity with a rise in temperature. The increase is linear over a fairly large temperature interval. The size of the temperature interval and its position on the temperature scale varies from metal to metal. As an example, the interval for copper is from $-200°C$ to $+300°C$ and the interval for platinum is from 0°C to 400°C.

Table 31-1 Resistivity and Temperature Coefficient of Various Metals.

Metal	Resistivity at 20°C (ohm-m)	Temperature coefficient of resistivity per C°
Aluminum	2.8×10^{-8}	4.0×10^{-3}
Copper	1.7×10^{-8}	4.0×10^{-3}
Gold	2.4×10^{-8}	3.4×10^{-3}
Iron	1.0×10^{-7}	5.0×10^{-3}
Nickel	7.8×10^{-8}	6.0×10^{-3}
Platinum	9.8×10^{-8}	3.6×10^{-3}
Silver	1.5×10^{-8}	3.8×10^{-3}
Tungsten	5.5×10^{-8}	4.5×10^{-3}

The ***temperature coefficient of resistivity*** may be defined as follows:

> The temperature coefficient of resistivity is the fractional change in resistivity per degree change in temperature.

The definition is written mathematically as follows:

$$\alpha_{20} = \frac{\Delta \rho}{\rho_{20} \Delta t} \qquad (31\text{-}9)$$

or

$$\rho = \rho_{20}(1 + \alpha_{20} \Delta t), \qquad (31\text{-}10)$$

where Δt is the temperature interval.

The index α_{20} means the temperature 20°C has been chosen as the reference level. Table 31-1 gives the temperature coefficient of resistance for various metals.

Because the changes in physical dimensions are much smaller than the change in resistivity with temperature, we can neglect dimensional changes in considering the change in resistance with temperature of an extended material, and write Eq. (31-9) as

$$\alpha_{20} = \frac{\Delta R}{R_{20} \Delta t}, \qquad (31\text{-}11)$$

since $\rho = RA/l$ and the quantity A/l appears both in numerator and the denominator when we substitute RA/l for ρ in Eq. (31-9).

It should be pointed out that while the resistance of most conductors increases with an increase in temperature, there are a group of materials called intrinsic semiconductors, where the resistance decreases with increase in temperature. Carbon is an example. Also, the resistance of some conductors disappears at very low temperatures. These materials are called superconductors. There are also alloys with a resistance independent of temperature over a wide temperature range. Examples are constantan and manganin, which are used in making standard resistance coils.

31-3 RESISTANCES IN SERIES AND PARALLEL

We will now begin a discussion of direct current circuitry. First we will discuss resistance combinations. Consider three resistances connected in series as illustrated in Figure 31-2. The current must be the same in each resistance since there are no other paths through which some of the current might go. Hence, the potential drops across the resistances R_1, R_2, and R_3 will be

$$\Delta V_1 = IR_1, \quad \Delta V_2 = IR_2, \quad \Delta V_3 = IR_3.$$

The total potential drop (ΔV) across the three resistances will be the sum of the individual potential drops, and will equal the current times the total resistance (R), that is,

$$\Delta V = IR = IR_1 + IR_2 + IR_3$$

and

$$R = R_1 + R_2 + R_3. \qquad (31\text{-}12)$$

Therefore, the total resistance of a number of resistances in series is the sum of the individual resistances.

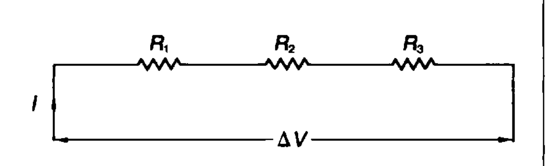


Figure 31-2 Resistances in series.

Next, consider three resistances connected in parallel, as illustrated in Figure 31-3. The total current through the unbranched part of the circuit splits up when it reaches the branch point, with a fraction of the total current passing through each branch. Therefore, the total current I must be equal to the sum of branch currents I_1, I_2, and I_3, that is,

$$I = I_1 + I_2 + I_3.$$

The potential drop in each branch must be the same since all the branches rejoin at a common

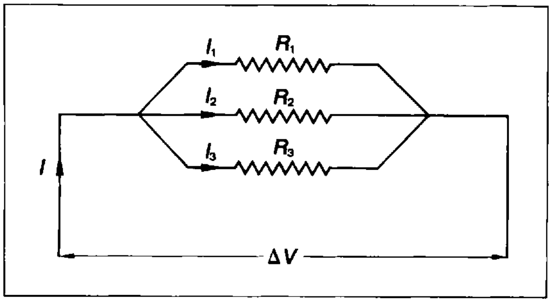


Figure 31-3 Resistances in parallel.

junction. Thus, the currents in each branch are

$$I_1 = \frac{\Delta V}{R_1}, \quad I_2 = \frac{\Delta V}{R_2}, \quad I_3 = \frac{\Delta V}{R_3},$$

and the total current $I = \Delta V/R$, where R is the total resistance. Combining these equations gives

$$\frac{\Delta V}{R} = \frac{\Delta V}{R_1} + \frac{\Delta V}{R_2} + \frac{\Delta V}{R_3}$$

or

$$\frac{1}{R} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3}. \qquad (31\text{-}13)$$

Therefore, the reciprocal of the total resistance of a number of resistances in parallel is the sum of the reciprocals of the individual resistances.

The total resistance of a number of resistances in parallel is less than that of the smallest individual resistance. Why?

31-4 ENERGY AND POWER IN ELECTRIC CIRCUITS, ELECTROMOTIVE FORCE

Now we will consider energy and power in electric circuits. When a constant electric current I passes through an ohmic resistance, the energy dissipated appears in the form of heat. When the potential drop across a resistance R is ΔV, the energy W dissipated in the resistance per unit charge equals ΔV, that is,

$$\Delta V = \frac{W}{Q}.$$

Since

$$Q = I\Delta t,$$

where Δt is the time interval, we have

$$W = \Delta V I \Delta t.$$

The rate at which energy is dissipated, or the power P, is

$$P = \frac{W}{\Delta t} = \Delta V I = I^2 R = \frac{\Delta V^2}{R}. \qquad (31\text{-}14)$$

When a charge is taken around a circuit, work is done on it. To provide this work, a source of energy is required. The source of energy may be chemical as in a battery, mechanical as in a generator, or thermal as in a thermocouple, although the latter situation is of less practical importance. When considering sources of electrical energy in a circuit, a term called the *electromotive force* ($\mathscr{E}$) of the source is used and is defined as follows:

> **The electromotive force (emf) in a circuit is equal to the energy supplied per unit positive charge in carrying the charges around the circuit.**

It should be pointed out that the work done on the charge or the energy dissipated in the circuit may occur within the source of the emf as well as outside it. In a steady state, the rate at which electric energy is put into an electric circuit is equal to the rate of dissipation of the electric energy in the circuit. Consider a simple circuit containing a battery B of emf ($\mathscr{E}$) and an external resistance R, as illustrated in Figure 31-4. In the battery, the electrolyte and the plates have a certain resistance, referred to as the internal resistance R_i. The power supplied must equal the power dissipated and, therefore,

$$I\mathscr{E} = I^2R + I^2R_i.$$

The potential difference across the external circuit,

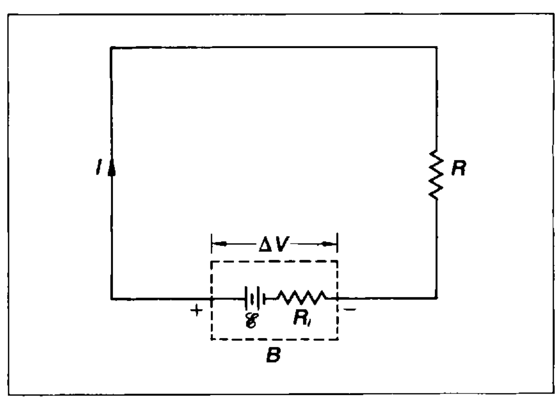


Figure 31-4 A circuit with a battery of emf $\mathscr{E}$, terminal voltage V, and internal resistance R_i, in series with an external resistance R.

which is the potential difference across the terminals of the battery, is given by $\Delta V = IR$. Using this in the above relation and dividing by I, we have

$$\mathscr{E} = \Delta V + IR_i. \qquad (31\text{-}15)$$

When current is being drawn from the battery, the potential difference between the terminals is less than the emf by an amount IR_i. The potential difference equals the emf when the current has been reduced to zero.

31-5 ELECTRIC CIRCUITS AND KIRCHHOFF'S RULES

Many circuits are more complex than the simple circuit illustrated in Figure 31-4. There are sometimes several conducting paths each with resistors, batteries, or generators. Many complex circuits can be broken down into somewhat more simple series and parallel combinations allowing a somewhat simple solution for the current in the circuit and the difference of potential across each element. However, in many complicated circuits this is not possible.

A complicated network can be solved with the help of two general principles, known as Kirchhoff's rules. They are:

1. The total current flowing into a junction of two or more conducting paths in a circuit is equal to the total current flowing out of the junction.
2. Around any closed loop in a circuit, the algebraic sum of the potential differences is equal to the albegraic sum of the emf's in the loop.

In the application of Kirchhoff's rules, we first assign a current direction in each loop. This choice may be correct or incorrect. If it is incorrect, the current comes out negative, which means that the current flows in a direction opposite to the chosen direction. It is also necessary to choose the signs of the emf's and potential differences as we go around the loop or, in other words, we must adopt a set of sign conventions. The sign conventions are the following:

1. An emf is taken as positive if we proceed from the positive to the negative terminal of the source in going around the loop, and negative if we proceed from the negative to the positive terminal.
2. A potential difference is taken as positive if we proceed in the same direction as the assigned loop current and negative if we proceed in an opposite direction.

The method of applying Kirchhoff's rules is best illustrated by an example.

Example 2. Find the current through the 50 ohm resistor in Figure 31-5.

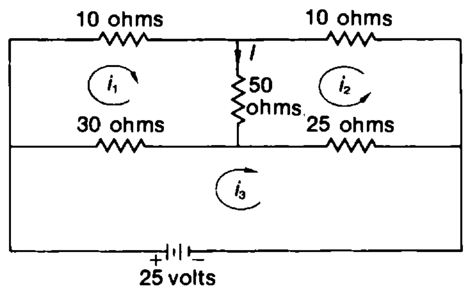


Figure 31-5 Circuit with a divided current path.

SOLUTION

Apply Kirchhoff's first rule to the junction A:

$$i_1 + i_2 = I.$$

Applying Kirchhoff's second rule to the three loops gives

$$\pm(\Omega)(\text{amp}) \pm (\Omega)(\text{amp}) \pm (\Omega)(\text{amp}) = \pm(\text{volts})$$

$$10i_1 + 50(i_1 + i_2) + 30(i_1 - i_3) = 0,$$

$$10i_2 + 50(i_1 + i_2) + 25(i_2 + i_3) = 0,$$

$$30(i_3 - i_1) + 25(i_2 + i_3) = 25.$$

These three equations may be written

$$\pm(\Omega)(\text{amp}) \pm (\Omega)(\text{amp}) \pm (\Omega)(\text{amp}) = \pm(\text{volts})$$

$$90i_1 + 50i_2 - 30i_3 = 0,$$

$$50i_1 + 85i_2 + 25i_3 = 0,$$

$$-30i_1 + 25i_2 + 55i_3 = 25.$$

Solving for i_1 and i_2 from the above three equations, gives $i_1 = -\frac{76}{75}i_2$ and $i_2 = -1.244$ amp. Substituting i_1 and i_2 into the equation from Kirchhoff's first law, gives

$$I = 0.017 \text{ amp downward.}$$

31-6 THE WHEATSTONE BRIDGE AND THE POTENTIOMETER

We shall now discuss two important measuring instruments—namely, the Wheatstone bridge and the potentiometer. The Wheatstone bridge is widely used for measuring resistance. It consists of a circuit as illustrated in Figure 31-6. R_1, R_2, and R_3 are adjustable calibrated resistors, and R_x is an unknown resistance. The resistors are adjusted so that no current flows through the galvanometer *G*, which is a current detecting instrument. Since no current flows in *bd*, the potential difference between *b* and *d* is zero. Therefore,

$$V_{ab} = V_{ad} \quad \text{and} \quad V_{bc} = V_{dc},$$

where V_{ab} is the potential difference between points *a* and *b*, etc.

Using Ohm's law, these equations can be rewritten as

$$I_1R_1 = I_2R_2 \quad \text{and} \quad I_1R_x = I_2R_3.$$

When the second equation is divided by the first, we find

$$\frac{R_x}{R_1} = \frac{R_3}{R_2}. \tag{31-16}$$

If we know R_1, R_3, and R_2 we can calculate R_x. Actually, it is only necessary to know one of the resistances and the ratio of the other two.

Several types of the Wheatstone bridge are made commercially. A simple form of the Wheatstone bridge used in elementary work has two of the resistors replaced by a bare wire of uniform diameter. The wire is divided into two sections by a sliding knife edge, as illustrated in Figure 31-7. Since the wire is of uniform diameter the resistance of either section is proportional to its length. Therefore, when the bridge is balanced,

$$\frac{R_x}{R_1} = \frac{R_3}{R_2} = \frac{Kl_2}{Kl_1}$$

or

$$\frac{R_x}{R_1} = \frac{l_2}{l_1}, \tag{31-17}$$

where R_1 is a resistor of known resistance.

A potentiometer is an instrument that can be used to compare the emf's of two sources, or to measure the emf of one source if the other is known since it does not draw any current from the sources. A simple form of the potentiometer is shown in Figure 31-8. *XY* is a high resistance bare wire, *Z* is a sliding point contact, and *B* is a battery that provides a steady current through *XY*. It must have a greater terminal voltage than the emf of either the unknown battery *X* or the standard battery *S*. The point contact is positioned so that no current flows

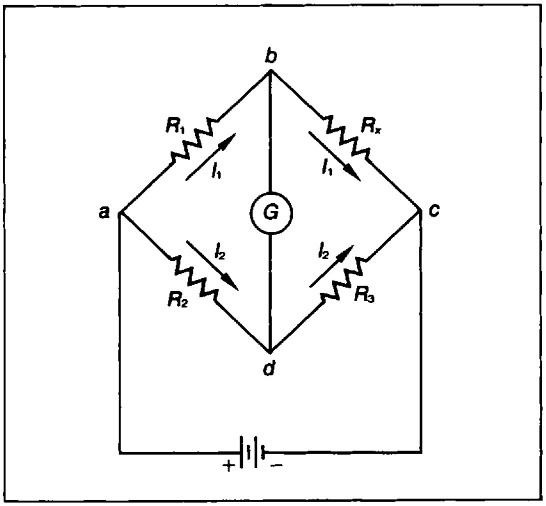


Figure 31-6 Wheatstone bridge circuit.

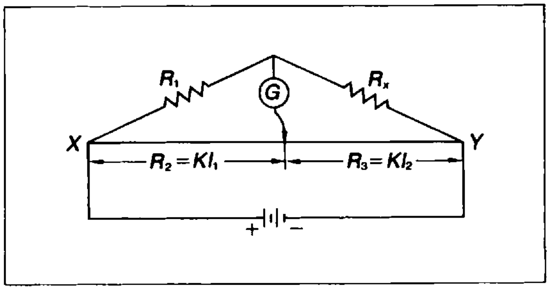


Figure 31-7 Slide wire bridge.

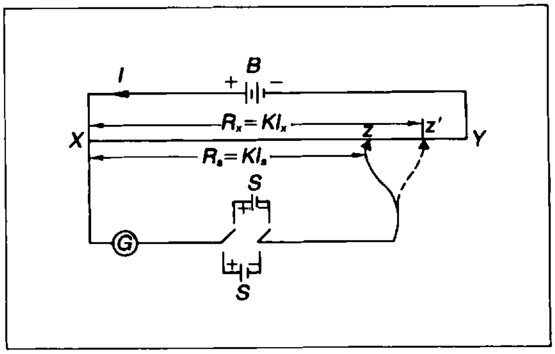


Figure 31-8 Simple potentiometer.

through the galvanometer when S is in the circuit and then repositioned so that no current flows in the galvanometer when X is in the circuit. Let Z and Z' be these positions. This indicates that the potential drop across the wire section l_s and l_x is equal to the emf of the batteries S and X, respectively. Therefore,

$$\mathscr{E}_x = IR_x = IKl_x \quad \text{and} \quad \mathscr{E}_s = IR_s = IKl_s \qquad (31\text{-}18)$$

or

$$\frac{\mathscr{E}_x}{\mathscr{E}_s} = \frac{R_x}{R_s} = \frac{l_x}{l_s}.$$

31-7 SERIES CIRCUIT WITH RESISTANCE AND CAPACITANCE

In the application of Kirchhoff's rules to DC circuits, only batteries, generators, motors, and resistors are important. When a capacitor is in a DC circuit in series with a resistor, there will be a transient current while the capacitor charges or discharges. When a steady state is reached, the current through the capacitor will be zero.

Consider a capacitor of capacitance C in series with a resistor R connected to a battery of voltage† V, then

$$V = \frac{Q}{C} + RI \qquad (31\text{-}19)$$

or

$$\frac{Q}{C} + R\frac{dQ}{dt} - V = 0. \qquad (31\text{-}20)$$

To solve the differential Eq. (31-20), multiply by C and rearrange the terms

$$\frac{dQ}{CV - Q} = \frac{dt}{RC}. \qquad (31\text{-}21)$$

Integration of Eq. (31-21), gives

$$-\ln(CV - Q) = \frac{t}{RC} + \text{constant}. \qquad (31\text{-}22)$$

The constant is determined by setting $Q = 0$ at $t = 0$, and Eq. (31-22) becomes

$$-\ln(CV - Q) = \frac{t}{RC} - \ln(CV)$$

or

$$-\frac{t}{RC} = \ln\left(\frac{CV - Q}{CV}\right). \qquad (31\text{-}23)$$

If we now take the antilogarithm of both sides of this equation, we obtain

$$e^{-t/RC} = \frac{CV - Q}{CV} = 1 - \frac{Q}{CV}$$

or

$$Q = CV(1 - e^{-t/RC}). \qquad (31\text{-}24)$$

This equation shows that the charge on the capacitor approaches exponentially to a value equal to CV. At a time $t = RC$, the charge has increased to within $1/e$ of its final value. The product RC is called the time constant (τ) of the circuit. The current at any time is given by

$$I = \frac{dQ}{dt} = \frac{V}{R}e^{-t/RC} = \frac{V}{R}e^{-t/\tau}. \qquad (31\text{-}25)$$

This shows that the current decreases exponentially, and has the value I/e at $t = \tau$.

If we have an isolated charged capacitor, initially at a voltage V_0, and resistance R is then connected across it at time $t = 0$, the differential equation for the charge on the capacitor at any time as it discharges is given by Eq. (31-20) with $V = 0$, which is

$$\frac{Q}{C} + R\frac{dQ}{dt} = 0.$$

A solution of this equation gives

$$Q = CV_0e^{-t/RC} = CV_0e^{-t/\tau}, \qquad (31\text{-}26)$$

which shows that the charge decays exponentially to zero. The discharge current is

$$I = \frac{dQ}{dt} = -\frac{V_0}{R}e^{-t/RC} = -\frac{V_0}{R}e^{-t/\tau}, \qquad (31\text{-}27)$$

and is the same as the expression for the charging current except that it has a minus sign indicating that it is oppositely directed.

Example 3. In the circuit of Figure 31-9, $V = 10$ volts, $R = 10^6$ ohms, and $C = 2$ microfarads. After the capacitor is fully charged, switch S_2 is opened and S_1 is closed.

(a) Find the charge on the plates of C at times $t = 0$, 2 seconds, and ∞.

(b) What is the initial value of I?

†For simplicity in notation we use the term voltage and the symbol V rather than the term potential difference and the symbol ΔV. Both notations, however, refer to the same concept.

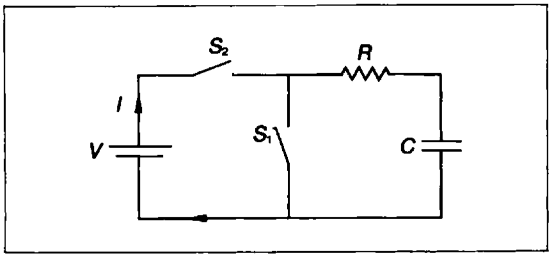


Figure 31-9

SOLUTION

(a) Apply Eq. (31-26); namely, $Q = CVe^{-t/RC}$. At time $t = 0$,

$$Q = (2 \times 10^{-6}\ \text{farads})(10\ \text{volts})[e^{-(0\ \text{sec})/[(10^6\ \Omega)(2\times 10^{-6}\ \text{farads})]}]$$
$$= (2 \times 10^{-6}\ \text{farads})(10\ \text{volts})$$
$$= 2 \times 10^{-5}\ \text{coul.}$$

At time $t = 2$ seconds,

$$Q = (2 \times 10^{-6}\ \text{farads})(10\ \text{volts})[e^{-(2\ \text{sec})/[(10^6\ \Omega)(2\times 10^{-6}\ \text{farads})]}]$$
$$= (2 \times 10^{-6}\ \text{farads})(10\ e^{-1}\ \text{volts})$$
$$= 7.32 \times 10^{-6}\ \text{coul.}$$

At time $t = \infty$

$$Q = (2 \times 10^{-6}\ \text{farads})(10\ \text{volts})[e^{-(\infty\ \text{sec})/[(10^6\ \Omega)(2\times 10^{-6}\ \text{farads})]}]$$
$$= (2 \times 10^{-6}\ \text{farads})(0\ \text{volts})$$
$$= 0\ \text{coul.}$$

(b) Apply Eq. (31-25); namely, $I = \frac{V}{R} e^{-t/RC}$,

$$I = \frac{10\ \text{volts}}{10^6\ \Omega} e^{-(0\ \text{sec})/[(10^6\ \Omega)(2\times 10^{-6}\ \text{farads})]} = \frac{10\ \text{volts}}{10^6\ \Omega}$$
$$= 10^{-5}\ \text{amps.}$$

PROBLEMS

1. A wire of 5 ohms resistance is stretched to increase its length by 10%, the volume and conductivity of the wire remaining constant. What is the new resistance?
2. A rod of material A ($\rho = 2 \times 10^{-7}$ ohm-m) of uniform cross section 0.5 cm^2 and 2 m long is connected in series with a rod of material B ($\rho = 1.5 \times 10^{-7}$ ohm-m) of cross section 0.3 cm^2 and 1.8 m long. If a current is passed through the combination, find the ratio of the potential drop across A to the potential drop across B.
3. At 0°C, the resistance of a certain wire is 200 ohms; at 40°C, it is 220 ohms. What is the temperature coefficient of resistance of this wire?
4. A current of 1 ampere flows in the 15 ohm resistance of the circuit shown in Figure 31-10. What is the potential difference between X and Y?

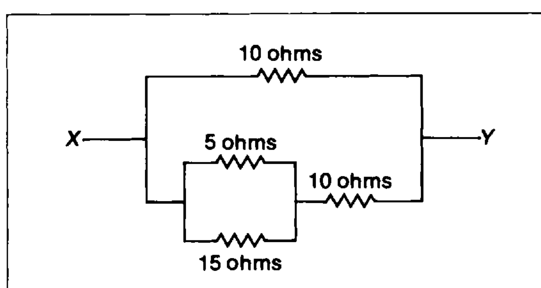


Figure 31-10

5. Three 4-ohm resistors are connected as in Figure 31-11. What is the resistance between the points A and B?

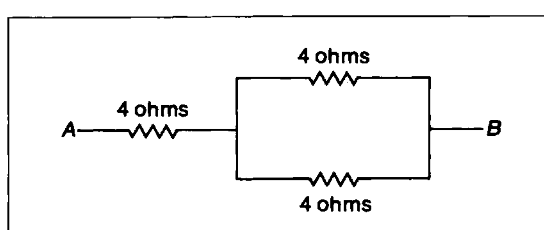


Figure 31-11

6. Two cells, one of emf 1.5 volts and internal resistance 0.2 ohms, the other of emf 3 volts and internal resistance 0.3 ohms, are connected in parallel. The combination is connected in series with an external resistance of 5 ohms. What current flows through the external resistance?

7. The potential difference across the terminals of a battery is 10.8 volts when there is a current of 4 amperes in the battery from the negative to the positive terminal. When the current is 3 amperes in the reverse direction, the potential difference becomes 12.9 volts.
   (a) What is the internal resistance of the battery?
   (b) What is the emf of the battery?

8. When the switch S is open in the circuit shown in Figure 31-12, the voltmeter V, connected across the terminals of the dry cell, reads 1.52 volts. When the switch is closed, the voltmeter reading drops to 1.37 volts and the ammeter reads 1.5 amperes. Find the emf and the internal resistance of the cell, assuming that the voltmeter reading with the switch open is the emf of the dry cell.

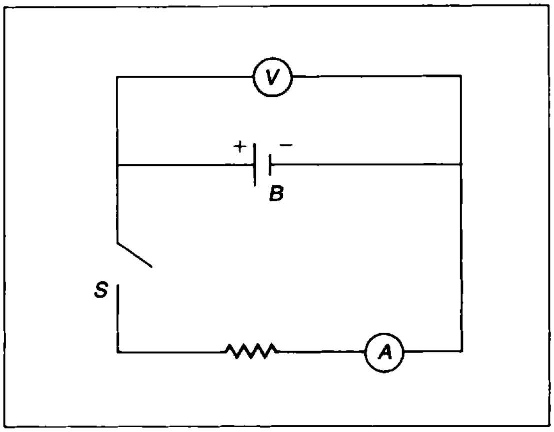


Figure 31-12

9. A series circuit consists of a 6 volt battery with an internal resistance of 0.4 ohms, a 3 volt battery with an internal resistance of 0.6 ohms connected so as to aid the 6 volt battery, and a resistor of 5 ohms. What will be the reading of a voltmeter requiring negligible power to operate connected across the 3 volt battery?

10. A 600 ohm resistor and a 400 ohm resistor are connected in series across a 90 volt line. A voltmeter across the 600 ohm resistor reads 45 volts.
    (a) Find the voltmeter resistance.
    (b) Find the reading of the same voltmeter if connected across the 400 ohm resistor.

11. In a resistor of 100 ohms, there is a constant current of 0.50 amperes for 3 minutes.
    (a) What is the power expended in the resistor?
    (b) How much energy is expended in the resistor?

12. How would you change, with the help of a resistor, a 10 volt voltmeter of internal resistance 1000 ohms to an ammeter which has a full scale reading of 1 ampere. Show the circuit diagram and compute the magnitude of the resistor.

13. The uniform wire of a slide wire Wheatstone bridge is 100 cm long, and a balance is obtained when the sliding contact is 25 cm from one end. If the known resistor of 200 ohms connects to the slide wire at the end nearest the sliding contact, what is the resistance of the unknown resistor?

14. In the DC circuit shown in Figure 31-13,

$I_2 = 1.5$ amps	$R_1 = 2$ ohms
$I_3 = 1$ amp	$R_2 = 2$ ohms
$\mathscr{E}_1 = 4$ volts	$R_3 = 3$ ohms
$\mathscr{E}_2 = 12$ volts	$R_4 = 3$ ohms
$r_2 = 2$ ohms	$R_5 = 1$ ohm

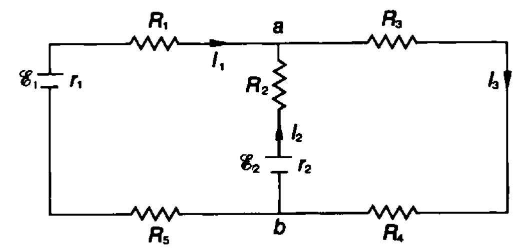


Figure 31-13

Find:
(a) the current I_1,
(b) the power expended in the resistance R_2,
(c) the potential difference V_{ab},
(d) the internal resistance r_1 of the 4 volt battery, and
(e) the terminal voltage across the battery in part (d).

15. In the circuit shown in Figure 31-14,

$I_1 = 3$ amps	$R_3 = 8$ ohms
$V_{ab} = 21$ volts	$R_4 = 4$ ohms
$R_1 = 3$ ohms	$r = 2$ ohms
$R_2 = 6$ ohms	

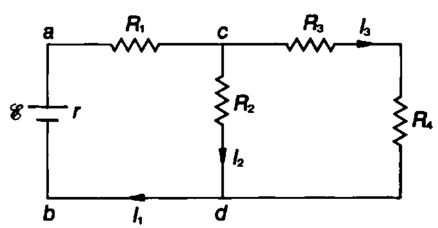


Figure 31-14

Find:
(a) the emf of the battery,
(b) the power expended in the resistance R_1,
(c) the effective resistance of R_2, R_3, and R_4,
(d) the potential difference V_{cd} between c and d, and
(e) the current I_2 through the resistance R_2.

16. Draw the circuit diagram of a potentiometer, label emf's, currents, resistances, and use Kirchhoff's laws to write down equations from them from which solutions could be made for the currents in the circuit.

17. A capacitor of 2 microfarads charged so that the potential difference across its terminals is 120 volts, is suddenly connected to a resistor of 250,000 ohms. What is the time constant of this circuit?

18. A 5 m length of a potentiometer wire is required to balance the emf of a battery. When a 10 ohm resistor is connected across the battery, the length required for balance is 4 m. What is the internal resistance of the battery?

19. The differential equation for the instantaneous charge Q of a capacitor after being disconnected from a source and connected to a resistance R is given by the equation

$$\frac{Q}{C} + R\frac{dQ}{dt} = 0.$$

Show that

$$Q = CVe^{-t/RC} \quad \text{and} \quad I = \frac{V}{R}e^{-t/RC}.$$

20. In the circuit shown in Figure 31-15, $V = 40$ volts, $R = 4 \times 10^6$ ohms, and $C = 10$ microfarads.

(a) Find the initial value of the current when S_1 is closed, the maximum charge on the capacitor plates, and the time constant of the circuit.

(b) After the capacitor is fully charged, S_1 is opened and S_2 is closed. Find the charge on the plates of the capacitor C 4 seconds after closing S_2.

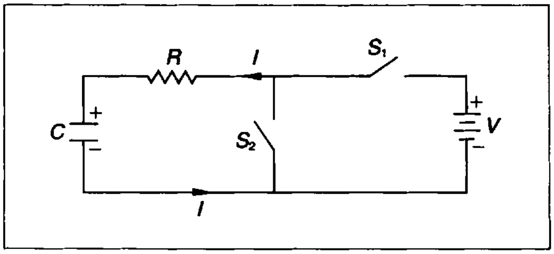


Figure 31-15

32 Magnetic Fields

32-1 MAGNETISM

We turn now to another facet of electromagnetic interactions—namely, magnetism and magnetic fields. Some knowledge of magnetism existed at the time of the ancient Greeks, about 600 B.C. Experimental studies on magnetism began about the sixteenth century. In 1600, William Gilbert published a collection of experimental facts about magnetism. We shall outline here the experimental facts known by Gilbert.

It was observed that certain natural stones called lodestones possess the property of attracting pieces of iron. Also a piece of iron can be made into a magnet by stroking it with a lodestone. It was also found that the attractive force of a magnet is concentrated at regions called poles. When a magnetized needle is freely suspended on a vertical axis, one end of the needle points north and the other end south. The end of the needle that points north is called the north-seeking pole or north pole. The other end is called the south-seeking pole or south pole. Like poles repel and unlike poles attract each other. The magnetic properties of a magnet can be destroyed in several ways, one of which is by heating it.

The space outside a magnet in which its influence can be detected contains a magnetic field. The shape of the magnetic field can be mapped by scattering iron filings in the vicinity of the magnet, as illustrated in Figure 32-1. The direction of the magnetic field at any point is chosen as the direction that a north pole of a freely suspended magnet, or compass needle, points.

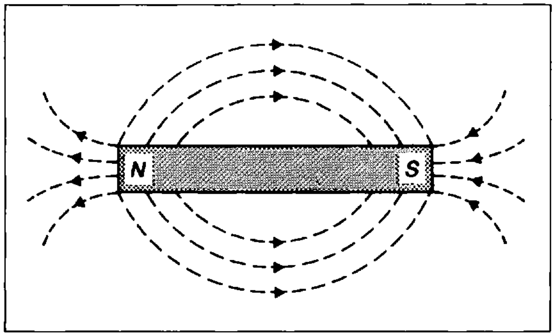


Figure 32-1 Magnetic field in the vicinity of an isolated bar magnet.

32-2 MAGNETIC FORCE ON A CURRENT

The first experiments on the magnetic effects of currents were performed by Ampere in 1820. The

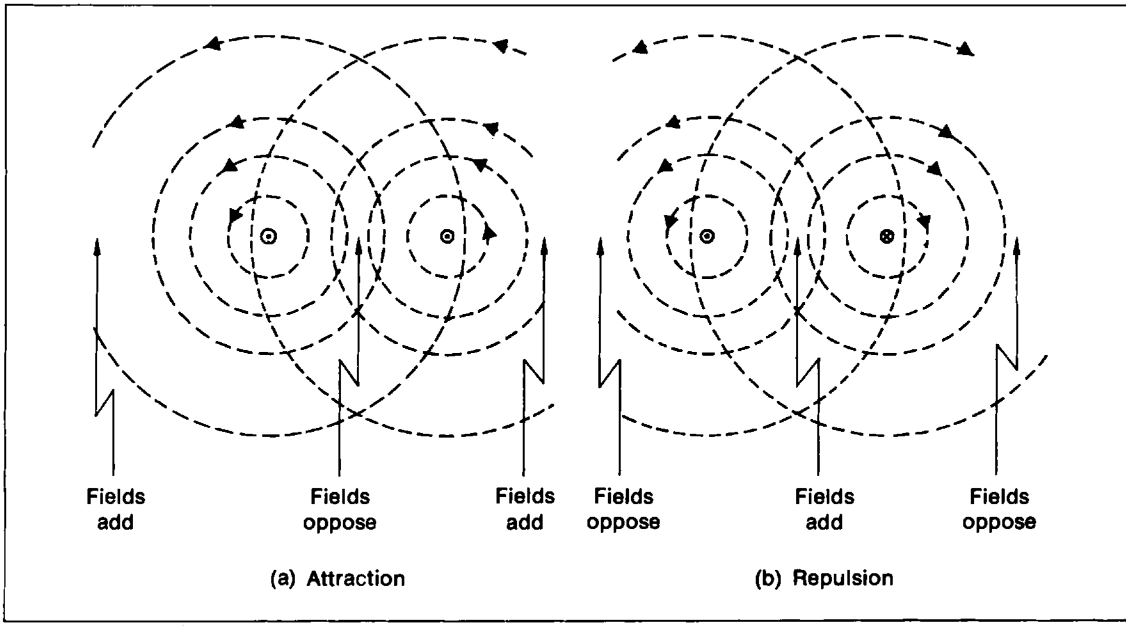


Figure 32-2 Magnetic fields around two wires carrying current.

work was continued by Oersted and others. They found that two long wires carrying currents in the same direction attract one another, whereas when carrying currents in the opposite direction they repel one another.

The magnetic field around a wire carrying a current forms closed circular loops around the wire. The direction is that in which a north pole of a freely suspended magnet or compass needle points, and can easily be remembered by means of the *right-hand rule.* With the thumb pointing in the direction of the conventional current, the fingers of the right hand encircling the wire point in the direction of the magnetic field.

The reason why wires carrying currents attract or repel one another is best illustrated with the aid of Figure 32-2. In Figure 32-2(a), the currents are flowing in the same direction (out of the paper, as illustrated in the figure by a dot · representing the head of an arrow). The fields oppose one another or are in the opposite direction in the space between the wires and reinforce one another or are in the same direction outside the wires. Therefore, the field is stronger outside the wires. If the field lines are considered to represent magnetic flux lines, it may be said that the magnetic flux density is greater outside the wires. It is experimentally observed that the wires move toward one another. In Figure 32-2(b), the currents are flowing in the opposite direction, as is illustrated in the figure by a dot representing the head of an arrow and a cross ⊗ representing the tail of an arrow. The fields here are in the same direction between the wires and in the opposite direction outside the wires. In this case, it is experimentally observed that the wires move apart.

Let us now consider a quantitative relation for the force on a current element in a magnetic field. Let an element of length $d\mathbf{l}$ which carries a current I, be placed in a magnetic field of flux density $\mathbf{B}$ at an angle θ to the direction of the field, as illustrated

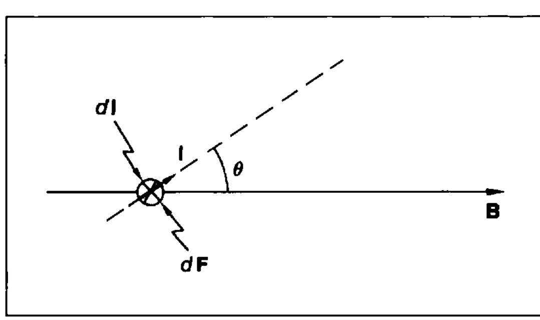


Figure 32-3 The force on a current element in a magnetic field.

in Figure 32-3. The force on this element as inferred from experiments on a closed current loop is written in terms of the vector-product or cross-product as

$$d\mathbf{F} = I d\mathbf{l} \times \mathbf{B}, \qquad (32\text{-}1)$$

where $d\mathbf{F}$ is the force on the element $d\mathbf{l}$ in newtons, I is the current in amperes, $\mathbf{B}$ is the magnetic flux density or magnetic induction in newtons/ampere-meter, and $d\mathbf{l}$ is the element of length in meters. The direction of $d\mathbf{F}$ is perpendicular to the plane containing $d\mathbf{l}$ and $\mathbf{B}$, and is directed into the paper.

32-3 THE BIOT-SAVART LAW

Experiments of Ampere and others showed that the basic relation for magnetic flux density or magnetic induction ($\mathbf{B}$) at a point P produced by a current carrying element, as illustrated in Figure 32-4, is given by

$$d\mathbf{B} = C\frac{I d\mathbf{l} \times \hat{r}_0}{r^2}, \qquad (32\text{-}2)$$

where C is a constant of proportionality, r is the distance from the current carrying element to P, and $\hat{r}_0$ is a unit vector in the direction of r. Equation (32-2) is known as the Biot–Savart law.

The constant of proportionality C is given as

$$C = \frac{\mu}{4\pi},$$

where μ is called the *permeability* constant. The permeability of free space is $\mu_0 = 4\pi \times 10^{-7}$ nt/amp^2. Now, Eq. (32-2) may be written as

$$d\mathbf{B} = \frac{\mu_0}{4\pi}\frac{I d\mathbf{l} \times \hat{r}_0}{r^2}. \qquad (32\text{-}3)$$

For the entire circuit,

$$\mathbf{B} = \frac{\mu_0 I}{4\pi}\oint \frac{d\mathbf{l} \times \hat{r}_0}{r^2}. \qquad (32\text{-}4)$$

From Eqs. (32-1) and (32-3) the force on an element of current $I_1 d\mathbf{l}_1$, due to the magnetic induction from another current element $I_2 d\mathbf{l}_2$, is

$$d\mathbf{F}_1 = I_1 d\mathbf{l}_1 \times d\mathbf{B}_2,$$

where

$$d\mathbf{B}_2 = \frac{\mu_0}{4\pi} I_2 \frac{d\mathbf{l}_2 \times \hat{r}_0}{r^2}.$$

Then

$$d\mathbf{F}_1 = \frac{\mu_0 I_1 I_2}{4\pi} d\mathbf{l}_1 \times \frac{d\mathbf{l}_2 \times \hat{r}_0}{r^2}$$

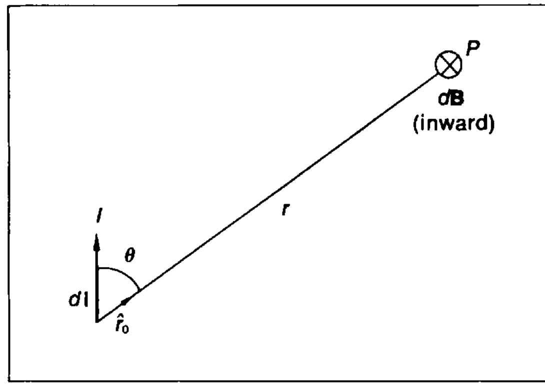


Figure 32-4 Magnetic flux density $d\mathbf{B}$ at a point P due to a current element $d\mathbf{l}$.

or

$$d\mathbf{F}_1 = \frac{\mu_0 I_1 I_2}{4\pi r^2}[d\mathbf{l}_1 \times (d\mathbf{l}_2 \times \hat{r}_0)]. \qquad (32\text{-}5)$$

Example 1. What is the magnitude of the force per unit length on a long wire carrying a current of 1 ampere placed parallel to and 1 m away from another long wire carrying the same current?

SOLUTION

We shall first find an expression for $|\mathbf{B}|$ at a perpendicular distance R meters from a long straight wire.

From Eq. (32-3), the magnetic flux density at P due to the current element dl has the magnitude

$$|d\mathbf{B}| = \frac{\mu_0}{4\pi}\frac{|I d\mathbf{l} \times \hat{r}_0|}{r^2}.$$

The magnitude of the total magnetic flux density at P is then

$$|\mathbf{B}| = \int_{-\infty}^{\infty} \frac{\mu_0 I \sin\theta dl}{4\pi r^2}.$$

Here,

$$dl \sin\theta = r d\theta$$

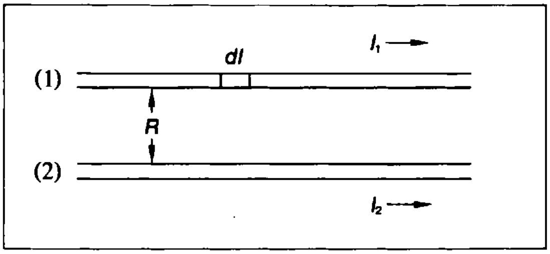
and

$$r = \frac{R}{\sin\theta},$$

Therefore,

$$B = \int_0^\pi \frac{\mu_0 I}{4\pi R} \sin\theta \, d\theta$$
$$= \frac{\mu_0 I}{4\pi R}[-\cos\theta]_0^\pi$$
$$= \frac{\mu_0 I}{2\pi R}.$$

Now we shall consider the two parallel wires as shown



The magnetic flux density a distance R from wire (2) has the magnitude

$$|\mathbf{B}_2| = \frac{\mu_0 I_2}{2\pi R}.$$

The force due to $\mathbf{B}_2$ on $d\mathbf{l}$ of wire (1) is

$$d\mathbf{F} = I_1 d\mathbf{l} \times \mathbf{B}_2$$
$$= \mathbf{n} I_1 I_2 \frac{\mu_0}{2\pi R}\, dl,$$

where $\mathbf{n}$ is a unit vector along R pointing toward wire (2). [If the currents were in the opposite direction, $\mathbf{n}$ would be negative or pointing away from wire (2).] Therefore,

$$\mathbf{F} = \mathbf{n} I_1 I_2 \frac{\mu_0}{2\pi R}\int_0^l dl = \mathbf{n} I_1 I_2 \frac{\mu_0 l}{2\pi R}$$

and

$$\frac{\mathbf{F}}{l} = \mathbf{n}\frac{I_1 I_2 \mu_0}{2\pi R}.$$

With $I_1 = I_2 = 1$ ampere and $R = 1$ m,

$$\frac{\mathbf{F}}{l} = \mathbf{n}\frac{\mu_0 I_1 I_2}{2\pi R} = \left(\frac{4\pi \times 10^{-7}\ \text{nt/amp}^2}{2\pi}\right)\left(\frac{(1\ \text{amp})(1\ \text{amp})}{(1\ \text{m})}\right)$$
$$= (2 \times 10^{-7}\ \text{nt/m})\mathbf{n}.$$

32-4 THE MAGNETIC FIELD OF AND THE FORCES ON A MOVING CHARGE

If a charge q is displaced an amount $d\mathbf{l}$ with a velocity $\mathbf{v}$, we have

$$I d\mathbf{l} = \left(\frac{dq}{dt}\right)\mathbf{v}\,dt = dq\,\mathbf{v}.$$

The Biot–Savart law, namely,

$$d\mathbf{B} = \frac{\mu_0}{4\pi}\frac{I d\mathbf{l} \times \hat{r}_0}{r^2},$$

becomes

$$\mathbf{B} = \int \frac{\mu_0 I d\mathbf{l} \times \hat{r}_0}{4\pi r^2} = \int \frac{\mu_0 dq(\mathbf{v} \times \hat{r}_0)}{4\pi r^2}. \tag{32-6}$$

For a point charge, $\mathbf{v}$ and r are constants in the integration over dq, and we have

$$\mathbf{B} = \frac{\mu_0 q(\mathbf{v} \times \hat{r}_0)}{4\pi r^2}. \tag{32-7}$$

The force on a current element in an external magnetic field is given by Eq. (32-1), namely,

$$d\mathbf{F} = I d\mathbf{l} \times \mathbf{B}.$$

For a moving charge, $I d\mathbf{l} = dq\,\mathbf{v}$ and

$$\mathbf{F} = \int dp\,\mathbf{v} \times \mathbf{B}$$

or

$$\mathbf{F} = q(\mathbf{v} \times \mathbf{B}). \tag{32-8}$$

If an electric field is also present, the total electromagnetic force on q is

$$\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}), \tag{32-9}$$

which is often called the Lorentz force law. Consider the motion of a charge in a uniform magnetic field of flux density $\mathbf{B}$. If the charge is initially moving in a plane normal to $\mathbf{B}$, then the force on it (assuming $\mathbf{E} = 0$) is also in this plane and perpendicular to its direction of motion. The charge, therefore, will move in a circle in this plane.

By Newton's second law, $F = ma$. Since $a = v^2/r$ and $F = qvB$, we obtain

$$\frac{mv^2}{r} = qvB, \tag{32-10}$$

where m is the mass of the charged particle and r is the radius of the circle. The radius of the circle r and the angular velocity ω of the particle will be given by

$$r = \frac{mv}{qB} \tag{32-11}$$

and

$$\omega = \frac{v}{r} = B\left(\frac{q}{m}\right). \qquad (32\text{-}12)$$

The equation shows that the angular velocity and, hence, the time taken to make one revolution is independent of the velocity of the particle so long as the relativistic change of mass with velocity is neglected.

32-5 THE CYCLOTRON

The cyclotron, an ion accelerator developed by E. O. Lawrence in 1931, illustrates Eqs. (32-10) to (32-12). A cyclotron consists of an evacuated chamber placed between the poles of a large electromagnet. The chamber contains two semicircular hollow electrodes called "dees" as illustrated in Figure 32-5. A high frequency alternating potential difference is applied to the dees, which produces an alternating electric field between them. Ions enter the chamber at the ion source F placed between the dees.

The ions are accelerated across the gap by the electric field between the dees, but are shielded inside from the action of the field. The magnetic field causes each moving ion to follow a semicircular path, the radius of which depends on the velocity, mass, and charge of the ion, and the magnetic field strength according to Eq. (32-11). The velocity of the ion increases each time it crosses a gap. As the velocity increases, the radius of the ion path increases. After many trips around, the high energy particles are ejected out of the window W. From Eq. (32-11), the velocity of the particles are given by

$$v = \frac{Brq}{m} \qquad (32\text{-}13)$$

and the final energy of the ions will be given by

$$E = \frac{1}{2}mv^2 = \frac{B^2r^2q^2}{2m}. \qquad (32\text{-}14)$$

One might think that by increasing the radius that an unlimited energy could be obtained. This is not so, since when the velocity of the ion approaches the speed of light, the relativistic increase in mass causes the ions to cross the space between the dees out of step with the changing electric field. As a result, the gain in velocity and, thus, in energy does not occur.

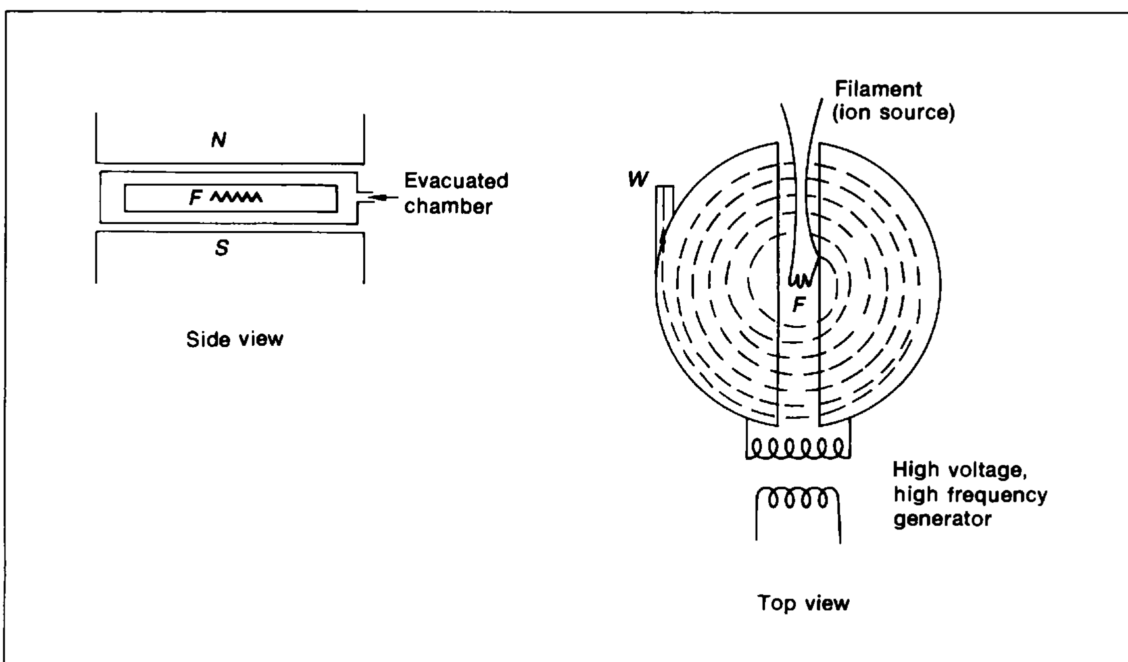


Figure 32-5 Schematic diagram of a cyclotron.

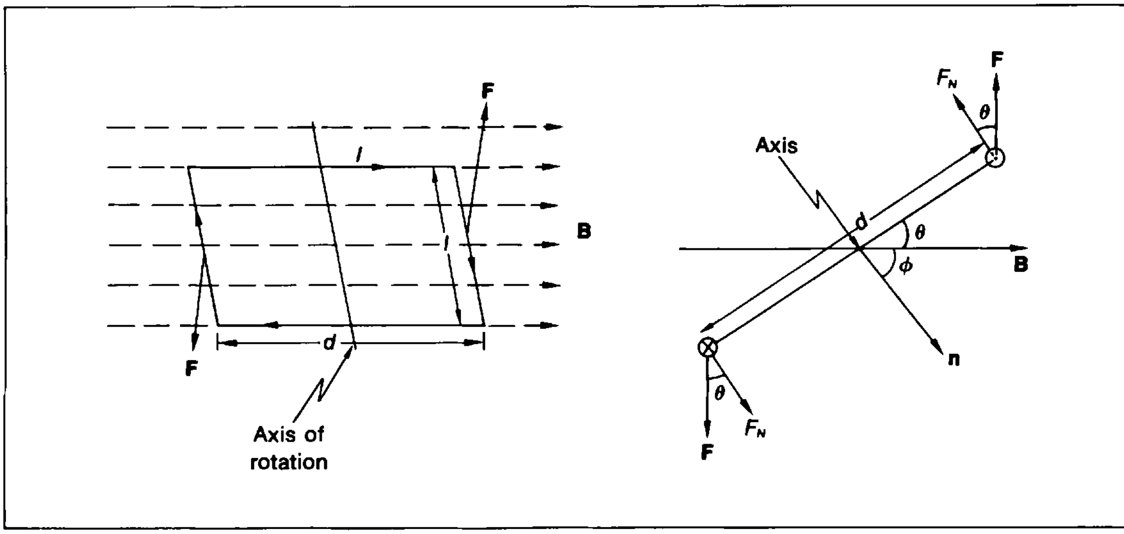


Figure 32-6 Rectangular loop in a magnetic field.

32-6 TORQUE ON A LOOP, DIPOLE MOMENT

We discuss now the torque on a current carrying wire loop in a magnetic field, which is the principle underlying the operation of electric meters. Consider a rectangular loop, carrying a current I, with sides of length l and d situated in a magnetic field of uniform flux density $\mathbf{B}$, as illustrated in Figure 32-6.

The force on any element of the loop is

$$d\mathbf{F} = Id\mathbf{l} \times \mathbf{B}.$$

The component of the force perpendicular to the loop F_n is

$$F_n = F\cos\theta = IB\cos\theta \int_0^l dl = IBl\cos\theta,$$

where θ is the angle between the plane of the loop and the direction of $\mathbf{B}$. The magnitude of the total torque τ on the loop is given by

$$\tau = F_n d = IBld\cos\theta.$$

Since ld equals the area of the loop A, we can write

$$\tau = IBA\cos\theta.$$

The product IA is defined as the magnetic dipole moment (m) of the loop. Since $m = IA$, the torque can be expressed as

$$\tau = mB\cos\theta = mB\sin\phi,$$

where ϕ is the angle between the normal to the plane of the loop and the direction of $\mathbf{B}$.

If the magnetic moment is regarded as a vector $\mathbf{m}$ with direction $\mathbf{n}$, where $\mathbf{n}$ is a unit vector perpendicular to the plane of the loop, the torque relation can be expressed in the form

$$\tau = \mathbf{m} \times \mathbf{B}, \qquad (32\text{-}15)$$

where

$$\mathbf{m} = \mathbf{n}m = \mathbf{n}IA \qquad (32\text{-}16)$$

is the magnetic dipole moment of the loop. The direction of $\mathbf{m}$ is determined by the right-hand rule for a loop or coil. Grasp the loop with the fingers of the right hand following the loop and pointing in the direction of the current flow, then the extended thumb points in the direction of the magnetic field due to current in the coil or in the direction of the magnetic dipole moment $\mathbf{m}$. The effect of the torque is to rotate the loop toward its equilibrium position with its magnetic moment $\mathbf{m}$ in the same direction as the field $\mathbf{B}$.

32-7 ELECTRIC METERS

Meters make use of the torque on a current carrying coil in a magnetic field. Figure 32-7 is a top view of a meter movement.

The pole pieces are shaped so that the field is parallel to the plane of the coil for any position of the coil. When the current passes through the coil,

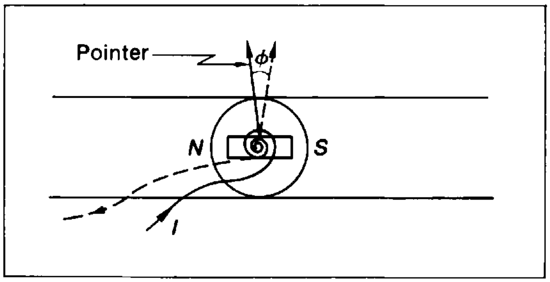


Figure 32-7 Schematic diagram of a meter movement.

the magnitude of the torque on the coil is given by the relation

$$\tau = NBIA, \qquad (32\text{-}17)$$

where N is the number of turns on the coil. The coil will turn until the torsional restoring torque due to the spring balances the magnetic torque τ. The restoring torque is proportional to the angle of deflection ϕ and can be put equal to $K\phi$, where K is the torsion constant of the spring. Therefore, the coil comes to rest when

$$K\phi = NBIA$$

or

$$\phi = \frac{NBIA}{K}. \qquad (32\text{-}18)$$

Since N, B, A, and K are all constants for a particular meter,

$$\phi = CI \qquad (32\text{-}19)$$

or the deflection of the pointer is proportional to the current. Galvanometers, which are instruments used to measure currents, work on the above principle. Therefore, Figure 32-7 also shows the rudiments of a galvanometer.

A voltmeter used to measure potential differences in a circuit is a galvanometer with a large number of turns on the moving coil so that there is a large deflection of the coil for small currents. It must be designed to draw as little current as possible so as not to affect significantly the current in the rest of the circuit. A high resistance is put in series with the coil as illustrated in Figure 32-8.

The current through the coil is given by

$$I = \frac{\Delta V}{R + r},$$

where r is the resistance of the coil and R is the

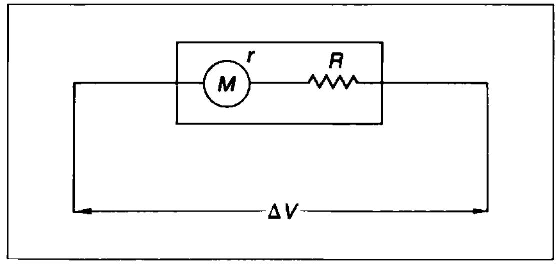


Figure 32-8 Internal connections of a voltmeter.

series resistance. Since the deflection is proportional to I, it is also proportional to ΔV.

The range of the meter can be modified by varying the resistance R. If we wish the meter to have a full scale reading ΔV_f with a current I_m through the meter movement, we must have a series resistance R so that

$$I_m = \frac{\Delta V_f}{r + R} \qquad (32\text{-}20)$$

or the series resistance R required is

$$R = \frac{\Delta V_f}{I_m} - r. \qquad (32\text{-}21)$$

An ammeter, which is used to measure currents in a circuit, is a galvanometer with a very low resistance, so as to cause a low potential drop across the instrument. This is achieved by putting a low resistance (R) in parallel or shunt with the coil of the meter as illustrated in Figure 32-9. Only a small fraction of the current goes through the coil so a coil with a large number of turns is necessary.

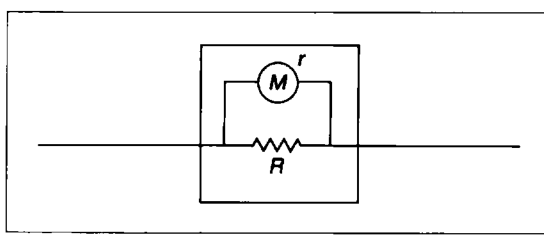


Figure 32-9 Internal connections of an ammeter.

As in the case of the voltmeter, the range of the meter can be modified by varying the resistance R. The shunt resistance R required to give a full-scale reading I_f with a current I_s through the shunt resistance and a current I_m through the meter movement is determined as follows: Since the same potential difference exists across the meter movement and the shunt resistance,

$$I_m r = I_s R,$$

where r is the resistance of the meter movement or coil. Also,

$$I_f = I_m + I_s.$$

Solving these two equations for R, gives

$$R = \frac{I_m r}{I_f - I_m}. \qquad (32\text{-}22)$$

The total resistance of the meter is the sum of R and r in parallel and is less than R, which is much less than the resistance r of the meter movement.

Example 2. The moving coil of a meter movement of height 3 cm and width 2 cm has 100 turns and is pivoted in a field of 5000 nt/amp-m.

(a) Calculate the torque when 2 milliamperes flow in the circuit.

(b) If the movement is controlled by springs that exert a restoring torque of 0.02 nt-m/deg, what is the steady deflection produced by a current of 2 milliamperes?

SOLUTION

(a) Assume that the field is parallel to the plane of the coil for any position of the coil. Apply Eq. (32-17),

$$\tau = NBIA = (100)(5000\text{ nt/amp-m})(2\times 10^{-3}\text{ amp}) \times (0.03\text{ m})(0.02\text{ m}) = 0.6\text{ nt-m}.$$

(b) Deflection produced by the current of 2 milliamperes is obtained by applying Eq. (32-19),

$$\phi = \frac{\tau}{C} = \frac{0.6\text{ nt-m}}{0.02\text{ nt-m/deg}} = 13\text{ deg}.$$

Example 3. A meter movement has a resistance of 100 ohms and takes 3 milliamperes for full scale deflection.

(a) Calculate the series resistance necessary if the meter is to act as a voltmeter having a full scale deflection of 6 volts.

(b) Calculate the shunt resistance necessary if the meter is to be used as an ammeter with a full scale deflection of 2 amperes.

SOLUTION

(a) Applying Eq. (32-21),

$$R = \frac{V_f}{I_m} - r = \frac{6\text{ volts}}{3\times 10^{-3}\text{ amp}} - 100\,\Omega = 1900\text{ ohms}.$$

(b) Applying Eq. (32-22),

$$R = \frac{I_m r}{I_f - I_m} = \frac{3\times 10^{-3}\text{ amp}\times 100\,\Omega}{(2 - 3\times 10^{-3})\text{ amp}} = 0.15\text{ ohms}.$$

PROBLEMS

1. What is the force on a conductor of length 15 m carrying a current of 5 amperes in a field of 4 nt/amp-m at right angles to the conductor?
2. Calculate the force on the conductor of Problem 1 if the conductor is
   (a) parallel to the field and
   (b) 30° to the direction of the field.
3. With 10 closely spaced turns carrying a current of 5 amperes, a square coil 40 cm × 40 cm is suspended vertically in an east-west plane in a region, where there is a uniform horizontal magnetic flux density **B**, of 1.5 nt/amp-m directed from west to east. What is the torque acting on this coil?
4. A uniform magnetic field of 2 nt/amp-m is directed vertically upward. A proton (charge = + 1.60×10^{-19} coul) is directed into this field at a speed of 3×10^{7} m/sec. Describe quantitatively how it will move if it is projected
   (a) eastward in a horizontal plane,
   (b) northward in a horizontal plane,
   (c) vertically downward, and
   (d) at an angle of 30° with the upward direction of **B**.

5. A single flat loop of area 25 cm^2 lies in the x-y plane, carrying a current of 5 amperes, clockwise as viewed from above. A uniform magnetic field of density 4 nt/amp-m is in the positive z-direction.
(a) What is the torque exerted by the magnetic field on the loop?
(b) What is the magnetic moment of the loop?

6. A long, straight wire carries a current of 1.5 amperes. An electron (charge = -1.60×10^{-19} coul) travels with a velocity of 5×10^6 cm/sec parallel to the wire, and in the same direction as the conventional current. What is the force (magnitude and direction) exerted on the electron by the magnetic field?

7. A circular loop of 0.2 m^2 area in the x-y plane carries a clockwise current of 4 amperes.
(a) What is the magnitude and direction of its magnetic moment?
(b) What is the torque on the loop due to a uniform magnetic field of 0.25 nt/amp-m parallel to the x-z plane and making an angle of 30° with the positive x-direction?

8. (a) What is the maximum torque on a rectangular galvanometer coil 5 cm × 12 cm of 600 turns when carrying a current of 10^{-5} amperes in a field of $\mathbf{B} = 0.10$ nt/amp-m?
(b) What is the torque on the coil when the normal to the plane of the coil makes an angle of 60° with the direction of **B**?

9. (a) What current must an infinitely long straight wire carry in order that $\mathbf{B} = 2.5 \times 10^{-4}$ nt/amp-m at a distance of 2 cm from it?
(b) What current must a circular coil of 100 turns and 10 cm radius carry in order that $\mathbf{B} = 0.001$ nt/amp-m at its center?

10. A long, straight wire carries a current of 100 amperes.
(a) What is the magnetic flux density at a point 5 cm from the wire?
(b) What is the force per unit length on another wire parallel to and 5 cm from it carrying a current of 10 amperes in the opposite direction?

11. A straight wire 25 cm long is at right angles to a uniform magnetic field of 0.7 nt/amp-m. What current must flow in the wire such that the force on it be 0.4 nt?

12. For the meter of Problem 13, calculate the shunt resistance necessary if the meter is to be used as an ammeter with a full scale deflection of 1 ampere.

13. A meter movement has a resistance of 50 ohms and takes 1.2 milliamperes for full scale deflection. Calculate the series resistance necessary if the meter is to act as a voltmeter having a full scale deflection of 3 volts.

14. The magnetic field density in a cyclotron, which is accelerating protons (charge = $+1.60 \times 10^{-19}$ coul, mass = 1.67×10^{-27} kg), is 2 nt/amp-m.
(a) Calculate the number of revolutions per second of the protons.
(b) If the maximum radius of the cyclotron is 0.30 m, what is the maximum kinetic energy of the proton?
(c) Through what potential difference would the proton have to be accelerated in order for it to have the same kinetic energy obtained as with the cyclotron?

15. What shunt resistance is required in a milliammeter with a movement resistance of 100 ohms giving full scale deflection for 1 milliampere in order for it to give full scale deflection for 60 milliamperes?

16. What is the velocity of a beam of electrons (charge on electron = -1.60×10^{-19} coul) when the simultaneous influence of an electric field of intensity 34×10^4 volts/m and a magnetic field of flux density 2×10^{-3} nt/amp-m, both fields being normal to the beam and to each other, produces no deflection of the electrons?

17. The resistance of a galvanometer is 30 ohms and the current for full scale deflection is 10 milliamperes.
(a) What shunt resistance is required to make it into an ammeter of 1 ampere full scale deflection?
(b) What series resistance is required to make it into a voltmeter of 150 volts full scale deflection?

18. How would you change a 10 volt voltmeter with an internal resistance of 1000 ohms to an ammeter of current carrying capacity 1 ampere with the help of a resistor? Compute the magnitude of the resistor.

19. In Figure 32-10, P and Q are cross sections of two long, parallel, straight wires perpendicular to the plane of the paper with P carrying a current of 20 amperes into the plane of the paper and Q carrying an equal current out of the plane of the paper.

Compute the magnetic flux density (**B**) at point R.

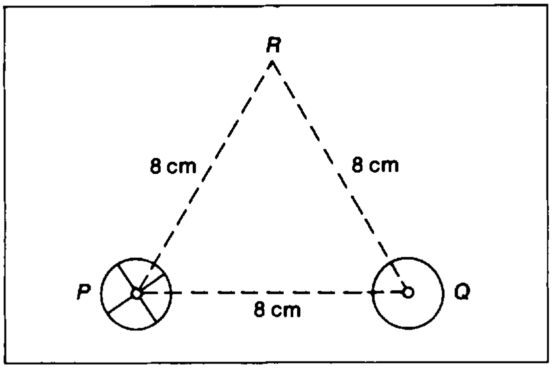


Figure 32-10

20. The earth's magnetic field at a certain location has a northerly component of $\mathbf{B}_N = 0.1 \times 10^{-4}$ nt/amp-m and an upward component of $\mathbf{B}_V = 0.4 \times 10^{-4}$ nt/amp-m. Calculate the magnitude and direction of the force a straight 1 m long wire would experience, if it were oriented parallel to the south-north direction and was carrying a current of 10 amperes toward the north.

33 Magnetic Field of a Current

In this chapter, we shall discuss magnetic fields and energy stored in a magnetic field along with other topics. In addition, we shall point out the strong analogy with electric fields.

33-1 MAGNETIC FLUX

In Chapter 32, we mentioned briefly the terms magnetic flux and magnetic flux density. However, more should be said about both these terms. The total number of magnetic field lines that cut any area A as illustrated in Figure 33-1 is the magnetic flux (ϕ) through the area, that is,

$$\phi = \oint_A \mathbf{B} \cdot \mathbf{n}\, dA, \qquad (33\text{-}1)$$

where $\mathbf{n}$ is a unit vector perpendicular to A.

The unit of ϕ in the MKS system of units is the weber, which is equal to a nt-m/amp, and the unit of the flux density $\mathbf{B}$ is webers/m^2.

Experimentally, when we plot the magnetic field about a wire as illustrated in Figure 32-2, we see that the magnetic flux lines are closed lines. This implies that if a closed surface A is taken in the field, every line entering the surface must leave it. This means that if we calculate the net flux through the surface by integrating over the surface, the value of the integral must be zero.

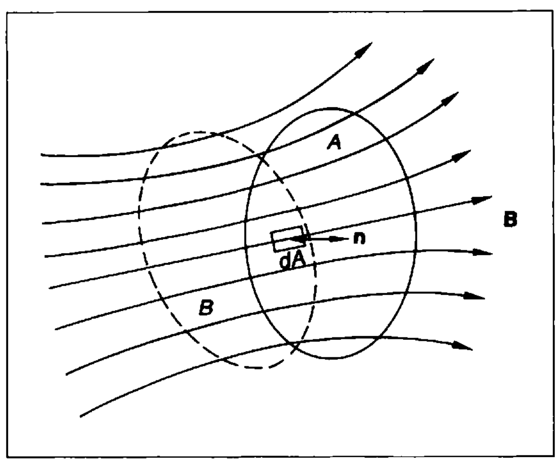


Figure 33-1 Magnetic flux through an area *A*.

$$\phi_{net} = \oint_A \mathbf{B} \cdot \mathbf{n}\, dA = 0.$$

33-2 AMPERE'S LAW

In this section, we will discuss the relationship between the line integral of $\mathbf{B}$ around any closed path and the current I enclosed by the path. In the example given in Chapter 32, we saw that the magnitude of **B** at a point a distance R from an infinitely long straight wire was given by

$$B = \frac{\mu_0 I}{2\pi R},$$

and that the direction of **B** was perpendicular to the wire. Now let us calculate the integral of **B** along a circular arc of radius R about a current carrying wire as illustrated in Figure 33-2.

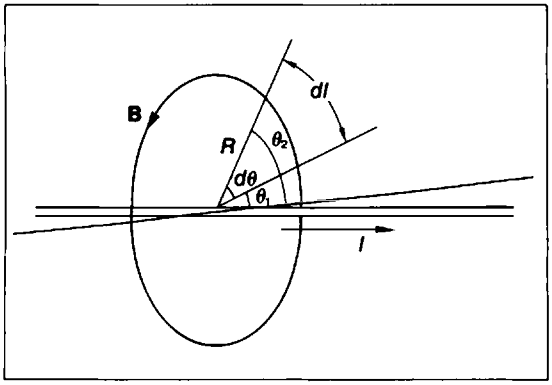


Figure 33-2 A circular path for integrating **B** around a wire. The current links the path of integration.

Referring to Figure 33-2,

$$dl = R d\theta$$

$$\int_{\theta_1}^{\theta_2} \mathbf{B} \cdot d\mathbf{l} = \mu_0 \frac{I}{2\pi} \int_{\theta_1}^{\theta_2} \frac{1}{R} R d\theta = \mu_0 I \frac{(\theta_2 - \theta_1)}{2\pi},$$

which is independent of R or

$$\mathbf{B} \cdot d\mathbf{l} = \mu_0 I \frac{d\theta}{2\pi}.$$

Integrating completely around the circle gives

$$\oint \mathbf{B} \cdot d\mathbf{l} = \mu_0 I. \qquad (33\text{-}2)$$

This equation states that the integral of $\mathbf{B} \cdot d\mathbf{l}$ around any closed path is equal to μ_0 times the current passing through any area bounded by the path of integration. This equation is called Ampere's law. If there is no enclosed current,

$$\oint \mathbf{B} \cdot d\mathbf{l} = 0. \qquad (33\text{-}3)$$

It should be pointed out that the above derivation is not a general proof of Ampere's law. A general proof is beyond the scope of this book. The student will have to accept here that the law is true in general for any magnetic field shape, for any assembly of currents, and for any path of integration.

33-3 ELECTROMAGNETIC INDUCTION; FARADAY'S LAW AND LENZ'S LAW

Faraday showed experimentally that when the magnetic flux through a closed circuit is changing, there is an emf induced in the circuit that is directly proportional to the time rate of change of the flux.

In analytical form, Faraday's law states that

$$\mathscr{E} = -\frac{d\phi}{dt}, \qquad (33\text{-}4)$$

where $\mathscr{E}$ is in volts, ϕ is in webers, and t is in seconds. The minus sign is an indication of the direction of the induced emf. The direction of the induced emf is a consequence of the principle of conservation of energy. The induced current always flows in such a direction that its magnetic field opposes the change in conditions giving rise to the induced current. This statement is known as Lenz's law. The induced current is set up by the induced emf.

We shall now consider some experimental verifications of both Faraday's law and Lenz's law. First, we shall consider a current in a coil of wire induced by a moving magnet, as illustrated in Figure 33-3. If the north pole of the magnet is pushed into the coil, the induced current will flow in such a direction so as to set up a magnetic field that opposes the insertion of the magnet, as illustrated in Figure 33-3(a). If the magnet is pulled out of the coil, the induced current flows in the opposite direction, or in such a direction that its magnetic field opposes the withdrawal of the magnet, as illustrated in Figure 33-3(b). The magnitude of the induced current will be proportional to the speed with which the magnet is pushed into or pulled out of the coil.

Let us consider for the moment the reason for the above experimental facts based on energy conservation. The coil has a resistance and the induced current produces heat, which requires a source of energy. The only other source of energy is the mechanical work done on the magnet in pushing it into the coil. In order for work to be done on the magnet, it must experience an opposing force as it enters the coil. This means that the field produced by the current in the coil must be in such a direction

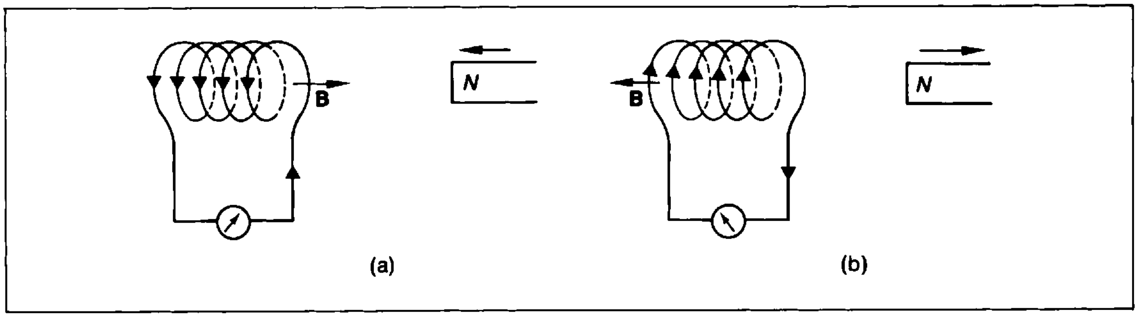


Figure 33-3 Currents induced by moving magnets.

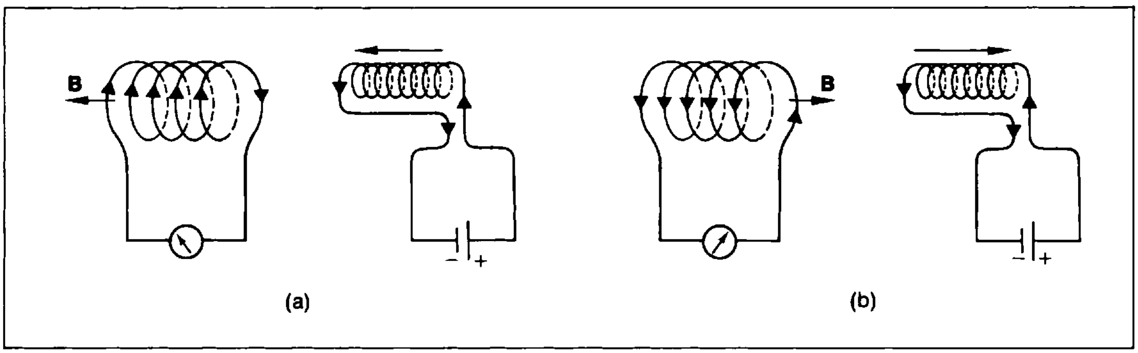


Figure 33-4 Currents induced by currents.

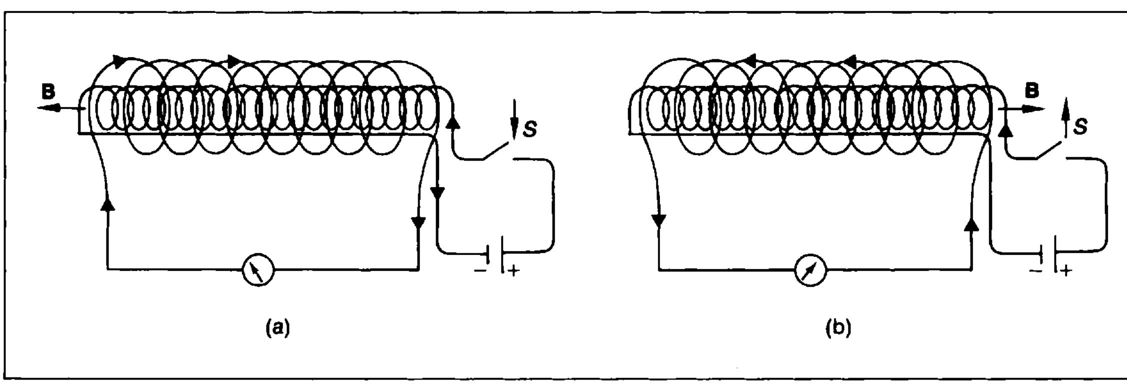


Figure 33-5 Currents induced by closing and breaking the circuit.

as to repel the magnet. If the magnet were attracted by the field in the coil, mechanical energy and heat energy would be produced by the system without any work being done on it, which is a contradiction of the principle of conservation of energy. A similar argument can be presented for the direction of the induced current when the magnet is withdrawn from the coil.

If we insert or withdraw a current carrying coil instead of a magnet into another coil, an induced current will flow in the other coil. The magnitude and direction of the induced current will be governed by Faraday's and Lenz's laws and the effects are the same as with the magnet. This is illustrated in Figure 33-4.

Let us consider what happens if, instead of inserting and withdrawing the current carrying coil into the larger coil, we leave it inside the larger coil and open and close the circuit leading to the inside coil as illustrated in Figure 33-5. It is then observed that the induced current flows in the opposite direction to the current producing it when the circuit is closed, and in the same direction when the circuit is broken. Why?

33-4 MUTUAL INDUCTANCE AND SELF-INDUCTANCE

Let us next consider two coils arranged so that part of the flux produced by a current in the first (primary) coil threads the other (secondary) coil. What emf will be induced in the secondary coil by this flux linkage with the primary coil? (If the current in the primary coil is varied, an induced emf is set up in the secondary coil.) In general, the emf induced in the secondary coil is directly proportional to the time rate of change of the current in the primary coil. Symbolically,

$$\mathscr{E}_s \alpha \frac{dI_p}{dt}.$$

We can write this as an equation by introducing a proportionality constant

$$\mathscr{E}_s = -M\frac{dI_p}{dt}, \tag{33-5}$$

where M is the coefficient of mutual inductance of the two coils. The minus sign indicates that the emf in the secondary coil is opposite in direction to the emf in the primary coil.

Two coils have a mutual inductance of 1 henry if a change in current of 1 ampere per second in the primary coil induces an emf of 1 volt in the secondary coil.

The value of $\mathscr{E}_s$ is also given by Eq. (33-4). Equating Eqs. (33-4) and (33-5) gives

$$\mathscr{E}_s = -M\frac{dI_p}{dt} = -\frac{d\phi_s}{dt}$$

or

$$M = \frac{d\phi_s}{dI_p}, \tag{33-6}$$

where $d\phi_s$ is the total flux linkage between the two coils produced by the current change dI_p in the primary coil.

An induced emf is also set up in single coil by a current change in the coil. The induced emf opposes the change in current and is sometimes called a "back emf" ($\mathscr{E}_b$). It is proportional to the time rate of change of the current, which permits us to write the following equation:

$$\mathscr{E}_b = -L\frac{dI}{dt}, \tag{33-7}$$

where L is the coefficient of self-inductance of the coil and is expressed in henrys.

A coil has a coefficient of self-inductance of 1 henry if a change in the current of 1 ampere per second induces a back emf of 1 volt in the coil.

The value of $\mathscr{E}_b$ is also given by Eq. (33-4). Equating Eqs. (33-4) and (33-7) gives

$$L = \frac{d\phi}{dI}, \tag{33-8}$$

where $d\phi$ is the total change in flux and dI is the total change in current.

33-5 ENERGY AND THE MAGNETIC FIELD

When we build up a current in an inductor, an external source of emf must do work against the induced emf. This suggests the possibility of energy storage in the inductor. Neglecting resistance, the rate at which the external source does work is

$$P = -\mathscr{E}_b I = LI\frac{dI}{dt}.$$

The total work done in building the current up to a value of I amperes in time t is

$$W = \int_0^t P\,dt = L\int_0^I I\,dI$$

or

$$W = \frac{1}{2}LI^2. \tag{33-9}$$

As an example, consider a solenoid as illustrated in Figure 33-6. At a point inside the solenoid and sufficiently far from the ends, the magnetic field is parallel to the axis and uniform. At a point outside the field is very weak. Let us calculate the line integral of the induction around the rectangular path shown in Figure 33-6. Over the two radial sides, the path is perpendicular to the field, so their contribution is zero. The same will be true for that part of the field that is parallel to the axis, outside, since the field there is very weak. The only contribution then comes from that part of the path of length d parallel to the axis and inside the solenoid, where

$$\oint \mathbf{B}\cdot d\mathbf{l} = Bd = \mu_0 n dI$$

or

$$B = \mu_0 nI, \tag{33-10}$$

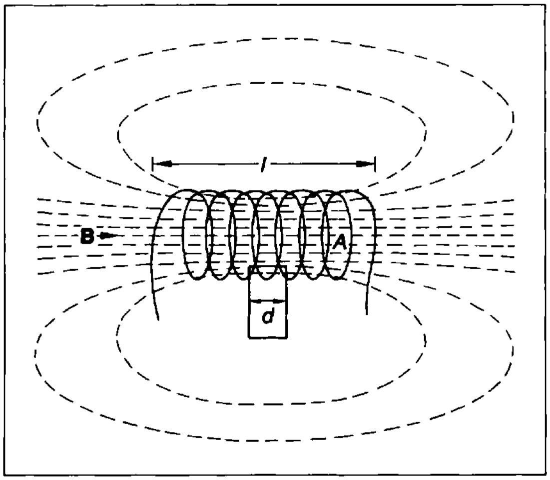


Figure 33-6 Magnetic field in a solenoid.

where n is the number of turns of wire per unit length on the solenoid. The induced emf in the solenoid is

$$\mathscr{E}_b = -L\frac{dI}{dt} = -nl\frac{d\phi}{dt}, \quad (33\text{-}11)$$

where l is the length of the solenoid. The magnetic flux through the coil from Eq. (33-10) is

$$\phi = \oint \mathbf{B}\cdot\mathbf{n}\,dA = BA = \mu_0 nIA, \quad (33\text{-}12)$$

where A is the cross-sectional area of the solenoid. From Eq. (33-11), the induced emf is

$$\mathscr{E}_b = -\mu_0 n^2 lA\frac{dI}{dt}. \quad (33\text{-}13)$$

Comparing Eqs. (33-11) and (33-13), we obtain a relation for the self-inductance of the solenoid

$$-L\frac{dI}{dt} = -\mu_0 n^2 lA\frac{dI}{dt}$$

or

$$L = \mu_0 n^2 lA. \quad (33\text{-}14)$$

If we substitute Eq. (33-14) into Eq. (33-9), we get the energy stored in the solenoid to be

$$W = \frac{1}{2}\mu_0 n^2 lAI^2. \quad (33\text{-}15)$$

From this, we can conclude that this energy is stored in the magnetic field. It should be pointed out that this is not a proof. However, all experimental work supports it. Since the quantity lA is the volume of the solenoid, the energy density or the energy per unit volume stored in the solenoid is

$$\frac{W}{v} = \frac{1}{2}\mu_0 n^2 I^2.$$

Since

$$B = \mu_0 nI,$$

from Eq. (33-10),

$$\frac{W}{v} = \frac{1}{2}\frac{B^2}{\mu_0}. \quad (33\text{-}16)$$

We may, therefore, state that the energy density stored in a magnetic field is given by Eq. (33-16). This is the analogue of the expression for the energy density in an electric field of a capacitor $\frac{1}{2}\epsilon_0 E^2$ with free space as a dielectric.

33-6 GENERATORS AND MOTORS

Generators and motors both make use of a coil called the armature, rotating in a magnetic field. The magnetic field of a generator is usually created by electromagnets, which consist of soft iron wound with many turns of wire called the field coils through which pass an electric current. A generator is a machine which, when driven mechanically, generates an electromotive force, while a motor is a machine that does mechanical work at the expense of electrical energy.

We shall first discuss the operation of two principle kinds of generators. These two kinds are: alternating current (AC) and direct current (DC). If a single coil is placed between the north and south poles of two magnets and rotated, an induced emf is set up in the coil as illustrated in Figure 33-7. The emf and, consequently, the current changes direction every half revolution, and an alternating current is generated. Figure 33-8 is a graph of the induced emf versus position of the coil as determined in Section 32-6. If an alternating current is desired in the external circuit, the current is removed from the coil by collecting rings and brushes, as illustrated in Figure 33-9, where B_1 and B_2 are the brushes and C_1 and C_2 are the collecting rings. The armature of a practical alternating current generator consists of a large number of coils. If a direct current is desired in the external circuit, a commutator and brushes are required to remove the current, as illustrated in Figure 33-10. The commutator C is a divided ring. The segments are insulated from one another, and the brushes B_1 and B_2 are so arranged that they slip from one segment to

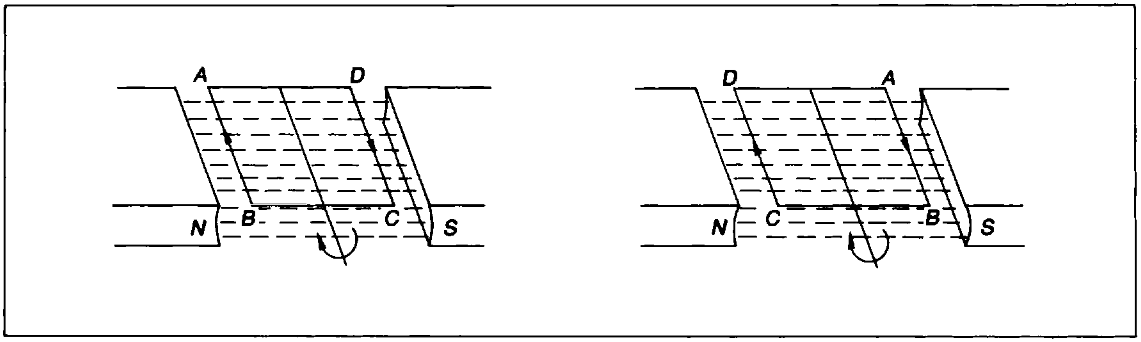


Figure 33-7 Rotating coil in a magnetic field.

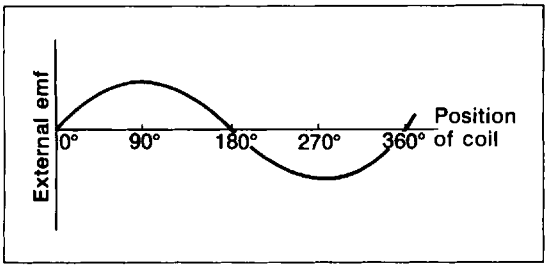


Figure 33-8 External emf versus position of coil for a coil rotating in a magnetic field.

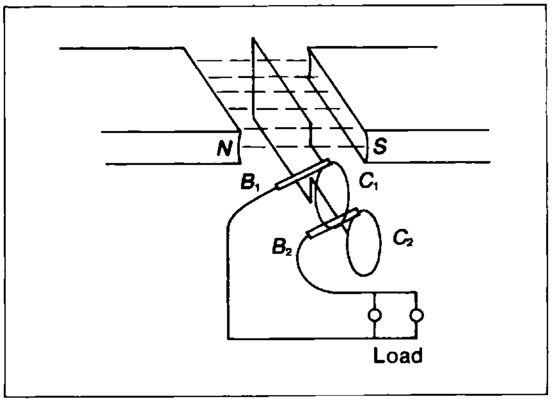


Figure 33-9 Simple alternating current generator.

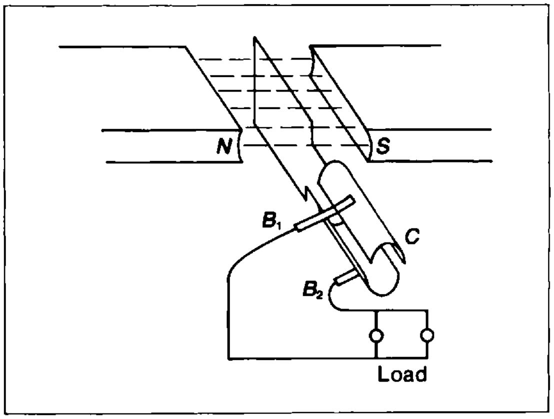


Figure 33-10 Simple direct current generator.

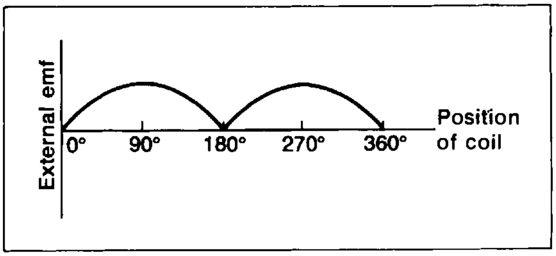


Figure 33-11 External emf versus position of coil for a single coil rotating in a magnetic field when a commutator is used.

the other at the instant the current in the coil of the armature changes direction. Figure 33-11 is a graph of the emf in the external circuit versus position of the coil when such a commutator is used.

To obtain a practically steady direct current from a direct current generator, a large number of coils are placed on the armature, and the commutator is divided into a correspondingly large number of segments. Figure 33-12 is a graph of the emf versus position of the armature for a direct current generator with a number of coils on the armature.

We shall now briefly discuss electric motors. An electric motor is the reverse of an electric generator. Any direct current generator can serve as a direct current motor because if a current is passed through the armature from an external source there

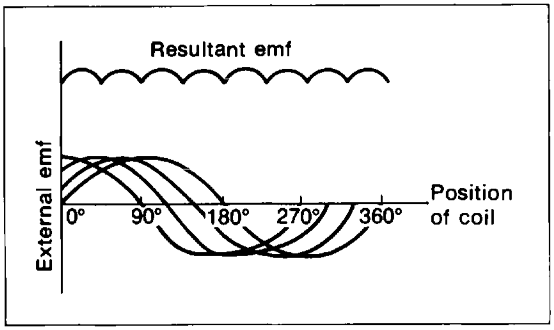


Figure 33-12 External emf for a four coil armature with a commutator.

will be a force on each armature coil and the coil will start to turn. Some motors, called universal motors, will run either on alternating or direct current. In order to have a motor run on alternating current, the armature and field currents must be in phase with each other so that they reverse at the same instant.

In a motor, in addition to the applied emf, we have an induced or back emf when the armature rotates, since the coils are cutting across lines of magnetic flux. This back emf opposes the impressed or applied emf. If V is the external applied voltage, and $\mathscr{E}_b$ the back emf, then the resultant voltage available to cause a current in the armature is $V - \mathscr{E}_b$. Therefore,

$$V - \mathscr{E}_b = IR \quad \text{or} \quad V = \mathscr{E}_b + IR, \qquad (33\text{-}17)$$

where I is the current in the armature, and R is the resistance of the armature coils, or the sum of the resistance of the armature and field coils, depending upon the kind of motor.

33-7 SERIES CIRCUIT WITH RESISTANCE AND INDUCTANCE

We must now discuss the effects that the insertion of an inductor in a circuit entails. Consider an inductor of inductance L in series with a resistor of resistance R connected to a battery of voltage V.

$$V = IR + L\frac{dI}{dt} \qquad (33\text{-}18)$$

or

$$\frac{L\,dI}{V - IR} = dt. \qquad (33\text{-}19)$$

Integrating both sides, we obtain

$$-\frac{1}{R}\ln(V - IR) = \frac{t}{L} + \text{constant}. \qquad (33\text{-}20)$$

Let us assume that there is no current flowing at $t = 0$. We then obtain the constant of integration, and Eq. (33-20) becomes

$$\ln(V - IR) - \ln V = -\frac{Rt}{L},$$

or

$$\ln\frac{V - IR}{V} = -\frac{Rt}{L}. \qquad (33\text{-}21)$$

If we now take the antilogarithm of both sides of Eq. (33-21), we obtain

$$1 - \frac{IR}{V} = e^{-Rt/L}$$

or

$$I = \frac{V}{R}(1 - e^{-Rt/L}). \qquad (33\text{-}22)$$

Equation (33-22) shows that the current increases in the same way as the charge on a capacitor in the R-C circuit. The quantity L/R is called the time constant (τ) of the inductor, and Eq. (33-22) may be written

$$I = \frac{V}{R}(1 - e^{-t/\tau}). \qquad (33\text{-}23)$$

The larger the value of τ, the slower is the build up of current in the inductor.

If, after a steady current is reached in the inductor, the battery or inductor is shorted out, then $V = 0$, and Eq. (33-19) becomes

$$-\frac{L\,dI}{IR} = dt, \qquad (33\text{-}24)$$

which can be solved in a similar manner to that above to give

$$I = \frac{V}{R}e^{-Rt/L}. \qquad (33\text{-}25)$$

Equation (33-25) shows that the current decreases exponentially to zero, and that this decay is governed by the same time constant $\tau = L/R$.

Example 1.

(a) Find the flux through a coil of 10 turns of average diameter 2 cm when placed in a field of 1000 webers/m^2.

(b) If the field strength is reduced to 200 webers/m^2 in 1 second, what is the induced emf?

SOLUTION

(a) The total flux is

$$\phi = BA = (1000\ \text{webers/m}^2)(\pi)((0.01)^2\text{m}^2) = 0.314\ \text{webers}.$$

(b) The induced emf is

$$\mathscr{E} = N\frac{\Delta\phi}{\Delta t} = N\frac{\Delta BA}{\Delta t} = \frac{(10)((1000-200)\ \text{webers/m}^2)(\pi(0.01)^2\text{m}^2)}{1\ \text{sec}} = 2.51\ \text{volts}.$$

Example 2. A long solenoid of cross-sectional area 10 cm^2 is wound with 1000 turns per meter. The windings carry a current of 0.50 amperes. A secondary winding of 10 turns encircles the solenoid. When a switch in the circuit providing the current to the solenoid is opened, the magnetic field of the solenoid becomes zero in 0.10 seconds. What is the average induced emf in the secondary?

SOLUTION

The change in field through the solenoid when the switch is opened is

$$\Delta B = \mu_0 n I - 0 = (4\pi\times 10^{-7}\ \text{nt/amp}^2)(1000\ \text{m}^{-1})(0.50\ \text{amp}) = 6.28\times 10^{-4}\ \text{webers/m}^2.$$

The induced emf in the secondary is

$$\mathscr{E} = N_s A\frac{\Delta B}{\Delta t} = (10)(10^{-3}\ \text{m}^2)\left(\frac{6.28\times 10^{-4}\ \text{webers/m}^2}{(0.10\ \text{sec})}\right) = 6.28\times 10^{-5}\ \text{volts}.$$

Example 3. An inductor of inductance 2 henrys, and resistance 20 ohms, is connected to a 6 volt battery of zero internal resistance. Find:

(a) the initial rate of current increase in the circuit.

(b) the rate of increase of current at the instant when the current is 0.2 amperes,

(c) the current 0.1 seconds after the circuit is closed, and

(d) the final constant current.

SOLUTION

(a) Apply Eq. (33-18) in the form

$$\frac{dI}{dt} = \frac{V}{L} - \frac{IR}{L}.$$

At

$$t = 0, \quad I = 0,$$

and

$$\frac{dI}{dt} = \frac{V}{L} = \frac{6\ \text{volts}}{2\ \text{henrys}} = 3\ \text{amp/sec}.$$

(b) At $I = 0.2$ amperes,

$$\frac{dI}{dt} = \frac{V}{L} - \frac{IR}{L} = \left(\frac{6\ \text{volts}}{2\ \text{henrys}}\right) - \left[\frac{(0.2\ \text{amp})(20\ \text{ohms})}{2\ \text{henrys}}\right], = 1\ \text{amp/sec}.$$

(c) The current 0.1 seconds after the switch is closed is obtained by applying Eq. (33-22).

$$I = \frac{V}{R}(1 - e^{-Rt/L}) = \frac{6\ \text{volts}}{2\ \text{ohms}}(1 - e^{-[(20\ \text{ohms})(0.1\ \text{sec})/2\ \text{henrys}]}) = 3\left(1 - \frac{1}{e}\right)\text{amp} = 1.90\ \text{amp}.$$

(d) The final constant current is obtained by substituting $t = \infty$ into Eq. (33-22), which gives

$$I = \frac{V}{R} = \frac{6\ \text{volts}}{2\ \text{ohms}} = 3\ \text{amp}.$$

PROBLEMS

1. A proton of mass $= 1.67\times 10^{-27}$ kg, charge $= 1.60\times 10^{-19}$ coul, moves in a circular path of 12 cm radius perpendicular to a uniform magnetic field. The velocity of the proton is 1×10^7 m/sec. What is the total magnetic flux encircled by the circular path of the proton?
2. Derive an expression for the magnetic flux density inside a long, straight conductor of circular cross section, having a radius R. The current is uniformly distributed over the cross-sectional area of the conductor so that the current density $j = I/\pi R^2$.
3. (a) Derive an equation for **B** at the center of a circular coil of radius a.
   (b) What is the value of **B** at the center of a circular coil of radius 4 cm containing 20 turns and carrying a current of 10 amperes?

4. A coil of 5 turns has dimensions of 7 cm × 9 cm. It rotates at an angular velocity of 30π rad/sec in a field of 0.4 webers/m^2. What is the maximum emf induced in the coil if the axis of rotation is perpendicular to the field?

5. Two coils placed close together have a mutual inductance of 25 millihenrys. The current in coil A is increasing at the rate of 2 amp/sec. What is the emf generated in coil B by this changing current?

6. A particle of mass 1 gm and charge 5 microcoulombs is moving in a circle in a solenoid of 50 turns and length 10 cm carrying a current of 10 amperes. The radius of the circle is 10 cm. Find the linear speed.

7. How much energy is stored in the magnetic field of a coil whose self-inductance is 10 henrys, when it is carrying a steady current of 4 amperes?

8. A flat square coil of 10 turns has sides of length 12 cm. The coil rotates in a magnetic field whose flux density is 0.025 webers/m^2.
   (a) What is the angular velocity of the coil if the maximum emf produced is 20 millivolts?
   (b) What is the average emf at this velocity?

9. What is the time constant of an R-L circuit containing a pure resistance of 200 ohms in series with a 5 henry inductor whose resistance is 100 ohms?

10. A 500 ohm resistance and a 5 henry inductance are connected in series with a 20 volt battery.
    (a) What is the value of the time constant of this circuit?
    (b) How much energy is stored in the inductor after the switch has been closed for a period of time equal to the time constant?

11. An inductor of inductance L and a resistor of resistance R are connected in series with a battery that has set up a steady current I_0 in the circuit. At time $t = 0$, the battery is shorted out so that the series circuit consists of only an inductor and the resistor. Set up the differential equation relating L, R, I, and t, and integrate this equation to obtain the current through the resistor as a function of time.

12. An electron ($q = 1.60 \times 10^{-19}$ coul) circulates about a proton in an orbit of 0.52×10^{-10} m radius at a speed of 2.5×10^6 m/sec.
    (a) To what current is this equivalent?
    (b) What is the magnetic field at the nucleus due to the orbital current?

13. (a) Show that self-inductance over resistance (L/R) and resistance times capacitance (RC) both have the dimensions of time.
    (b) Show that 1 weber per second equals 1 volt.

14. A closely wound rectangular coil of 50 turns has dimensions of 12 cm × 25 cm. The plane of the coil is rotated from a position where it makes an angle of 45° with a magnetic field of flux density 2 webers/m^2 to a position perpendicular to the field in time $t = 0.1$ seconds. What is the average emf induced in the coil?

15. (a) If the current increases uniformly from 0 to 2 amperes in 0.40 seconds, in a coil whose coefficient of self-induction is 0.60 henrys, what is the magnitude of the emf produced?
    (b) With the aid of a diagram, specify its direction.

16. A uniform field of magnetic flux density $\mathbf{B}$ is changing in magnitude at a constant rate $d\mathbf{B}/dt$. Given a mass m of conducting material (density d, resistivity ρ) drawn into a wire of radius r and formed into a circular loop of radius R, show that the induced current in the loop when placed normally in the above field does not depend upon the size of the wire or the loop, and is given by

$$I = \left(\frac{m}{4\pi\rho d}\right)\left(\frac{d\mathbf{B}}{dt}\right).$$

17. In Figure 33-13, V is 120 volts, L is 120 henrys, and R is 60 ohms.
   (a) To what value does the current build up after the switch is closed?
   (b) How much energy will be stored in the inductor when the current has reached its final value?
   (c) Calculate the current and the rate of change of the current when the potential difference across the resistor is 50 volts.

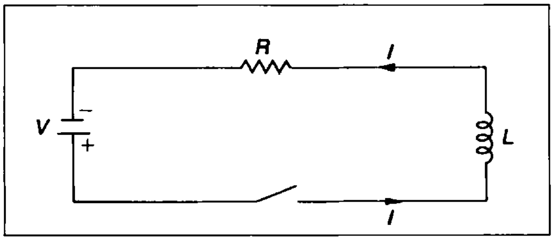


Figure 33-13

18. In Figure 33-14, $V = 6$ volts, $R_1 = 0.1$ ohms, $R_2 = 0.9$ ohms, and $L = 1$ henry. Find the potential across R_2 after switch S has been closed for 0.4 seconds. What is the power dissipated in the circuit 0.4 seconds after switch S has been closed?

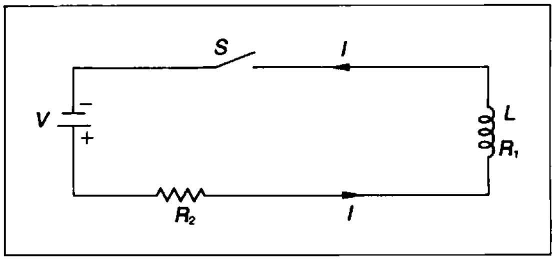


Figure 33-14

19. An inductor of inductance 4 henrys and resistance 4 ohms is connected to a battery of terminal voltage 12 volts. Find
   (a) the initial rate of increase of current in the circuit,
   (b) the rate of increase of current at the instant when the current is 1 ampere,
   (c) the current 0.4 seconds after the circuit is closed, and
   (d) the final steady current.
20. Derive the expression

$$B = \frac{RQ}{NA}$$

for a search coil used with a galvanometer to measure a magnetic field, where R is the total resistance of the coil and galvanometer, Q is the total charge passing through R, N is the number of turns of wire in the coil, and A is the area of the coil. The operation of a search coil is as follows: Originally the coil is placed in the magnetic field with the plane of the coil perpendicular to the field. It is then quickly removed from the field region. This induces a current in the coil which flows through the galvanometer attached to the coil.

34 Magnetic and Thermoelectric Properties of Matter

34-1 ORIGINS OF MAGNETISM

The earliest guess as to the origin of magnetism in magnets was that the material consisted of a number of particles each being a small magnet. These small magnets are oriented at random when the material is unmagnetized. When the material is magnetized they are aligned with unlike poles adjacent to one another within the material, thus leaving unlike poles at each end of the material, as illustrated in Figure 34-1.

After the discovery that a small coil carrying a current behaves like a small magnet, Ampere suggested the following: (1) magnetism is due to small circulating currents associated with each atom; (2) each atom possesses a magnetic dipole moment; and (3) the magnetic moment of any substance is equal to the vector sum of the magnetic dipole moments of its constituent atoms.

The magnetic moment of a current loop enclosing an area A is defined by

$$\mathbf{m} = \mathbf{n}IA, \tag{34-1}$$

where $\mathbf{n}$ is a unit vector in the direction of $\mathbf{m}$, which is normal to the plane of the loop and is in the direction a right-hand screw projects if turned in the direction of the current. Let us now list some of the properties of the magnetic moment $\mathbf{m}$ of the current associated with the orbiting electrons in an atom, which have been verified experimentally.

(1) The resultant magnetic moment of an atom is related to the vector sum of the individual magnetic moments of its electrons.

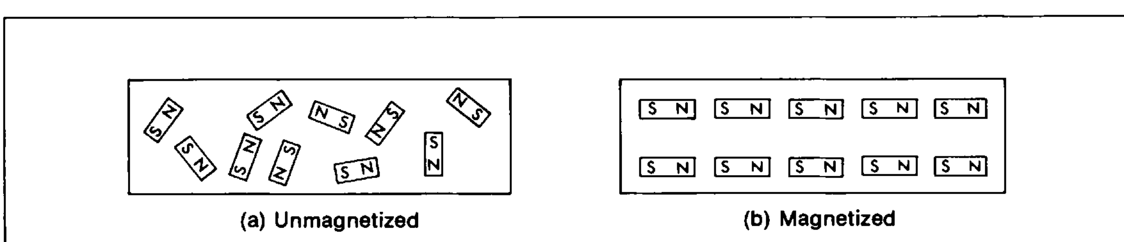


Figure 34-1 Materials in unmagnetized and magnetized states.

(2) An atom, ion, or molecule has zero permanent magnetic moment if the total angular momentum of the electrons is zero. If the total angular momentum is not zero, the particle has a permanent magnetic moment.
(3) Most free atoms have a permanent magnetic moment, while molecules nearly always have a zero permanent magnetic moment.

It should be clear to the reader that the above properties of the magnetic moment indeed support Ampere's suggestions.

34-2 MAGNETIZATION, MAGNETIC INTENSITY, SUSCEPTIBILITY, AND PERMEABILITY

If we place a material substance within a coil carrying a current, it is quite likely that the value of **B** will be different than when the substance is not present. It is possible to consider the resultant **B** within the coil material as the sum of two quantities $\mathbf{B}_0$ due to the current in the coil and a quantity $\mu_0\mathbf{M}$ from the magnetic properties of the substance in the coil. We may write

$$\mathbf{B} = \mathbf{B}_0 + \mu_0\mathbf{M}, \tag{34-2}$$

where μ_0 is referred to as the permeability of free space and **M** is called the magnetization of the substance. The magnetization **M** is the magnetic moment per unit volume and is analogous to the electric polarization **P**. Also, it is convenient to define another vector **H** called the magnetizing force or magnetic intensity just as in electrostatics it was convenient to define the vector **D**. **H** is defined by the relationship

$$\mathbf{H} = \frac{\mathbf{B}_0}{\mu_0}. \tag{34-3}$$

This allows us to write Ampere's law [Eq. (33-2)] as

$$\oint \mathbf{H} \cdot d\mathbf{l} = NI, \tag{34-4}$$

for any substance placed within a coil of N turns. In MKS units, **H** is measured in ampere turns per meter.

In an isotropic substance, **M** is directly proportional to **H** and parallel to it, so that we may write

$$\mathbf{M} = \chi_m\mathbf{H}, \tag{34-5}$$

where χ_m is known as the magnetic susceptibility and is a dimensionless quantity. If we substitute Eq. (34-3) into Eq. (34-2) for B_0, we get

$$\mathbf{B} = \mu_0(\mathbf{H}+\mathbf{M}) = \mu_0\left(1+\frac{\mathbf{M}}{\mathbf{H}}\right)\mathbf{H}. \tag{34-6}$$

Substituting Eq. (34-5) into Eq. (34-6), gives

$$\mathbf{B} = \mu_0(1+\chi_m)\mathbf{H} \tag{34-7}$$

or

$$\mathbf{B} = \mu\mathbf{H}, \tag{34-8}$$

where

$$\mu = \mu_0(1+\chi_m). \tag{34-9}$$

The ratio μ/μ_0 is called the relative permeability μ_r. It follows from Eq. (34-9) that

$$\mu_r = 1+\chi_m \quad \text{or} \quad \chi_m = \mu_r - 1. \tag{34-10}$$

The susceptibility of a vacuum equals zero, since $\mu_r = 1$ in a vacuum.

Since **M** can be interpreted as the magnetic moment per unit volume, χ_m is a quantity whose interpretation is in terms of a unit volume of the material. For comparison between experiment and theory, it is convenient to define mass susceptibility or specific susceptibility χ_{ms},

$$\chi_{ms} = \frac{\chi_m}{\rho}, \tag{34-11}$$

where ρ is the density of the material. Another useful quantity is obtained by referring χ_m to a particular number of molecules of the material. We therefore define the term molar susceptibility by the relation

$$\chi_{m\ \text{mole}} = \frac{M}{\rho}\chi_m, \tag{34-12}$$

where M is the molecular mass of the material.

Real materials are usually characterized by the sign of their susceptibility. In the next three sections we discuss the three types of magnetic materials.

34-3 DIAMAGNETISM

Diamagnetic substances are those for which χ_m is negative and μ_r is less than unity. In diamagnetic substances, χ_m is practically independent of temperature and field strength, and is approximately constant for a particular type of atom or ion.

The explanation of diamagnetism follows from Faraday's law and Lenz's law of electromagnetic induction. Atoms can be visualized as consisting of electrons moving in orbits around the nucleus. This orbital motion of the electrons constitutes a tiny current loop. For certain atoms, those which make up diamagnetic materials, the net magnetic moment for the orbital motion of the electrons is zero. If a magnetic field is applied to the substance, each electron experiences a force, and its orbital motion is altered so that it acquires an angular momentum and hence a magnetic moment. The change in the orbits will produce a net flux for the atom, which will oppose the flux due to $\mathbf{B}_0$. Thus, the net flux through the material will be less than if the material were not present, causing the resultant $\mathbf{B}$ to be less than $\mathbf{B}_0$, or for the material to have a negative χ_m.

Diamagnetism is present in all substances, but its presence is masked in substances where the atoms have a net magnetic moment. Superconductors, which are substances that have zero electrical resistance at very low temperatures, are highly diamagnetic when in the superconducting state. When a superconductor is placed in a magnetic field low enough not to destroy its superconductivity, currents are produced in the surface layers and the magnetic field in the interior of the sample remains zero. Thus, we may say that superconductors are perfectly diamagnetic.

34-4 PARAMAGNETISM

Paramagnetic substances are those for which χ_m is positive and μ_r is slightly greater than unity. Paramagnetism occurs in substances that possess a permanent magnetic moment. In the absence of an external magnetic field, the atomic dipoles point in random directions, and there is no resultant magnetization. When a field is applied to the substance, the atomic dipoles orient themselves parallel to the field. This gives a net magnetization parallel to the field and a positive contribution to the susceptibility. This positive contribution is affected by temperature, since thermal agitations affect the randomness of the orientations of the dipoles. At a low temperature, there is less thermal agitation, thus giving less random orientation, which permits greater magnetization for a given field. The temperature dependence of χ_m for paramagnetic substances is given by an empirical relation known as Curie's law,

$$\chi_m = \frac{C}{T}, \tag{34-13}$$

where C is a constant and T is the absolute temperature. This relation is good for paramagnetic gases at ordinary field strengths and temperatures, where the interaction of the molecules is negligible.

For liquids and solids, where the interaction of the molecules may be large, a modified Curie law called the Curie–Weiss law is used.

$$\chi_m = \frac{C}{T-\theta}, \tag{34-14}$$

where θ is the Weiss temperature and is characteristic of the substance. It may be positive or negative. The equation only holds for $T > \theta$, and for many substances no single equation holds over a wide temperature range. Of the common gases, only oxygen and nitric oxide are paramagnetic. Most solids have filled electron inner shells and are diamagnetic. Exceptions are compounds of transition elements, which have unfilled electron shells. It should be mentioned again that diamagnetism is also present in paramagnetic substances, but its presence is masked by the stronger paramagnetic effect. The magnetic behavior of metals is even more complex. In metals, the bound electrons are attached to positive ions, which (except in the transition group) have filled electron shells and are, therefore, diamagnetic. The conduction electrons give rise to a diamagnetic and paramagnetic effect, both of the same order of magnitude and both independent of temperature. Exceptions are ferromagnetic metals such as iron, nickel, cobalt, along with many alloys.

34-5 FERROMAGNETISM

Ferromagnetic substances are those for which χ_m is positive and large, and μ_r is much greater than unity. In ferromagnetic materials, the permeability is not a constant but is a function of temperature and magnetic field. Ferromagnetic substances are all solids. Their magnetic properties change abruptly at a certain temperature called the "Curie point." Above the Curie point, χ_m is independent of field strength and follows the Curie–Weiss law approximately. Below the Curie point, very large values of magnetization are produced by quite small fields, and the magnetization varies nonlinearly with field strength, which means that the permeability varies with field strength. Figure 34-2 is a typical magnetization curve for a ferromagnetic material. This curve is not reversible. If the material is initially unmagnetized, and an increasing field $\mathbf{H}$ is

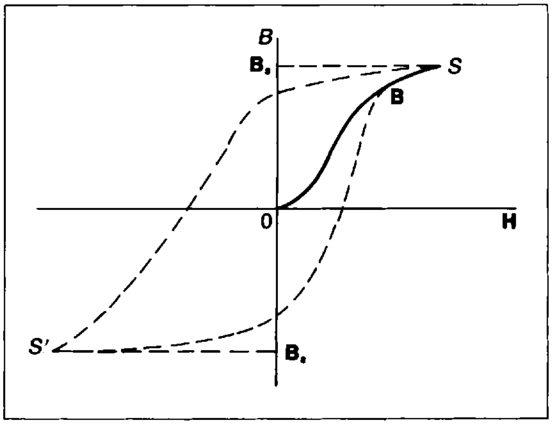


Figure 34-2 Magnetization curve for a ferromagnetic material.

applied, **B** follows the solid curve 0S and reaches a saturation value $\mathbf{B}_s$. If the field **H** is now reduced, **B** follows the broken curve SS'. When $\mathbf{H} = 0$, **B** still has a finite value, known as *residual induction.* As the field is increased in the reverse direction, **B** decreases and finally becomes zero. The value of this reversed field to make $\mathbf{B} = 0$ is called the *coercive force.* If we continue to increase the field **H** in the reverse direction, a reverse induction **B** is set up, which reaches a saturation value $\mathbf{B}_s' = -\mathbf{B}_s$. If the field is now again increased in the positive direction, **B** will follow the broken curve $S'S$, and we have what is called a hysteresis loop or cycle traced out by the broken curve.

A detailed explanation of ferromagnetism is beyond the scope of this book, so only a brief qualitative description will be attempted here. Ferromagnetic substances contain atoms with permanent magnetic dipoles. Groups of these atomic dipoles point in the same direction. These groups are called "domains." In the unmagnetized material, the domains are oriented at random, so there is no resultant magnetic moment in any direction. When a magnetic field is applied, the domains where the magnetization is parallel or at a small angle with the direction of the field grow at the expense of those where the magnetization is anti-parallel, and the boundaries of the domains are displaced. It is possible to map out domain boundaries and to detect the shifting of domain orientation experimentally.

There are two classes of ferromagnetic materials: (1) magnetically soft materials, which have a high permeability and are easily magnetized and demagnetized, and (2) magnetically hard materials, which have a relatively low permeability and are difficult to magnetize and demagnetize.

34-6 THERMOELECTRICITY

In 1821, Seebeck discovered that a current flows in a simple circuit composed of two different metals, when the junctions are maintained at different temperatures. A thermocouple, a temperature measuring device that utilizes this effect, is illustrated in Figure 34-3. The extended junctions are temperature probes composed of metals A and B twisted together. Since a steady current flows in the circuit, an emf must exist. This emf is called a thermal emf because it is produced by a difference in temperature between the two junctions. The effect is called the *Seebeck effect.*

The Seebeck current is due to the fact that the density of free electrons in a metal depends on the temperature and is different for different metals. Therefore, when two different metals are connected to form two junctions and the junctions are at different temperatures, electron diffusion at the junctions takes place at different rates, and there is a net flow of electrons from one junction to the other. In terms of energy considerations, energy is absorbed from the heat source at the hot junction, which is in part liberated to the surroundings at the cold junction, the remainder being converted into electrical energy in the circuit.

In 1834, Peltier discovered that if a current from an external source is sent through a circuit composed of two different metals (similar to a thermocouple), one junction warms up and the other cools, which means that heat is absorbed at one junction and liberated at the other. The heat absorbed or liberated at a junction between two metals depends on the direction of the current. If the Seebeck emf is from metal A to metal B at a

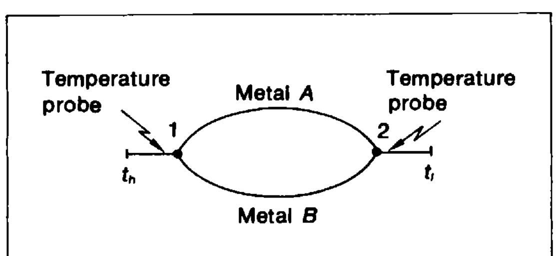


Figure 34-3 Thermocouple of metals A and B with junctions 1 and 2 at temperatures t_h and t_l ($t_h > t_l$).

hot junction, an external emf applied in this direction produces a cooling at this junction and a heating at the other junction. The heat absorbed or liberated at a junction is dependent on the quantity of charge that crosses the junction, on the temperature of the junction, and on the nature of the two metals.

It was discovered by Sir William Thomson (Lord Kelvin) that a thermal emf is set up in a metal having a temperature gradient. The emf in a section of the conductor is directly proportional to the difference in temperature between its ends. It is possible to explain this effect in terms of our electron gas model of conducting materials. The density of free electrons at the hot end of the conductor will be greater than at the cold end. This will cause the electrons to migrate from the hot end to the cold end of the conductor, leaving the hot end positive and the cold end negative. As a result, an emf should be produced in the conductor directed from the cold to the hot end. This is the case for most metals, but there are some metals in which the emf is in the opposite direction, an effect for which there is no simple explanation.

The Seebeck, Peltier, and Thomson effects are all completely reversible. The Peltier emf, called the Peltier coefficient (π_{AB}) of a junction of metals A and B, is defined as the quantity of heat absorbed or liberated per unit electric charge crossing the junction. Thus,

$$\pi_{AB} = \frac{H_P}{q}, \tag{34-15}$$

where H_P is the Peltier heat.

The Thomson emf of an infinitesimal length of wire that has a temperature difference dt may be written σdt, where σ is called the Thomson coefficient and is defined as the heat absorbed or liberated in this length of wire per unit electric charge transferred. Thus,

$$\sigma dt = \frac{H_T}{q}, \tag{34-16}$$

where H_T is the Thomson heat. The Seebeck emf in a circuit like the thermocouple circuit (Figure 34-3) is the resultant of two Peltier emf's and two Thomson emf's. Referring to this thermocouple circuit,

$$\mathscr{E}_{AB} = (\pi_{AB})_{t_h} + (\pi_{BA})_{t_l} + \int_{t_l}^{t_h} \sigma_A \, dt + \int_{t_h}^{t_l} \sigma_B \, dt. \tag{34-17}$$

In this equation, the direction of the emf is taken so that the current flows from B to A at the hot junction. The above equation is the "fundamental thermocouple equation." Since the emf is a function of the t_h and t_l, it is possible to use a thermocouple to measure temperature if either t_h or t_l is fixed.

Example 1. A toroid with a core of relative permeability 500 is 1 m in circumference. It has 1000 turns of wire carrying a current of 0.5 amperes.

(a) What is the magnetic intensity?

(b) What is the flux density in the core?

SOLUTION

(a) The magnetic intensity is given by Eq. (34-4),

$$H = \frac{NI}{l} = \frac{(1000)(0.5\text{ amp})}{1\text{ m}} = 500\text{ amp/m}.$$

(b) The flux density in the core is given by Eq. (34-8),

$$B = \mu H = (500\text{ nt/amp}^2)(500\text{ amp/m}) = 25 \times 10^4\text{ webers/m}^2.$$

PROBLEMS

1. What is the magnetic moment of a flat coil of 20 turns, having an area of 5 cm^2, and carrying a current of 8 amperes?
2. Consider a cylindrical rod of iron, length $L = 60$ cm, cross-sectional area $A = 5$ cm^2, which is bent into a ring with a small air gap $l = 0.1$ cm between the ends. If 400 turns of wire, carrying 6 amperes, are wound around the iron ring, and if the relative permeability of the iron is 5000, what is the magnetic flux inside the iron?
3. An iron ring of relative permeability 1200 has a mean circumference of 40 cm and an area of 5 cm^2. The ring is wound with 350 turns of wire and carries a current of 0.2 amperes.
   (a) What is the magnetic intensity in the ring?
   (b) What is the flux density in the ring?

4. A metal ring having 800 turns of wire and a mean diameter of 16 cm carries a current of 0.3 amperes. The relative permeability of the core is 650.
   (a) What is the flux density in the core?
   (b) What part of the flux density is due to electronic loop currents in the core?
5. The area of a hystersis loop is proportional to an energy loss. By the consideration of the physical dimensions of **B** and **H**, show that the area of the loop has the units joules/m^3.
6. An iron rod of cross-sectional area 4 cm^2 and relative permeability 1000 is bent into a circle with inside radius 9 cm and outside radius 11 cm. It is wound with 1000 turns of wire per meter and a current of 2 amperes is sent through the wire. Determine the magnetic flux in the iron.
7. A solenoid has a cross-sectional area of 6 cm^2 and the magnetic flux density (**B**) in its air core is 10^{-3} webers/m^2. What is the value of **B** and flux ϕ in the solenoid if an iron core of relative permeability 1000 replaces the air core?
8. The iron rod of Problem 6 is an alloy of relative permeability 10,000 and is wound with 1000 turns of wire carrying 1 ampere. Find **H**, **B**, and ϕ.
9. The hysteresis loops for four ferromagnetic alloys are shown in Figure 34-4, all to the same scale. Which of the materials would make the best permanent magnet? Explain your answer.

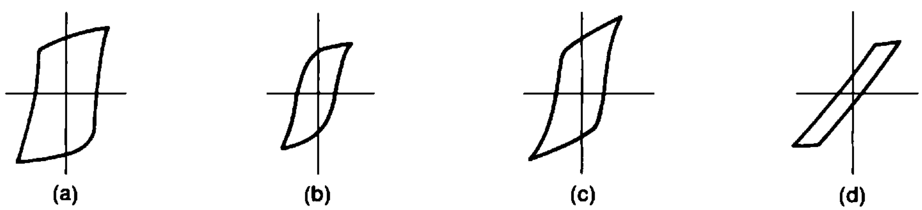


Figure 34-4

10. The current in the windings of a coil of 500 turns wound around an iron ring of relative permeability 1000 and circumference 50 cm is 1 ampere. Calculate:
    (a) the magnetic intensity,
    (b) the flux density,
    (c) the magnetization, and
    (d) the magnetic susceptibility.
11. A magnet has a coercive force of 6×10^3 amp/m. It is inserted in a solenoid 10 cm long, having 50 turns. What current must be carried in the solenoid in order to demagnetize it?
12. Why is Eq. (34-17) the result of two Peltier emf's and two Thomson emf's?
13. A closed thermocouple circuit consists of the two thermal junctions and a low resistance galvanometer all in series. The galvanometer has a resistance of 9 ohms and reads directly in millivolts across its terminals. The resistance of the rest of the circuit is 1 ohm, and the thermocouple develops an emf of 20 microvolts per degree difference of temperature between the junctions. When one junction is maintained at 0°C and the other is placed in a molten metal, the galvanometer reading is 5.40 millivolts. What is the temperature of the molten metal?

35 Alternating Currents

35-1 INTRODUCTION

An alternating current is one that flows alternately in opposite directions in a circuit. When the current rises from zero to a maximum, returns to zero, increases to a maximum in the opposite direction, and finally returns to zero again, it is said to have completed a *cycle*. The time required for the completion of one cycle is called the *period* (τ). The number of cycles per second is called the *frequency* (f) of alternation of the current. The maximum instantaneous current in either direction is called the *amplitude*.

Figure 35-1 is a graph of an alternating current versus time, where positive values of the instantaneous current i represent currents in one direction and negative values in the other direction. This graph can be represented by the equation

$$i = I_M \sin(2\pi ft + \phi), \qquad (35\text{-}1)$$

where I_M is the maximum value of the current and f is the frequency of alternation of the current. ϕ is a constant angle called the *phase angle*, which fixes the value of i at time $t = 0$ relative to its maximum value.

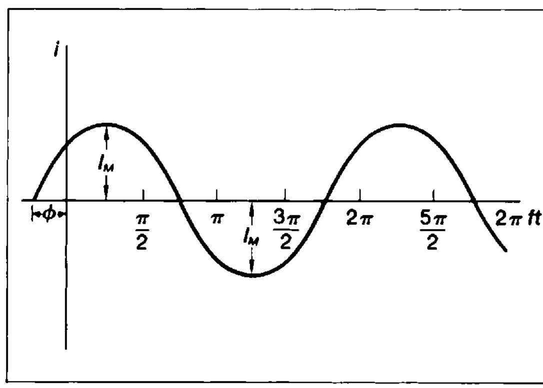


Figure 35-1 Graph of sinusoidal alternating current.

35-2 PROPERTIES AND EFFECTS OF ALTERNATING CURRENTS AND EMF'S

The effects of alternating currents and emf's are quite different from those due to direct currents and emf's. In general, it is possible for the applied emf and the resulting current not to be in phase. That is, the current and emf do not have maximum or minimum values at the same time. Alternating currents can flow continuously in a capacitor, while direct currents cannot. In an alternating current circuit, the simple form of Ohm's law must be modified to take care of capacitance and inductance.

Since an alternating current varies from zero to a maximum value, we may ask what is the effective value of this current, or, in other words, what is meant by an alternating current of 1 ampere. Experiment has shown that the heating effect of an electric current is independent of current direction. As a result, the heating effect of an alternating current is taken as the basis for the definition of an ampere.

An alternating current is said to have an effective value of 1 ampere when it will develop the same amount of heat in a given resistance as 1 ampere direct current at the same time. If we designate the effective alternating current by I, then

$$RI^2 = \text{average value of } Ri^2$$

or

$$I = (\text{average of } i^2)^{1/2}$$
$$= [\text{average of } I_M^2 \sin^2(2\pi ft + \phi)]^{1/2}.$$

The average value of $\sin^2\theta$ and $\cos^2\theta$ per cycle must be the same since both have the same shape, where $\theta = (2\pi ft + \phi)$. Since $\sin^2\theta + \cos^2\theta = 1$, the average value of both $\sin^2\theta$ and $\cos^2\theta$ is $\frac{1}{2}$. Therefore,

$$I = \frac{I_M}{\sqrt{2}} = 0.707 I_M. \qquad (35\text{-}2)$$

In a similar way, we define the effective value of the voltage V as

$$V = \frac{V_M}{\sqrt{2}} = 0.707 V_M, \qquad (35\text{-}3)$$

where V_M is the maximum or peak voltage. Alternating currents and voltages are almost always expressed in terms of their effective values, which are sometimes called their "root-mean-square" values.

35-3 CIRCUIT WITH RESISTANCE ONLY

In the case of an alternating current circuit that has only a pure resistance, the instantaneous current and voltage are connected by Ohm's law,

$$i = \frac{v}{R}. \qquad (35\text{-}4)$$

Therefore, if an alternating voltage $v = V_M \sin 2\pi ft$ is across the resistance, the instantaneous current is

$$i = \frac{V_M \sin 2\pi ft}{R} = I_M \sin 2\pi ft, \qquad (35\text{-}5)$$

and the current and voltage are both in phase since they both vary with $\sin 2\pi ft$. This is illustrated graphically in Figure 35-2(a). The instantaneous current and voltage values may be represented as the projection of vectors on the vertical axis, which are rotating about a common origin with an angular velocity $\omega = 2\pi f$. Figure 35-2(b) is such a vector representation showing current and voltage in phase with one another. Since the effective current and voltage are each given by $I = 0.707 I_M$ and $V =$

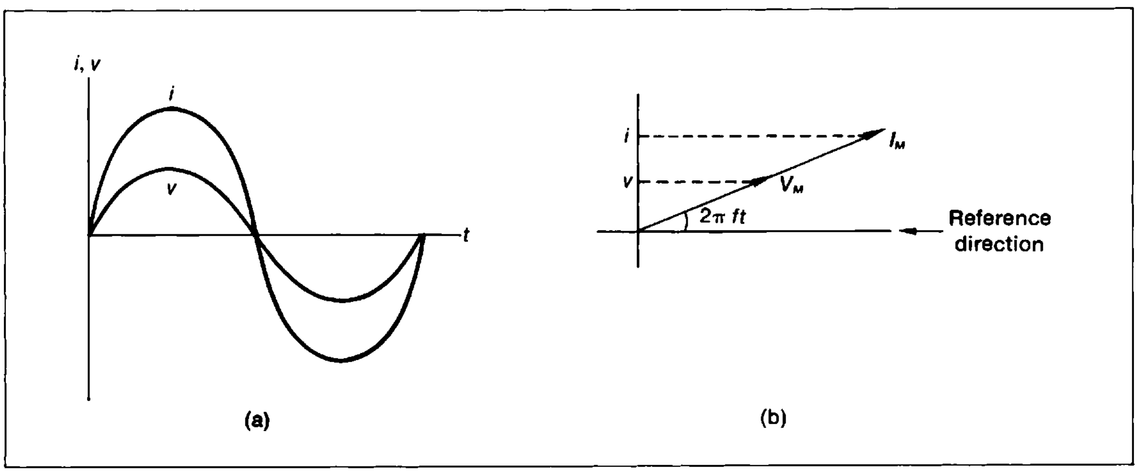


Figure 35-2

$0.707 V_M$, respectively, the relation

$$I = \frac{V}{R} \tag{35-6}$$

is also true.

35-4 CIRCUIT WITH CAPACITANCE ONLY

It was mentioned at the beginning of the chapter that an alternating current can flow continuously in a capacitor. It should be pointed out that no current flows through the dielectric between the plates of a capacitor. The current in the circuit is due to the charging and discharging of the capacitor in opposite directions. The current through the circuit is equal to the time rate of change of charge on the capacitor, since the current flowing to or from either plate is the amount of charge transferred per second. Therefore, if an alternating-current voltage is applied to the capacitor, the instantaneous charge on the capacitor is

$$q = Cv = CV_M \sin 2\pi ft,$$

and the current i is

$$i = \frac{dq}{dt} = \frac{d}{dt}(CV_M \sin 2\pi ft)$$
$$= 2\pi fCV_M \cos 2\pi ft.$$

The maximum current is obviously

$$I_M = 2\pi fCV_M \tag{35-7}$$

and

$$i = I_M \cos 2\pi ft. \tag{35-8}$$

Ohm's law in terms of the effective voltage and current is

$$I = \frac{V}{\frac{1}{2\pi fC}}. \tag{35-9}$$

The quantity $1/2\pi fC$ is called the *capacitive reactance* (X_C) of the capacitor and has the dimensions of resistance. Then

$$I = \frac{V}{X_C}. \tag{35-10}$$

Figure 35-3(a) is a graphical representation of the instantaneous voltage and current equations, and Figure 35-3(b) shows the phase difference by a vector representation.

The current and the voltage are out of phase by 90°. The current reaches its maximum one-quarter period ahead of the voltage. Thus, it may be said that in a capacitor the current leads the voltage by 90°.

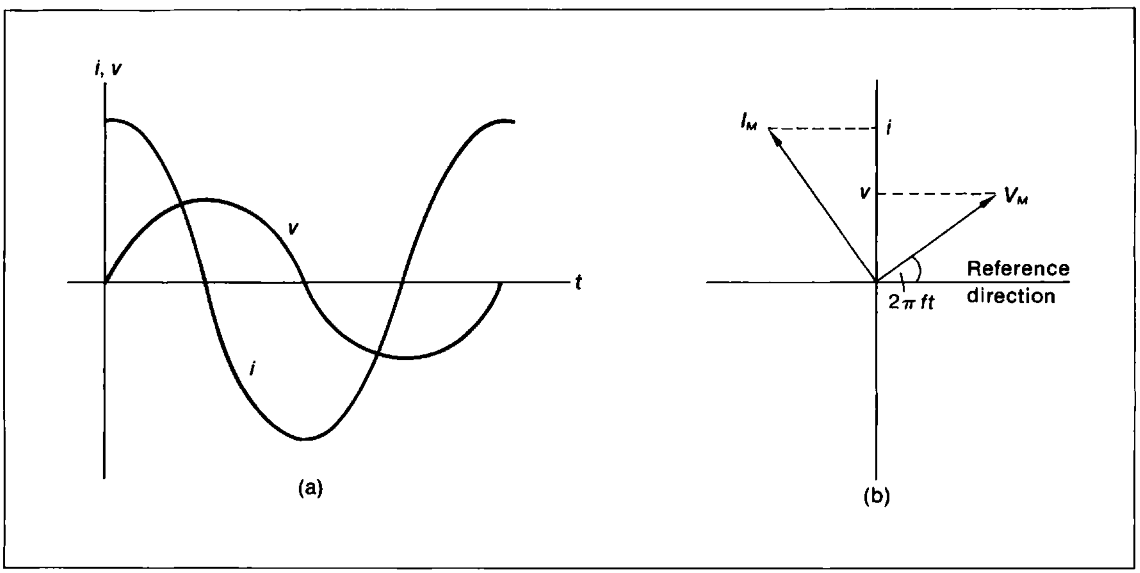


Figure 35-3

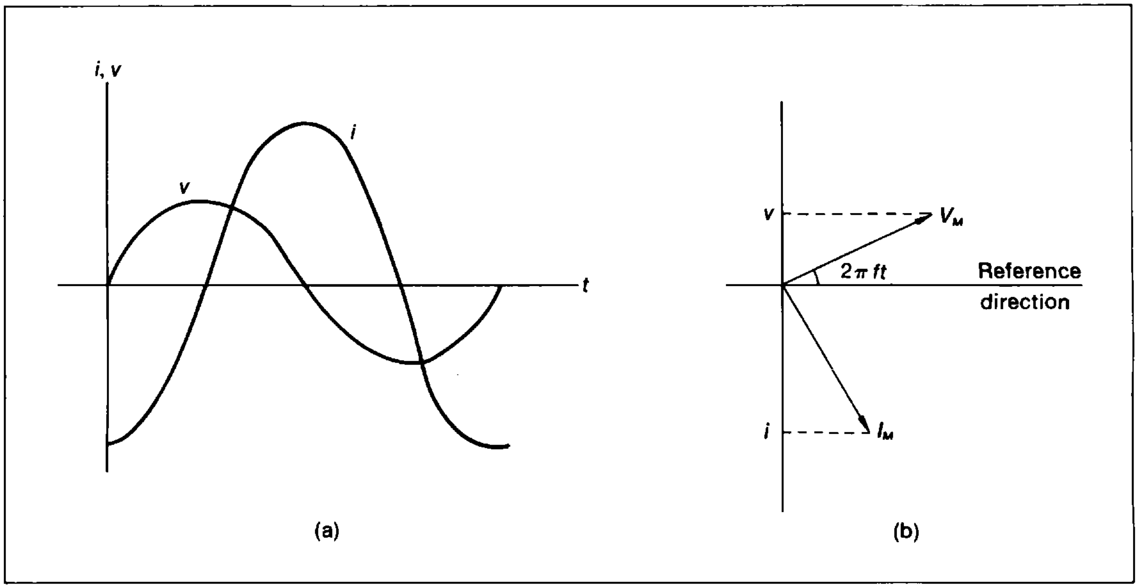


Figure 35-4

35-5 CIRCUIT WITH INDUCTANCE ONLY

If an alternating-current voltage is applied to a pure inductance L, the instantaneous potential difference across it is given by

$$v = V_M \sin 2\pi ft = L\frac{di}{dt}$$

and

$$di = \frac{V_M}{L} \sin 2\pi ft\, dt.$$

Integration of both sides, gives

$$i = -\frac{V_M}{2\pi fL} \cos 2\pi ft.$$

The maximum current is obviously

$$I_M = \frac{V_M}{2\pi fL} \tag{35-11}$$

and

$$i = -I_M \cos 2\pi ft. \tag{35-12}$$

Ohm's law in terms of effective voltage and current is

$$I = \frac{V}{2\pi fL}. \tag{35-13}$$

The quantity $2\pi fL$ is called the *inductive reactance* X_L of the inductance. Then,

$$I = \frac{V}{X_L}. \tag{35-14}$$

Figure 35-4(a) is a graphical representation of the instantaneous voltage and current equations and Figure 35-4(b) shows the phase difference by a vector diagram.

The current and the voltage are out of phase by 90°. The voltage reaches its maximum one-quarter period ahead of the current. Thus, it may be said that, in an inductance, the voltage leads the current by 90°.

35-6 SERIES CIRCUIT WITH RESISTANCE, CAPACITANCE, AND INDUCTANCE

We shall now consider a circuit with a resistance, capacitance, and inductance in series, as illustrated in Figure 35-5. The instantaneous potential difference v between the terminals a and b is equal to the sum of the potential differences across each circuit element. If we use the voltage-current relations

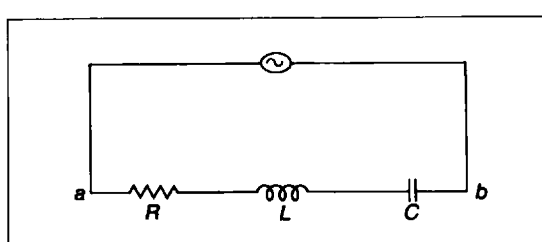


Figure 35-5 A series *R-L-C* circuit.

for each element, and use the vector diagram for this type of circuit as illustrated in Figure 35-6 as a guide, we can obtain a voltage–current relation for this circuit. Referring to the vector diagram, we see that

$$V = I\sqrt{R^2 + (X_L - X_C)^2} \tag{35-15}$$

or

$$I = \frac{V}{\sqrt{R^2 + \left(2\pi f L - \frac{1}{2\pi f C}\right)^2}} = \frac{V}{Z}. \tag{35-16}$$

We have introduced the quantity

$$Z = \sqrt{R^2 + \left(2\pi f L - \frac{1}{2\pi f C}\right)^2}, \tag{35-17}$$

which is called the *impedance* of the circuit.

The voltage in this circuit can either lead or lag the current, depending on the relative values of the inductance and capacitance. From Figure 35-6 the phase angle is given by the relation

$$\tan\phi = \frac{2\pi f L - \frac{1}{2\pi f C}}{R} \tag{35-18}$$

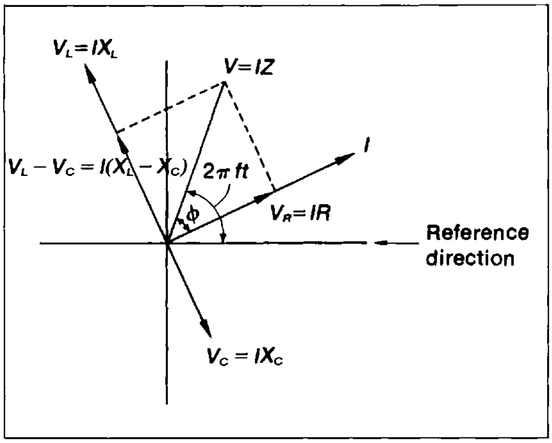


Figure 35-6 Vector diagram for a series *R-L-C* circuit.

Example 1. A series circuit connected across a 110 volt, 60 cycle line consists of a coil with 100 ohms resistance and 0.20 henrys inductance, and a capacitor of capacitance 100 microfarads. Calculate:

(a) the current in the circuit,

(b) the phase angle between the current and supply voltage, and

(c) sketch the vector diagram for this circuit.

SOLUTION

(a) $X_L = 2\pi f L = (2\pi)(60\text{ cps})(0.20\text{ henrys})$
$= 75.4\text{ ohms}.$

$$X_C = \frac{1}{2\pi f C} = \frac{1}{(2\pi)(60\text{ cps})(100 \times 10^{-6}\text{ farads})}$$
$$= 26.6\text{ ohms}$$

$$Z = \sqrt{R^2 + (X_L - X_C)^2}$$
$$= \sqrt{(100\Omega)^2 + [(75.4 - 26.6)\Omega]^2}$$
$$= 111\text{ ohms}$$

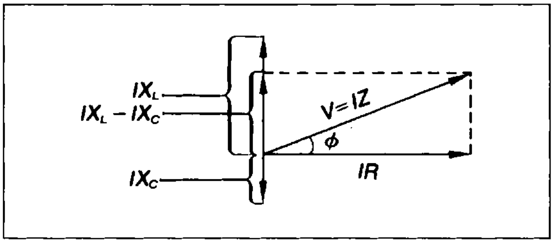
$$I = \frac{V}{Z} = \frac{110\text{ volts}}{111\Omega} = 0.991\text{ amp.}$$

(b) $\tan\phi = \frac{X_L - X_C}{R} = \frac{75.4\Omega - 26.6\Omega}{100\Omega} = 0.488$

$\phi = 26°$

Voltage leads the current by 26°.

(c)



35-7 PARALLEL CIRCUIT WITH RESISTANCE, CAPACITANCE, AND INDUCTANCE

Another circuit we will examine is one with resistance, capacitance, and inductance in parallel, as illustrated in Figure 35-7. In this circuit, the potential difference across each branch is the same, but

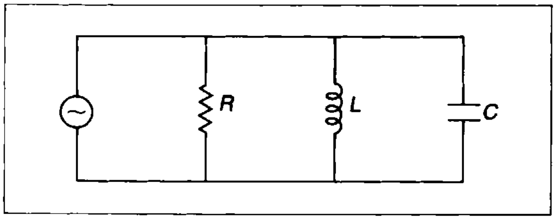


Figure 35-7 A parallel *R-L-C* circuit.

the current through each branch is generally different. If V is the effective value of the applied emf, then the current through each branch of the circuit is

$$I_R = \frac{V}{R},$$

$$I_C = \frac{V}{\frac{1}{2\pi fC}} = \frac{V}{X_C},$$

$$I_L = \frac{V}{2\pi fL} = \frac{V}{X_L}.$$

The vector diagram for this circuit is shown in Figure 35-8. The total effective current in the circuit is obtained by the vector addition of the component currents. The magnitude of this current is

$$I = V\left[\frac{1}{R^2} + \left(\frac{1}{X_C} - \frac{1}{X_L}\right)^2\right]^{1/2}, \qquad (35\text{-}19)$$

and the impedance is given by the relation

$$Z = \left(\frac{1}{R^2} + \left(2\pi fC - \frac{1}{2\pi fL}\right)^2\right)^{-1/2}$$

$$= \frac{1}{\sqrt{\frac{1}{R^2} + \left(2\pi fC - \frac{1}{2\pi fL}\right)^2}}. \qquad (35\text{-}20)$$

The phase angles (ϕ) may be computed with the aid of the diagram.

$$\tan\phi = \frac{\frac{1}{X_C} - \frac{1}{X_L}}{\frac{1}{R}} = \frac{2\pi fC - \frac{1}{2\pi fL}}{\frac{1}{R}}. \qquad (35\text{-}21)$$

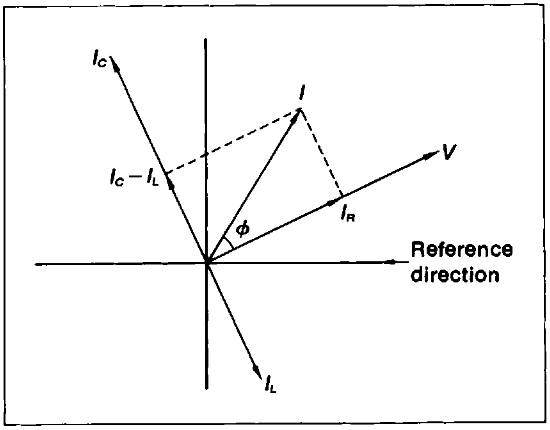


Figure 35-8 Vector diagram for a parallel *R-L-C* circuit.

It should be noted that the signs of X_C and X_L are reversed in comparison with their use in the series circuit. Why?

35-8 POWER IN ALTERNATING CURRENT CIRCUITS

The instantaneous power in an alternating current circuit is

$$p = vi = V_M I_M \sin 2\pi ft \sin(2\pi ft - \phi), \qquad (35\text{-}22)$$

where ϕ is the phase angle between the voltage and the current, which may be either positive or negative. The quantity of interest is not the instantaneous power, but the average power. We may transform Eq. (35-22) by the use of the trigonometric identity

$$\sin a \sin b = \frac{1}{2}[\cos(a-b) - \cos(a+b)],$$

and obtain the equation

$$p = \frac{1}{2} V_M I_M [\cos\phi - \cos(4\pi ft - \phi)]. \qquad (35\text{-}23)$$

In this equation, $\cos(4\pi ft - \phi)$ is periodic, alternating between positive and negative values, so that its average value is zero. Thus, the average power in the circuit is given by

$$P = \frac{1}{2} V_M I_M \cos\phi. \qquad (35\text{-}24)$$

In terms of the effective current I and voltage V, the average power may be expressed as

$$P = VI\cos\phi. \qquad (35\text{-}25)$$

The term $\cos\phi$ is called the *power factor* of the circuit. For a pure resistance circuit, $\phi = 0$ and the average power is $P = VI$. For a pure inductance or capacitance, $\phi = 90°$ and $P = 0$.

Equation (35-23) may be put in the form

$$p = VI[(1 - \cos 4\pi ft)\cos\phi + \sin 4\pi ft \sin\phi], \qquad (35\text{-}26)$$

by expanding the variable term and replacing V_M and I_M by V and I. The first term in the square brackets represents the fluctuating power dissipated into heat, for this term is non-negative and exists alone if the component has a pure resistance. The second term represents the power alternately stored in a circuit component and released to the circuit, for this term changes sign and is the only term present if the component has a pure reactance. An ideal capacitor stores energy in the form of an electric field and an ideal inductor—in the form of a magnetic field. The peak power involved in energy storage is

$$P_S = VI \sin \phi. \qquad (35\text{-}27)$$

This is called *reactive power*. To help distinguish reactive power from instantaneous or average power, different units are used. For instantaneous or average power, the unit is the watt if the current is in amperes and the electromotive force or potential difference is in volts, while, by convention, reactive power is left in volt-amperes.

35-9 RESONANT CIRCUITS

Equation (35-17) is the expression for the impedance of a series AC circuit. It is evident that for a given value of the resistance R, the impedance Z is a minimum when

$$2\pi f L = \frac{1}{2\pi f C}.$$

This happens when the frequency is

$$f_R = \frac{1}{2\pi\sqrt{LC}}, \qquad (35\text{-}28)$$

where f_R is called the *resonant frequency* of the circuit. We then have what is called a resonant circuit. For this particular frequency, the impedance reduces simply to the resistance, i and v are in phase, and the current–voltage relationship is as given by Ohm's law for a direct current circuit. Also, at the resonant frequency, the current is a maximum, and the voltages across the capacitance and inductance are equal and opposite, so the phase angle is zero.

It is also possible to have a resonant parallel circuit. Equation (35-20) is the expression for the impedance of a parallel AC circuit. It is evident that, for a given value of the resistance R, the impedance Z is a minimum when

$$2\pi f C = \frac{1}{2\pi f L},$$

which leads again to

$$f_R = \frac{1}{2\pi\sqrt{LC}}, \qquad (35\text{-}29)$$

It should be noted from Eqs. (35-19) and (35-20) that in a resonant parallel circuit the impedance is a maximum and the current is a minimum.

Since the capacitance and inductance of a circuit can be varied, the resonant frequency can also be varied. Resonant circuits are of great importance in radio and television. They provide a means by which one particular frequency can be selected so as to respond to the signal of a desired station.

35-10 THE TRANSFORMER

An alternating current has the advantage over direct current in that its voltage can be changed in any desired ratio by means of a transformer. This is the reason why alternating current is used almost universally for power transmission.

A transformer consists of two separate coils wound on opposite sides of a common iron core as illustrated in Figure 35-9. The coil to which the current is supplied is called the *primary coil*, and the coil from which the current is delivered is called the *secondary coil*. An alternating current is induced in the secondary coil by the changing magnetic flux in the primary coil due to the impressed alternating current in the primary coil. If we let N_p equal the number of turns in the primary coil and N_s the number of turns in the secondary coil, respectively, and assume that the magnetic flux (ϕ) is the same throughout the iron core, then the primary flux linkage is $N_p\phi$ and the secondary flux linkage is $N_s\phi$. Since the flux is varying, an induced emf is produced in both coils that is proportional to the rate of change of flux, and we have

$$\frac{V_s}{V_p} = \frac{N_s \dfrac{d\phi}{dt}}{N_p \dfrac{d\phi}{dt}} \qquad (35\text{-}30)$$

The rate of change of flux will be the same in both coils, since the change in both coils is due to the same source, an alternating current at a definite fre-

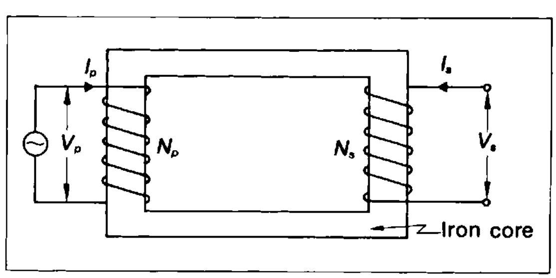


Figure 35-9 Schematic diagram of a transformer.

quency. Therefore, Eq. (35-30) becomes

$$\frac{V_s}{V_p} = \frac{N_s}{N_p}. \qquad (35\text{-}31)$$

Thus, the ratio of the secondary voltage to the primary voltage can be controlled by the choice of the number of turns on each coil.

The power loss in a transformer is usually quite small, and it can be shown that the current and voltage are nearly in phase. We can therefore arrive at the following approximate relations, namely,

$$V_s I_s = V_p I_p$$

and

$$\frac{I_s}{I_p} = \frac{V_p}{V_s} = \frac{N_p}{N_s}. \qquad (35\text{-}32)$$

Equation (35-32) points out that the voltage ratio is the same as the ratio of the number of turns, while the current ratio is the inverse ratio of the number of turns.

Example 2. A series circuit includes a resistor of resistance 20 ohms, a capacitor of capacitance 30 microfarads, and an inductor of inductance 0.30 henrys. Impressed upon this is a 60 cycle AC voltage whose effective value is 120 volts. Calculate:

(a) the inductive reactance,
(b) the capacitive reactance,
(c) the impedance,
(d) the effective current in the circuit,
(e) the phase angle,
(f) the power factor,
(g) the power dissipation in the circuit,
(h) the maximum instantaneous voltage applied to the circuit,
(i) if the frequency of the supply voltage could be changed, the resonant frequency,
(j) when in resonance the current in the circuit, and
(k) the voltage drop across the capacitor at resonance.

SOLUTION

(a) $X_L = 2\pi f L = (2\pi)(60 \text{ cps})(0.30 \text{ henrys}) = 113.1 \text{ ohms}.$

(b) $X_C = \dfrac{1}{2\pi f C} = \dfrac{1}{(2\pi)(60 \text{ cps})(30 \times 10^{-6} \text{ farad})} = 88.3 \text{ ohms}.$

(c) $Z_C = \sqrt{R^2 + (X_L - X_C)^2} = \sqrt{(20\Omega)^2 + (113.1\Omega - 88.3\Omega)^2} = 32 \text{ ohms}.$

(d) $I = \dfrac{V}{Z} = \dfrac{120 \text{ volts}}{32\Omega} = 3.75 \text{ amp}.$

(e) $\phi = \tan^{-1}\dfrac{X_L - X_C}{R} = \tan^{-1}\dfrac{24.8\Omega}{20\Omega} = 51.1°.$

(f) $\cos\phi = \cos 51.1° = 0.628.$

(g) $P = VI\cos\phi = (120 \text{ volts})(3.75 \text{ amp})(0.628) = 282 \text{ watts}.$

(h) $V_M = \sqrt{2}V = (1.414)(120 \text{ volts}) = 169.5 \text{ volts}.$

(i) $f_R = \dfrac{1}{2\pi\sqrt{LC}} = \dfrac{1}{2\pi\sqrt{(0.30 \text{ henrys})(30 \times 10^{-6} \text{ farad})}} = \dfrac{1}{6\pi \times 10^{-3}} = 53 \text{ sec}^{-1}.$

(j) $I = \dfrac{V}{R} = \dfrac{120 \text{ volts}}{20\Omega} = 6 \text{ amp}.$

(k) $V = IX_C = \dfrac{I}{\sqrt{2}\pi f_R C} = \dfrac{6 \text{ amp}}{\sqrt{(2\pi)(53 \text{ cps})(30 \times 10^{-6} \text{ farad})}} = 60 \text{ volts}.$

PROBLEMS

1. In a series R-L-C circuit, $R = 80$ ohms, $X_L = 100$ ohms, and $X_C = 40$ ohms. What is the impedance of the circuit?
2. An AC circuit has a resistance of 500 ohms, an inductor of inductance 5 henrys, and a capacitor of capacitance 20 microfarads all in series.
   (a) At a frequency of 60 cps, what will be the power supplied to this circuit if $V_M = 220$ volts?
   (b) At what frequency will resonance occur?

3. A resistor of resistance 8 ohms is connected in a series circuit to an inductor of inductive reactance 8 ohms.
   (a) What is the impedance of this circuit?
   (b) Does the current lag or lead the voltage?
   (c) What is the phase angle?

4. Every inductor has a resistance due to the wire in its windings. With reference to Figure 35-10.
   (a) Calculate the effective current in the circuit.
   (b) Calculate the average power in the circuit.

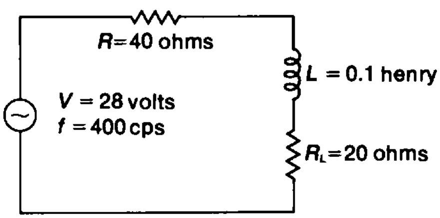


Figure 35-10

5. A series circuit consists of an 8 ohm resistor, a 12 ohm inductor, and a capacitor. A 60 cycle AC is impressed across this circuit.
   (a) What capacitance will cause this circuit to be in resonance?
   (b) What will be the phase angle between the current and the voltage when the circuit is in resonance?

6. A series AC circuit has an impedance of 50 ohms and a power factor of 0.60 at 60 cps, the voltage lagging the current.
   (a) Should an inductor or a capacitor be placed in series with the circuit to raise its power factor?
   (b) What size element will raise the power factor to unity?

7. (a) An AC series circuit contains a resistance of 200 ohms, an inductance of 10 henrys, and a capacitance of 2.5 microfarads in series with a variable frequency generator. What is the resonant frequency of the circuit?
   (b) If the applied voltage in part (a) has a *peak* value of 200 volts, what is the *effective* current at resonance?

8. An AC series circuit consists of a generator, which provides an effective voltage of 500 volts at 2000 rad/sec to a resistor with a resistance of 150 ohms, an inductor of inductance 0.4 henrys, and a capacitor of capacitance 0.5 microfarads. Calculate:
   (a) the reactance of the circuit,
   (b) the impedance of the circuit,
   (c) the effective current and current amplitude,
   (d) the power factor and average power dissipation, and
   (e) draw the vector diagram of the circuit roughly to scale.

9. A resistance of 80 ohms, an inductance of 0.4 henrys, and a capacitance of 100 microfarads are connected in series with an AC source whose angular frequency is 250 rad/sec. The effective current is 1 ampere.
   (a) Calculate the reactances X_L and X_C of the circuit.
   (b) Calculate the effective voltage and the voltage amplitude.
   (c) Draw the vector diagram of the circuit roughly to scale.

10. The mutual inductance of a transformer is 0.50 henrys. Find the expression for the voltage induced in the secondary coil when the primary coil carries an alternating current $I_p = 0.2 \sin \omega t$.

11. A series circuit includes a resistor of resistance 5 ohms, a capacitor of capacitance 20 microfarads, and an inductor of inductance 0.50 henrys. Impressed on this is an AC voltage whose effective value is 120 volts and whose frequency is 60 cps. Calculate:
    (a) the impedance,
    (b) the effective current,
    (c) the power factor, and
    (d) the power dissipation in the circuit.
12. For the circuit shown in Figure 35-11, it is desired that the current I_2 shall lead the current I_1 by 30°. What must be the value of R_2 if $L_1 = 12$ henrys, $L_2 = 7$ henrys, $R_1 = 3$ ohms, f is 1000 cps, and V_M is 68 volts?

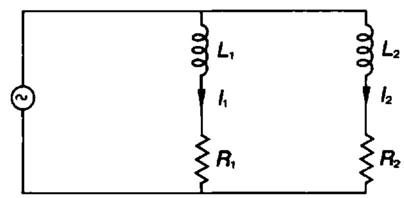


Figure 35-11

13. A resistance of 400 ohms, a pure inductance of 0.55 henrys, and a capacitance of 4 microfarads are connected in series with an AC source whose effective voltage is 100 volts. The angular frequency is 1000 rad/sec. Calculate:
    (a) the reactance of the circuit,
    (b) the impedance of the circuit,
    (c) the effective and peak currents,
    (d) the voltage across the resistor,
    (e) the voltage across the capacitor, and
    (f) the power dissipated in the circuit.
14. A resistance of 100 ohms, a capacitance of 10 microfarads, and a coil having a resistance of 200 ohms and an inductance of 0.5 henrys are connected in *parallel* to a 110 volt 60 cps power line. Calculate:
    (a) the current in the circuit and
    (b) the power supplied by the line.
15. A series circuit connected across a 200 volt 60 cps line consists of a capacitor of capacitive reactance 30 ohms, a resistor of 44 ohms, and a coil of inductive reactance 90 ohms and resistance 36 ohms. Calculate:
    (a) the current in the circuit and
    (b) the potential difference across each unit.
16. If the circuit in Problem 15 consisted of all the units in parallel, what would be the current in the circuit?
17. A step-down transformer operates on a 2200 volt line and supplies a load of 20 amperes. The ratio of the secondary to primary windings is 1:20. Determine:
    (a) the secondary voltage,
    (b) the primary current, and
    (c) the power output.
18. A toroidal coil has a 1000 turn primary and a 5 turn secondary wound on an iron ring. The rate of change of flux is 0.5 webers/sec in the core. Certain information is missing; hence, you cannot calculate all of the following quantities:
    (a) voltage across the secondary,
    (b) voltage across the primary, and
    (c) permeability of the core.
    For each quantity, calculate the value or list the missing data, and show how the calculation could be made had all the data been given.

36 Maxwell's Equations and Electromagnetic Waves

36-1 INTRODUCTION

Clerk Maxwell condensed all of electromagnetic field theory into four field equations. In recognition of this outstanding contribution to electromagnetic theory, they are called Maxwell's equations, and they govern all macroscopic electromagnetic phenomena. Maxwell developed these equations by correlating the expressions for electric and magnetic fields based on Gauss's, Ampere's, and Faraday's laws, which we have previously discussed. The equations used in either the differential or integral form are applicable to time dependent as well as constant fields, and lead to expressions for electromagnetic waves, which we shall discuss later in this chapter.

36-2 MAXWELL'S EQUATIONS†

Gauss's law was written in Chapter 30 in the following form:

$$\oint_A \mathbf{E} \cdot \mathbf{n} dA = \frac{\Sigma q}{\epsilon_0}, \qquad (36\text{-}1)$$

which may be written

$$\oint_A \mathbf{E} \cdot \mathbf{n} dA = \frac{Q}{\epsilon_0}, \qquad (36\text{-}2)$$

where $Q = \Sigma q$. This is one of Maxwell's equations in integral form. For a continuous charge distribution, either Eq. (36-1) or (36-2) may be written

$$\oint_A \mathbf{E} \cdot \mathbf{n} dA = \frac{1}{\epsilon_0} \oint_v \rho dv, \qquad (36\text{-}3)$$

where ρ is the volume charge density inside the volume v enclosed by the Gaussian surface.

The divergence theorem in vector analysis states that, for any vector function **F**,

$$\oint_A \mathbf{F} \cdot \mathbf{n} dA = \oint_v \nabla \cdot \mathbf{F} dv, \qquad (36\text{-}4)$$

where dv is any volume element enclosed by the surface area A. If we apply this theorem to Eq. (36-3), it becomes

$$\oint_A \nabla \cdot \mathbf{E} dv = \frac{1}{\epsilon_0} \oint_v \rho dv. \qquad (36\text{-}5)$$

Since we are integrating over the same volume, the integrands are equal, and we obtain the equation

$$\nabla \cdot \mathbf{E} = \frac{\rho}{\epsilon_0} \qquad (36\text{-}6)$$

or

$$\text{div}\, \mathbf{E} = \frac{\rho}{\epsilon_0},$$

†The vector relationships required in this chapter are discussed in Appendix I.

which is the differential form of Maxwell's Eq. (36-2).

Ampere's law for a complete circuit was given by Eq. (32-4) of Chapter 32 as

$$\mathbf{B} = \frac{\mu_0 I}{4\pi}\oint \frac{dl \times \hat{r}_0}{r^2}, \qquad (36\text{-}7)$$

which tells us that the lines representing **B** are circles about the axis of a current element. In Chapter 33, we saw that

$$\oint \mathbf{B} \cdot \mathbf{n}\, dA = 0, \qquad (36\text{-}8)$$

since magnetic induction lines are closed lines and every line entering a closed surface must leave it. Equation (36-8) is another of Maxwell's equations in integral form. Since, from the divergence theorem,

$$\oint_A \mathbf{B} \cdot \mathbf{n} dA = \oint_v \operatorname{div} \mathbf{B} dv,$$

thus

$$\operatorname{div} \mathbf{B} = 0 \qquad (36\text{-}9)$$

or

$$\nabla \cdot \mathbf{B} = 0,$$

which is another of Maxwell's equations, Eq. (36-8), in differential form.

In Chapter 33, we discussed Faraday's law, which is expressed in analytical form in that chapter by Eq. (33-4); namely,

$$\mathscr{E} = -\frac{d\phi}{dt}, \qquad (36\text{-}10)$$

where $\mathscr{E}$ is the induced emf and ϕ is the magnetic flux. Also

$$\phi = \oint_A \mathbf{B} \cdot \mathbf{n} dA \qquad (36\text{-}11)$$

and

$$\mathscr{E} = \oint \mathbf{E} \cdot d\mathbf{s}, \qquad (36\text{-}12)$$

Eq. (33-1) and Eq. (29-8), respectively. Substituting Eqs. (36-11) and (36-12) into Eq. (36-10), gives

$$\oint \mathbf{E} \cdot d\mathbf{s} = -\frac{d}{dt}\oint_A \mathbf{B} \cdot \mathbf{n} dA, \qquad (36\text{-}13)$$

which is another one of Maxwell's equations in integral form. By Stoke's theorem,

$$\oint \mathbf{E} \cdot d\mathbf{s} = \oint_A \operatorname{curl} \mathbf{E} \cdot \mathbf{n} dA. \qquad (36\text{-}14)$$

Furthermore, the time and space coordinates are independent variables. Therefore, Eq. (36-13) can be rewritten in the form

$$\oint_A \operatorname{curl} \mathbf{E} \cdot \mathbf{n} dA = -\oint_A \frac{d\mathbf{B}}{dt} \cdot \mathbf{n} dA. \qquad (36\text{-}15)$$

Since this equation holds for any surface area, the integrands must be equal, giving the differential form of Maxwell's equation [Eq. (36-13)]

$$\operatorname{curl} \mathbf{E} = -\frac{d\mathbf{B}}{dt} \qquad (36\text{-}16)$$

or

$$\nabla \times \mathbf{E} = -\frac{d\mathbf{B}}{dt}.$$

We will now derive the remaining Maxwell's equation. In Chapter 33, we derived Ampere's law—namely,

$$\oint \mathbf{B} \cdot d\mathbf{s} = \mu_0 I, \qquad (36\text{-}17)$$

where I is the total current passing through any surface bounded by the path. In general, I is the sum of the true or conduction current I_C and the displacement current I_D. Since $\mathbf{B} = \mu_0\mathbf{H}$ for free space, we may write Eq. (36-17) as

$$\oint \mathbf{H} \cdot d\mathbf{s} = I_C + I_D, \qquad (36\text{-}18)$$

which is the final Maxwell equation in integral form. It may be written in terms of current densities as

$$\oint \mathbf{H} \cdot d\mathbf{s} = \oint_A \mathbf{j} \cdot \mathbf{n} dA + \oint_A \frac{d\mathbf{D}}{dt} \cdot \mathbf{n} dA, \qquad (36\text{-}19)$$

where $\mathbf{j}$ is the conduction current density and $d\mathbf{D}/dt$ is the displacement current density. By Stoke's theorem, Eq. (36-19) can be written

$$\oint_A \operatorname{curl} \mathbf{H} \cdot \mathbf{n} dA = \oint_A \mathbf{j} \cdot \mathbf{n} dA + \oint_A \frac{d\mathbf{D}}{dt} \cdot \mathbf{n} dA. \qquad (36\text{-}20)$$

Since this equation holds for any surface, the integrands must be equal, giving the differential form of the above Maxwell equation

$$\operatorname{curl} \mathbf{H} = \mathbf{j} + \frac{d\mathbf{D}}{dt} \qquad (36\text{-}21)$$

or

$$\nabla \times \mathbf{H} = \mathbf{j} + \frac{d\mathbf{D}}{dt}.$$

Equations (36-19) and (36-21) tell us that the displacement current density $d\mathbf{D}/dt$ has the same effect in producing a magnetic field as does the conduction current density **j**. For steady currents, Eqs. (36-19) and (36-21) reduce to

$$\oint \mathbf{H} \cdot d\mathbf{s} = \oint_A \mathbf{j} \cdot \mathbf{n}\, dA \qquad (36\text{-}22)$$

and

$$\text{curl}\, \mathbf{H} = \mathbf{j}, \qquad (36\text{-}23)$$

respectively. Let us consider the space between the plates of a capacitor which is being charged. Here there is no conduction current between the plates but, since the voltage between the plates is changing with time, there is a displacement current between them; Eqs. (36-19) and (36-21) become

$$\oint \mathbf{H} \cdot d\mathbf{s} = \oint_A \frac{d\mathbf{D}}{dt} \cdot \mathbf{n}\, dA \qquad (36\text{-}24)$$

and

$$\text{curl}\, \mathbf{H} = \frac{d\mathbf{D}}{dt}, \qquad (36\text{-}25)$$

respectively.

36-3 THE WAVE EQUATION

One of the most important applications of Maxwell's equations is the derivation of the electromagnetic wave equation. It is possible to derive a wave equation for each of the field vectors **E**, **H**, **B**, and **D**. Let us derive the wave equation for the field vector **E**. We first take the curl of Maxwell's equation for the field vector **E**—namely, curl $\mathbf{E} = -d\mathbf{B}/dt$. This gives

$$\text{curl curl}\, \mathbf{E} = -\frac{d}{dt}\, \text{curl}\, \mathbf{B}. \qquad (36\text{-}26)$$

By the use of the theorem

$$\nabla^2 \mathbf{A} = \text{grad div}\, \mathbf{A} - \text{curl curl}\, \mathbf{A},$$

we can write

$$\text{curl curl}\, \mathbf{E} = \text{grad div}\, \mathbf{E} - \nabla^2 \mathbf{E}. \qquad (36\text{-}27)$$

Substituting Eq. (36-27), Maxwell's equation curl $\mathbf{H} = \mathbf{j} + (d\mathbf{D}/dt)$ and $\mathbf{B} = \mu \mathbf{H}$ into Eq. (36-26), we obtain

$$\text{grad div}\, \mathbf{E} - \nabla^2 \mathbf{E} = -\mu \frac{d\mathbf{j}}{dt} - \mu \frac{\partial^2 \mathbf{D}}{\partial t^2}. \qquad (36\text{-}28)$$

If the field is uniform, the medium obeys Ohm's law, and the susceptibility is independent of the field intensity, then we can show that

$$\text{div}\, \mathbf{E} = 0,$$

$$\mathbf{j} = \sigma \mathbf{E}, \quad \text{and} \quad \mathbf{D} = \epsilon \mathbf{E}.$$

By substitution of these equations into Eq. (36-28), we get

$$\nabla^2 \mathbf{E} - \mu\epsilon \frac{\partial^2 \mathbf{E}}{\partial t^2} - \mu\sigma \frac{\partial \mathbf{E}}{\partial t} = 0. \qquad (36\text{-}29)$$

This is the three-dimensional wave equation for **E** in an absorbing medium. It will be left as a problem to derive the wave equation for each of **H**, **B**, and **D**. These equations are similar to Eq. (36-29). All four equations can be written collectively as

$$\left(\nabla^2 - \mu\epsilon \frac{\partial^2}{\partial t^2} - \mu\sigma \frac{\partial}{\partial t}\right) \begin{matrix} \mathbf{E} \\ \mathbf{H} \\ \mathbf{B} \\ \mathbf{D} \end{matrix} = 0. \qquad (36\text{-}30)$$

In a medium in which the absorption or power loss to the medium is zero, the conductivity (σ) is also zero, and Eq. (36-29) may be written

$$\frac{\partial^2 \mathbf{E}}{\partial t^2} = \frac{1}{\mu\epsilon} \nabla^2 \mathbf{E}. \qquad (36\text{-}31)$$

Let us introduce a quantity v such that

$$v^2 = \frac{1}{\mu\epsilon}$$

and

$$\frac{\partial^2 \mathbf{E}}{\partial t^2} = v^2 \nabla^2 \mathbf{E}. \qquad (36\text{-}32)$$

Dimensionally, Eq. (36-32) in MKS units is

$$\frac{\text{volts}}{\text{meters seconds}^2} = v^2 \frac{\text{volts}}{\text{meters}^3}$$

so that

$$v = \frac{\text{meters}}{\text{second}}.$$

Thus, it appears that v has the dimensions of speed. This speed is characteristic of the medium, being dependent on the constants μ and ϵ. We may, therefore, write

$$v = \frac{1}{\sqrt{\mu\epsilon}}. \qquad (36\text{-}33)$$

The index of refraction n of a medium, as indicated in Eq. (10-26), is

$$n = \frac{c}{v} = \frac{\sqrt{\mu\epsilon}}{\sqrt{\mu_0\epsilon_0}} = \mu_r\epsilon_r, \qquad (36\text{-}34)$$

where c is the speed of electromagnetic radiation in space, μ_r is the relative permeability, and ϵ_r—the relative permittivity of the medium.

We have shown in this section that the electromagnetic field quantities have the attributes of wave phenomena. That is, they satisfy a wave equation and their speed of propagation is directly related to the medium in which they propagate. In the next section, we discuss in more detail the nature and energy of electromagnetic waves.

Example 1. Polystyrene has a relative permeability (μ_r) of 1 and a relative permittivity (ϵ_r) of 2.70. Find the index of refraction for polystyrene, and the speed of electromagnetic waves in polystyrene.

SOLUTION

$$n = \sqrt{\mu_r \epsilon_r} = \sqrt{(2.70)} = 1.65$$

$$\text{Speed} = \frac{c}{\sqrt{\mu_r \epsilon_r}} = \frac{3 \times 10^8 \text{ m/sec}}{1.65}$$

$$= 1.82 \times 10^8 \text{ m/sec.}$$

36-4 THE NATURE AND ENERGY OF ELECTROMAGNETIC WAVES

A system of traveling electric and magnetic fields comprises what is called an electromagnetic wave. There is no microscopic vibration of a material substance taking place, but only an alternation of the electric and magnetic field. An electromagnetic wave can travel through empty space, that is, energy can be transmitted through empty space. This energy transmission through space by an electromagnetic wave must be related to the energy associated with electric and magnetic fields, which we described previously. We saw in Chapters 30 and 33 that the energy per unit volume stored in an electric and magnetic field in empty space is equal to $\frac{1}{2}\epsilon_0 E^2$. and $(1/2\mu_0)B^2 = \frac{1}{2}\mu_0 H^2$, respectively. The simplest form of an electromagnetic wave is one in which the electric field **E** and the magnetic field **H** vary sinusoidally in space and time, are at right angles to one another, have the same frequency, and are in phase. This kind of electromagnetic wave is called a *plane polarized wave.* Here, the total energy per unit volume, W/v, at any point is given by

$$W/v = \frac{1}{2}\epsilon_0 E^2 + \frac{1}{2}\mu_0 H^2.$$

The **E** vector oscillates in one plane and the **H** vector in a plane perpendicular to it, as illustrated in Figure 36-1. Therefore, the energy flow associated with the electromagnetic wave also oscillates and is

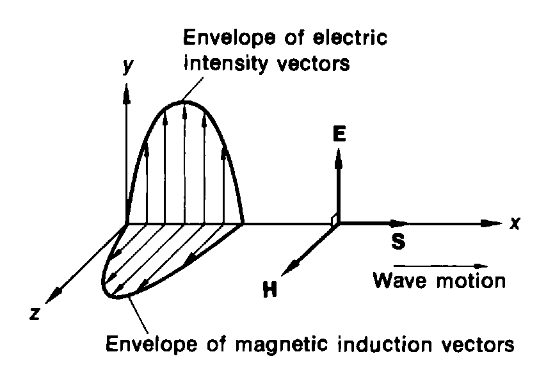


Figure 36-1 A simple electromagnetic wave.

best represented by a vector (**S** in Figure 36-1). The sense of this vector gives the direction of energy flow, and the length of the vector gives the magnitude of the energy flow per unit area, per unit time, across a plane perpendicular to the vector. Although we will not prove it here, the rate at which this energy per unit area is transported is given by

$$\mathbf{S} = \mathbf{E} \times \mathbf{H}.$$

The vector **S** is called the *Poynting vector.* We see that, for energy to be transported in an electromagnetic field, we must have an electric and a magnetic field not parallel to one another. Why? It is worth repeating that **E** and **H** are not independent of one another. The change in **E** gives rise to **H**, and the change in **H** gives rise to **E**. In electrostatics, the Poynting vector is zero since a static charge does not produce a magnetic field.

Example 2. If the earth has a downward electric field of 100 volts/m at the Equator and a northward magnetic field of 40 amperes per meter, what is the Poynting vector?

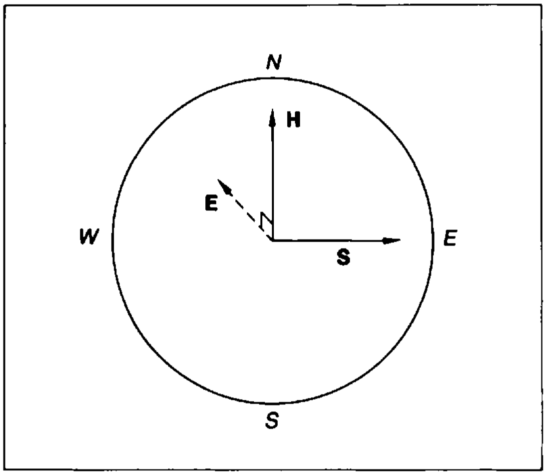
SOLUTION

Since **E** is perpendicular to **H**, the magnitude of **S** is EH. Therefore,

$$S = EH = 100 \text{ volts/m} \times 40 \text{ amp/m}$$

$$= 4000 \text{ watts/m}^2.$$

The direction of **S** by the right-hand screw rule for vector products applied to the diagram following is east.



36-5 THE ELECTROMAGNETIC SPECTRUM

Electromagnetic waves exist with all possible wavelengths, from the longest radio waves of about 10^6 m to the shortest gamma rays of about 10^{-14} m. The shorter wavelengths are usually measured in either angstroms ($1\ \text{Å} = 10^{-10}$ m) or X-ray units ($1\ \text{XU} = 10^{-13}$ m). The electromagnetic spectrum has no definite upper or lower limit, and the spectrum regions overlap. Table 36-1 gives the approximate wavelength limits of the different types of electromagnetic waves and how each is produced, detected, or used. Light is defined as the portion of the electromagnetic spectrum that is capable of affecting the eye's sense of vision and ranges from about 4300 Å to about 6900 Å.

We have previously discussed light in some detail. However, we have said little on the other portions of the electromagnetic spectrum. At the long wavelength end of the spectrum, we have radio waves. An antenna is used to produce the waves. The size of the antenna depends upon the wavelength to be transmitted. Antenna designers have found that antennas about one-quarter the wavelength give efficient transmission. Thus, antennas can vary in size from 100 m in length for long wave radio transmission to a few centimeters for some radar sets operating in the microwave region.

Thermal radiation is produced by heating a solid. When the solid is heated, it first appears a dull red, and its spectrum is almost wholly in the red. As its temperature is increased, its spectrum extends to include all the spectral colors from red to violet. The shorter the wavelength or the higher the frequency of the radiation emitted, the greater the energy. Figure 36-2 is a distribution of the energy in a continuous spectrum of thermal radiation for different temperatures of the source.

Infrared radiation is given out by most ordinary light sources. About 95% of the radiation given out by an ordinary incandescent lamp is in the infrared

Table 36-1 The Electromagnetic Spectrum.

Kind of radiation	Approximate units of wavelength	How produced	How detected or used
Radio waves	$10^6 - 10^{-4}$ m	Oscillating electronic circuit	Radio, radar, television
Thermal radiation	$10^{-3} - 8\times10^{-7}$ m	Hot bodies	Heat effects, thermopile
Infrared light	$5\times10^{-5} - 8\times10^{-7}$ m	Light sources	Heat effects, photography
Visible light	7000–4000 Å	Light sources	Eye, photography, photoelectric cell
Ultraviolet light	4000–50 Å	Electric spark, gas discharge, the sun	Photography, photoelectric cell, ionization, fluorescence
X-rays	50,000–100 XU	Electron impact on solid bodies	Photography, ionization, fluorescence
γ-rays	100–6 XU	Radioactive materials	Ionization, photography
Cosmic rays	0.7–0.4 XU	Bombardment of earth's atmosphere by high energy particles from outer space	Ionization, nuclear disintegration

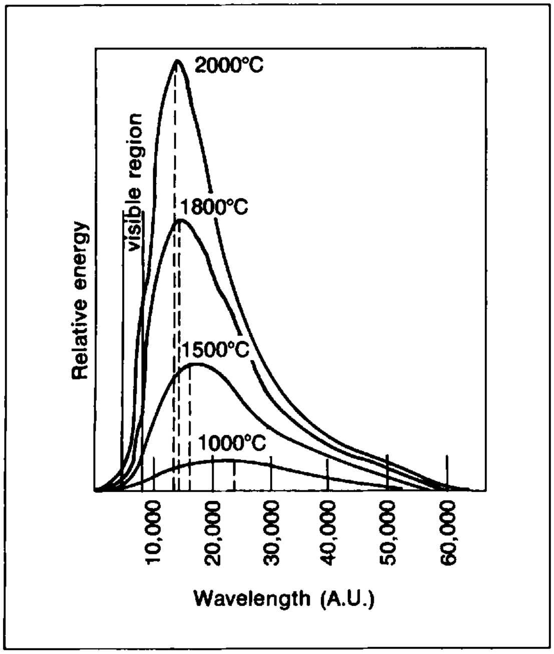


Figure 36-2 Distribution of energy in the continuous spectrum of thermal radiation, for different temperatures of the source.

region. The spectrum of both thermal radiation and infrared radiation can be studied by means of a thermopile. A thermopile consists of a large number of copper-bismuth thermocouple junctions in series, one set of junctions being exposed to the radiation, while the other set is protected. The current produced is proportional to the radiant energy received by the instrument. The spectrum of infrared radiation can also be studied by means of specially prepared photographic plates and coarse-ruled diffraction gratings.

The greatest source of ultraviolet radiation observed on the Earth is the Sun. Ultraviolet radiation in ordinary light has an important effect on animal health. When absorbed by the skin, it causes chemical reactions that are usually beneficial to the health. One of these reactions is the generation of vitamin D in both animal and vegetable tissue. However, overexposure to high frequency ultraviolet radiation can cause burns to the tissue, especially such sensitive tissue as that of the eye. As mentioned in Table 36-1, the ultraviolet portion of the spectrum can be studied by means of photography, ionization, and fluorescence. In fluorescence, the emitted wavelength is longer than the

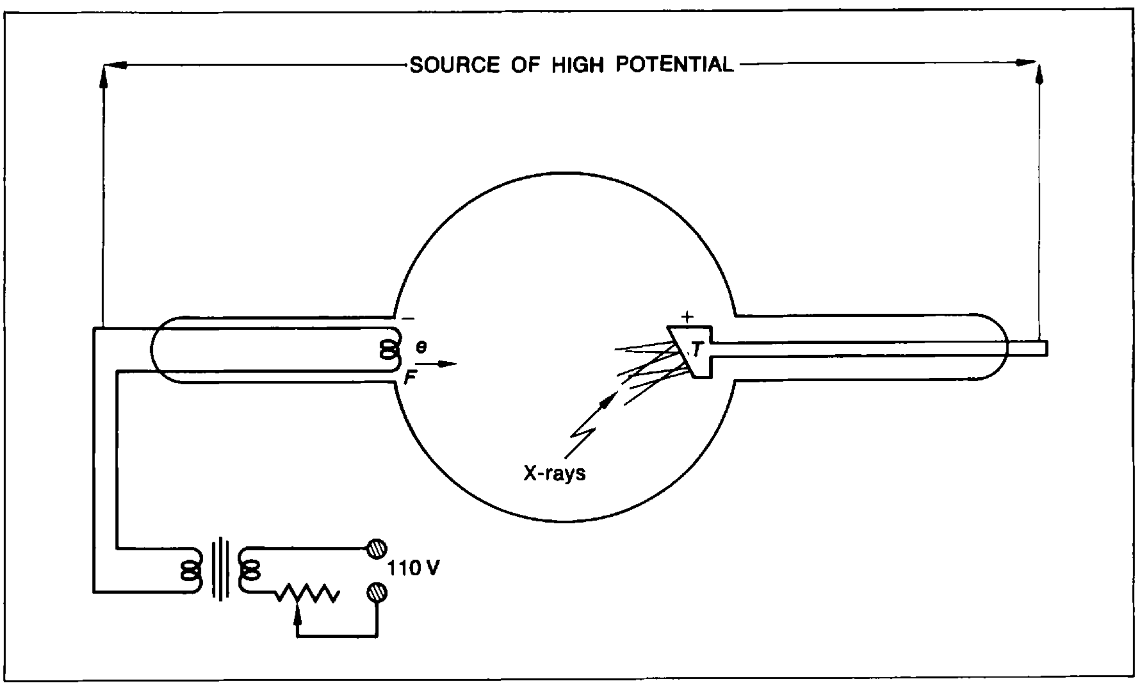


Figure 36-3 X-ray tube.

wavelength absorbed, so that when ultraviolet light falls on a fluorescent material, visible light is emitted permitting the detection and study of the ultraviolet radiation.

X-rays are produced by accelerating electrons in a vacuum tube. Figure 36-3 shows an X-ray tube. Electrons are emitted by the hot filament F, accelerated across the tube striking a target T, usually made of some heavy metal. The electrons are accelerated by a constant high potential placed across the tube. The high energy electrons, on hitting the target, cause it to radiate X-rays. X-rays affect a photographic plate, ionize gases, cause some materials to fluoresce, and can be detected and studied because of these properties. For the lighter elements, the absorption of X-rays is directly proportional to the electron density in the absorbing material. It is this property of materials that makes possible the successful application of X-rays to medicine and industry. For example, if a part of the body is placed between a beam of X-rays and a photographic plate, the photographic plate will show shadows of the bones, since the bones absorb X-ray radiation to a greater extent than the flesh.

Gamma (γ-) rays are emitted by radioactive materials and have a greater penetrating power than ordinary X-rays. They affect a photographic plate and will ionize gases like X-rays. The γ-rays of radioactive elements are very often used in treatment of cancer, although the use of very short wavelength X-rays for such treatments has increased.

Cosmic rays are present all over the Earth and come from outer space. In spite of many experimental investigations, the origin of cosmic rays is still not completely understood. When they reach the Earth's surface, the rays consist of electromagnetic radiation of shorter wavelength than γ-rays along with high energy charged particles. They can be detected by such devices as ionization chambers and Geiger counters.

PROBLEMS

1. A parallel plate capacitor of area A is connected to a voltage source that increases slowly and linearly with time. Find an expression for
   (a) the displacement current density and
   (b) the displacement current at a point between the plates.
2. Starting from Maxwell's equations, show that the electric field intensity due to a point charge q is given by
$$\mathbf{E} = \frac{1}{4\pi\epsilon_0}\frac{q}{r^2}\hat{r}_0.$$
3. Paraffin has a relative permittivity ϵ_r of 2 and a relative permeability μ_r of 1. Find:
   (a) the index of refraction for paraffin and
   (b) the speed of electromagnetic waves in paraffin.
4. Write out Maxwell's equations in terms of **E** and **H** only for a non-homogeneous medium in which ϵ and μ are functions of the coordinates.
5. Develop the wave equation in **E** for a plane wave traveling in the y-direction with **E** in the z-direction.
6. Given a plane monochromatic wave traveling in a linear, isotropic homogeneous medium, show that the electric and magnetic energy densities are equal.
7. A plane electromagnetic wave, having an electric field of magnitude $\mathbf{E} = 100$ volts/m, is traveling in a non-conducting medium of relative permittivity 9 and relative permeability unity. Find:
   (a) the velocity of the wave,
   (b) the index of refraction of the medium, and
   (c) the energy density of the wave in the medium. Hint: Refer to Problem 6.
8. A low frequency radio wave is incident upon a body of distilled water of relative permittivity 9 and relative permeability unity. Find:
   (a) the speed of the wave in the water and
   (b) the index of refraction of the water.

9. Derive Snell's law for light refraction through the consideration of polarized plane light waves incident obliquely at a plane dielectric interface.
10. A plane radio wave travels in the x-direction and is plane-polarized with its electric field vector in the z-direction. Its frequency is 10^6 cps. The average power propagation by the wave is 16 watts/m^2. Find:
    (a) the wavelength of the wave and
    (b) the effective values of **E** and **H** for this wave.
11. Calculate the Poynting vector for a plane wave in free space having an effective electric field of 50 microvolts/m.
12. A wire of diameter 2 mm, resistivity 10^{-3} ohm/m, carries a current of 5 amp. Calculate the component of the Poynting vector (S) perpendicular to the wire.
13. Lines of magnetic induction are always closed. They neither originate nor terminate at any point, and as many flux lines enter a volume as leave it. Assuming the preceding statement is true, prove Maxwell's equation

$$\nabla \cdot \mathbf{B} = 0.$$

14. A radar transmitter emits a 10 kilowatt pulse lasting for 10^{-6} seconds. At a certain distance from the transmitter, the area of the wavefront is 0.5 km^2. Find the average energy density within the pulse.

Appendices

APPENDIX I VECTOR ANALYSIS

1 Vector Quantities and Vectors

Vector quantities have magnitude and direction, for example, force, displacement, electric field, etc. The complete specification of a vector quantity requires (1) a unit for the quantity, (2) a number stating how many times the unit is contained in the quantity, and (3) a statement of direction.

Vectors are directed line segments or arrows used to represent vector quantities. The length of the line segment gives the magnitude and its sense the direction of the vector quantity.

2 Addition and Subtraction of Vectors

Vectors may be added by two methods: the polygon method and the method of components, which will be discussed briefly in the next section. In the polygon method, the tail of the second vector is joined to the head of the first vector, the tail of the third to the head of the second, and so on, keeping their direction as given. The sum of the vectors is a vector drawn from the tail of the first vector to the head of the last vector. Figure 1 represents the addition of two vectors **A** and **B**. **C** is the sum of the two vectors

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

or

$$\mathbf{B} + \mathbf{A} = \mathbf{C}. \quad (1)$$

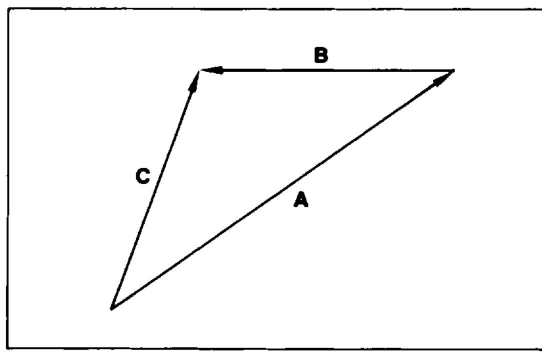


Figure 1 Vector addition.

Therefore, the sum of the two vectors is *commutative.*

Subtraction of the vector **B** from the vector **A** is defined as the addition of the negative vector $-\mathbf{B}$ to **A**. Thus,

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

The sum of any number of vectors is associative. That is, vectors may be added in any desired manner.

$$\mathbf{R} = (\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C}) = (\mathbf{A} + \mathbf{C}) + \mathbf{B}. \quad (2)$$

3 Components of a Vector

Vectors may be conveniently represented in Cartesian coordinates by means of the unit vector concept. A unit vector is one of unit magnitude.

Let the vectors **i**, **j**, **k** be unit vectors along with x-, y-, z-axes, respectively. A vector **A** can be considered to be the sum of the vectors $\mathbf{i}A_x$, $\mathbf{j}A_y$, and $\mathbf{k}A_z$, each of which lies along one of these axes.

$$\mathbf{A} = \mathbf{i}A_x + \mathbf{j}A_y + \mathbf{k}A_z, \tag{3}$$

where A_x, A_y, and A_z are known as the Cartesian components of **A**. The sum of vectors **A**, **B**, **C** in terms of their components can be written

$$\mathbf{A} + \mathbf{B} + \mathbf{C} = \mathbf{i}(A_x + B_x + C_x) + \mathbf{j}(A_y + B_y + C_y) + \mathbf{k}(A_z + B_z + C_z), \tag{4}$$

since the components along the x-axis add directly, as do those along the y- and z-axis. The magnitude of the length of a vector **A** is

$$A = \sqrt{A_x^2 + A_y^2 + A_z^2}. \tag{5}$$

Proof:

Refer to Figure 2.

$$A^2 = (OP)^2 + A_z^2,$$

$$(OP)^2 = A_x^2 + A_y^2,$$

$$A^2 = A_x^2 + A_y^2 + A_z^2$$

or

$$A = \sqrt{A_x^2 + A_y^2 + A_z^2}.$$

4 The Products of Two Vectors

(a) Scalar (Dot) Product. The scalar (dot) product is the product of the magnitude of one vector by the magnitude of the projection of another vector upon it. The result is a scalar, and the process is commutative; that is,

$$\mathbf{A} \cdot \mathbf{B} = \mathbf{B} \cdot \mathbf{A}. \tag{6}$$

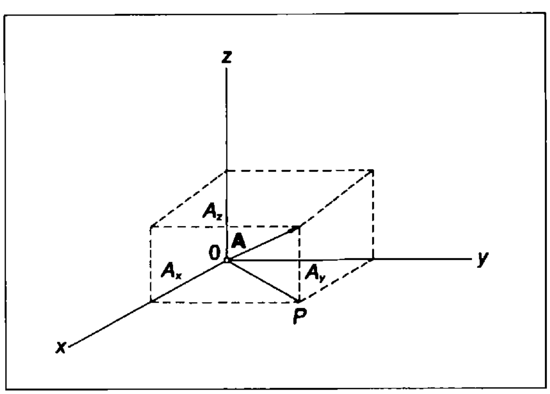


Figure 2 Cartesian components of a vector.

Consider the scalar product of the two vectors **A** and **B** as shown in Figure 3.

$$\mathbf{A} \cdot \mathbf{B} = \mathbf{B} \cdot \mathbf{A} = AB \cos \theta. \tag{7}$$

If $\mathbf{A} \cdot \mathbf{B} = 0$, then $\mathbf{A} = 0$, $\mathbf{B} = 0$, or **A** is perpendicular to **B**. If $\mathbf{A} \cdot \mathbf{B} = AB$, then **A** and **B** are parallel.

The scalar products of the unit Cartesian vectors obey the relations

$$\mathbf{i} \cdot \mathbf{j} = \mathbf{j} \cdot \mathbf{k} = \mathbf{k} \cdot \mathbf{i} = 0$$

$$\mathbf{i} \cdot \mathbf{i} = \mathbf{j} \cdot \mathbf{j} = \mathbf{k} \cdot \mathbf{k} = 1. \tag{8}$$

The scalar product of two vectors is the sum of the products of the corresponding components. Consider the two vectors **A** and **B** written in terms of their components:

$$\mathbf{A} = \mathbf{i}A_x + \mathbf{j}A_y + \mathbf{k}A_z \quad \text{and} \quad \mathbf{B} = \mathbf{i}B_x + \mathbf{j}B_y + \mathbf{k}B_z.$$

Multiplying out in the usual way, and then substituting the values of the products of the unit vectors, the scalar product becomes

$$\mathbf{A} \cdot \mathbf{B} = A_xB_x + A_yB_y + A_zB_z. \tag{9}$$

The scalar product is distributive, that is,

$$\mathbf{A} \cdot (\mathbf{B} + \mathbf{C}) = \mathbf{A} \cdot \mathbf{B} + \mathbf{A} \cdot \mathbf{C}, \tag{10}$$

which we shall not prove here since its proof is available in any book on vector analysis.

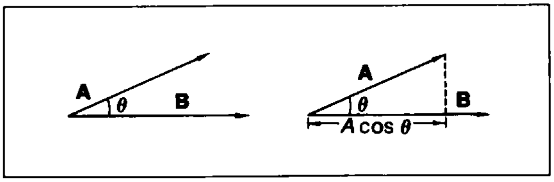


Figure 3 Scalar product of two vectors.

(b) Vector (Cross) Product. The vector, or cross product, of two vectors **A** and **B** is written **A** × **B**. It is defined as the vector perpendicular to the plane determined by **A** and **B**, as illustrated in Figure 4, and is equal in magnitude to $AB \sin \theta$, where θ is the angle between **A** and **B**, that is,

$$\mathbf{V} = \mathbf{A} \times \mathbf{B} = \mathbf{n}AB \sin \theta, \tag{11}$$

where **n** is a unit vector in the direction of **V**. It is so directed that a right-handed rotation about **V** through an angle θ carries **A** into **B**. **A** × **B** is equal in magnitude to the area of the parallelogram. **A** × **B**

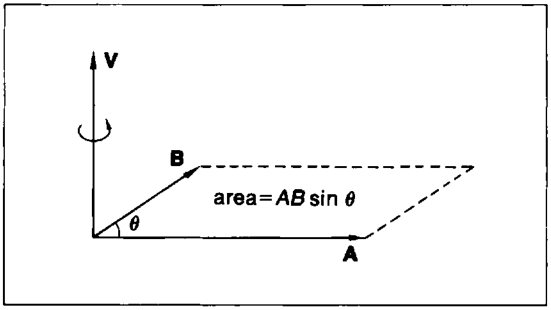


Figure 4 Vector product of two vectors.

is non-commutative

$$\mathbf{A}\times\mathbf{B} = -\mathbf{B}\times\mathbf{A}, \quad (12)$$

since rotation from **B** to **A** is opposite to that from **A** to **B**.

If **A** and **B** are parallel, $\theta = 0°$ or $180°$ and $\mathbf{A}\times\mathbf{B} = 0$. Conversely, if the vector product is zero, one of the vectors is zero or else the two are parallel. If **A** and **B** are perpendicular, $\mathbf{A}\times\mathbf{B} = \mathbf{n}AB$, in which case the two vectors and their product are mutually at right angles.

The vector products of the unit Cartesian vectors are seen to obey the following relations:

$$\begin{aligned}
\mathbf{i}\times\mathbf{i} &= \mathbf{j}\times\mathbf{j} = \mathbf{k}\times\mathbf{k} = 0,\\
\mathbf{i}\times\mathbf{j} &= -\mathbf{j}\times\mathbf{i} = \mathbf{k},\\
\mathbf{j}\times\mathbf{k} &= -\mathbf{k}\times\mathbf{j} = \mathbf{i},\\
\mathbf{k}\times\mathbf{i} &= -\mathbf{i}\times\mathbf{k} = \mathbf{j}.
\end{aligned} \quad (13)$$

The vector (cross) product of **A** and **B** may be expressed in terms of their components and the unit vectors.

$$\begin{aligned}
\mathbf{A}\times\mathbf{B} &= (\mathbf{i}A_x + \mathbf{j}A_y + \mathbf{k}A_z)\times(\mathbf{i}B_x + \mathbf{j}B_y + \mathbf{k}B_z),\\
&= (A_yB_z - A_zB_y)\mathbf{i} + (A_zB_x - A_xB_z)\mathbf{j}\\
&\quad + (A_xB_y - A_yB_x)\mathbf{k}.
\end{aligned} \quad (14)$$

As a determinant, this can be written

$$\mathbf{A}\times\mathbf{B} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ A_x & A_y & A_z \\ B_x & B_y & B_z \end{vmatrix}. \quad (15)$$

The vector product, like the scalar product, is distributive; that is,

$$\mathbf{A}\times(\mathbf{B}+\mathbf{C}) = \mathbf{A}\times\mathbf{B} + \mathbf{A}\times\mathbf{C}. \quad (16)$$

5 The Products of Three Vectors

(a) Scalar Triple Product [A · (B × C)]. $\mathbf{A}\cdot(\mathbf{B}\times\mathbf{C})$ is a scalar and represents the volume of a parallelopiped as illustrated in Figure 5. Any face of the solid may be taken as the base. The three equivalent expresssions for the volume are:

$$\mathbf{A}\cdot(\mathbf{B}\times\mathbf{C}) = \mathbf{B}\cdot(\mathbf{C}\times\mathbf{A}) = \mathbf{C}\cdot(\mathbf{A}\times\mathbf{B}). \quad (17)$$

Since the order of the terms in a scalar product is immaterial, these relations are equivalent to

$$(\mathbf{B}\times\mathbf{C})\cdot\mathbf{A} = (\mathbf{C}\times\mathbf{A})\cdot\mathbf{B} = (\mathbf{A}\times\mathbf{B})\cdot\mathbf{C}$$

or

$$(\mathbf{A}\times\mathbf{B})\cdot\mathbf{C} = (\mathbf{B}\times\mathbf{C})\cdot\mathbf{A} = (\mathbf{C}\times\mathbf{A})\cdot\mathbf{B}. \quad (18)$$

Therefore, in the scalar triple product, the dot and cross can be interchanged without changing the value of the result.

The sign of the scalar triple product is unchanged as long as the cyclic order of the factors is unchanged. For every change in cyclic order, a minus sign is introduced, for example,

$$\mathbf{A}\cdot(\mathbf{B}\times\mathbf{C}) = -\mathbf{A}\cdot(\mathbf{C}\times\mathbf{B}).$$

Therefore, three vectors have six identical scalar triple products.

The scalar triple product may be expressed in determinant form. Consider the product $\mathbf{A}\cdot(\mathbf{B}\times\mathbf{C})$.

$$\mathbf{B}\times\mathbf{C} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ B_x & B_y & B_z \\ C_x & C_y & C_z \end{vmatrix},$$

$$\mathbf{A} = \mathbf{i}A_x + \mathbf{j}A_y + \mathbf{k}A_z,$$

$$\mathbf{A}\cdot(\mathbf{B}\times\mathbf{C}) = \begin{vmatrix} A_x & A_y & A_z \\ B_x & B_y & B_z \\ C_x & C_y & C_z \end{vmatrix}. \quad (19)$$

When three vectors are in a plane, the volume of the parallelopiped formed by them is zero. Hence,

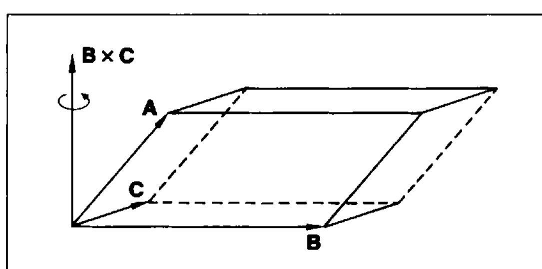


Figure 5 The scalar triple product.

the condition for three vectors to be coplanar is that their scalar triple product equals zero.

(b) Vector Triple Product [A × (B × C)]. $\mathbf{A}\times(\mathbf{B}\times\mathbf{C})$ is a vector. The sign of the product changes every time the order of the factors $\mathbf{A}$ and $\mathbf{B}\times\mathbf{C}$ is changed in $\mathbf{A}\times(\mathbf{B}\times\mathbf{C})$ or whenever the order of the factors $\mathbf{B}$ and $\mathbf{C}$ is changed in $(\mathbf{B}\times\mathbf{C})$. The vector $\mathbf{B}\times\mathbf{C}$ is perpendicular to the plane containing $\mathbf{B}$ and $\mathbf{C}$. The vector $\mathbf{A}\times(\mathbf{B}\times\mathbf{C})$ is perpendicular to the plane containing $\mathbf{A}$ and $(\mathbf{B}\times\mathbf{C})$ and is, therefore, in the same plane as $\mathbf{B}$ and $\mathbf{C}$, as illustrated in Figure 6.

The product $\mathbf{A}\times(\mathbf{B}\times\mathbf{C})$ can be reduced to a simpler form as follows: Let the x-axis be along vector $\mathbf{B}$, the y-axis in the plane $\mathbf{B}$ and $\mathbf{C}$, and z-axis along $\mathbf{B}\times\mathbf{C}$, as in Figure 6.

$$\mathbf{A} = \mathbf{i}A_x + \mathbf{j}B_y + \mathbf{k}B_z, \tag{20}$$

$$\mathbf{B} = \mathbf{i}B_x, \tag{21}$$

$$\mathbf{C} = \mathbf{i}C_x + \mathbf{j}C_y, \tag{22}$$

$$\mathbf{B}\times\mathbf{C} = \mathbf{k}B_xC_y, \tag{23}$$

$$\mathbf{A}\times(\mathbf{B}\times\mathbf{C}) = (\mathbf{i}A_x + \mathbf{j}A_y + \mathbf{k}A_z)\times(\mathbf{k}B_xC_y)$$

$$= \mathbf{i}A_yB_xC_y - \mathbf{j}A_xB_xC_y. \tag{24}$$

But

$$\mathbf{A}\cdot\mathbf{C} = (\mathbf{i}A_x + \mathbf{j}A_y + \mathbf{k}A_z)\cdot(\mathbf{i}C_x + \mathbf{j}C_y)$$

$$= A_xC_x + A_yC_y \tag{25}$$

and

$$\mathbf{A}\cdot\mathbf{B} = (\mathbf{i}A_x + \mathbf{j}A_y + \mathbf{k}A_z)\cdot(\mathbf{i}B_x)$$

$$= A_xB_x. \tag{26}$$

From Eqs. (24), (21), and (22),

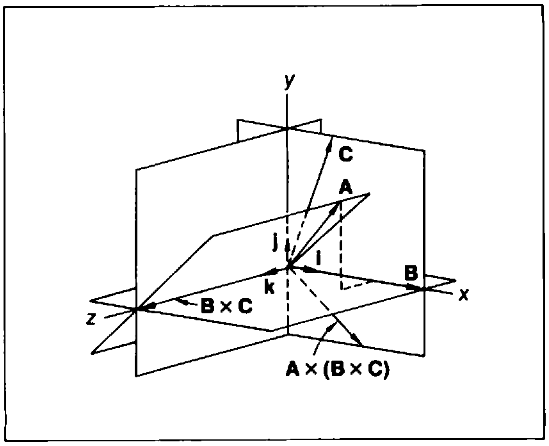


Figure 6 The vector triple product.

$$\mathbf{A}\times(\mathbf{B}\times\mathbf{C}) = A_yC_y\mathbf{B} - \mathbf{j}A_xB_xC_y$$

$$= (A_yC_y + A_xC_x)\mathbf{B} - \mathbf{i}A_xB_xC_x - \mathbf{j}A_xB_xC_y$$

$$= (A_xC_x + A_yC_y)\mathbf{B} - A_xB_x(\mathbf{i}C_x + \mathbf{j}C_y).$$

From Eqs. (22), (25), and (26),

$$\mathbf{A}\times(\mathbf{B}\times\mathbf{C}) = (\mathbf{A}\cdot\mathbf{C})\mathbf{B} - (\mathbf{A}\cdot\mathbf{B})\mathbf{C}. \tag{27}$$

Each term of the product involves the external factor in a scalar product first with the extreme and then with the middle factor.

6 Vector Integration

(a) Line Integral. Suppose ds is an element of length along a curve, where $s = s(t)$ is the equation of a curve. Let $\mathbf{V}$ denote a vector at an angle θ with that of the length element, as illustrated in Figure 7. It is possible to form the integrals

$$\int \mathbf{V}\cdot d\mathbf{s} = \int V\cos\theta\, ds \tag{28}$$

and

$$\int \mathbf{V}\times d\mathbf{s} = \mathbf{n}\int V\sin\theta\, ds, \tag{29}$$

each being a line integral along the curve C. Such integrals occur frequently, for example, if $\mathbf{V}$ is a force on a particle and ds an element of its displacement, Eq. (28) denotes the work done on the particle; while if $\mathbf{V}$ is a force on an extended body free to rotate about a given axis and ds is an element of its moment arm, Eq. (29) denotes the torque on the body. Also, the integral $\int S ds$ could be formed where S is a scalar function.

(b) Surface and Volume Integrals. Consider a surface A drawn in any vector field, as illustrated in Figure 8. Let $\mathbf{V}$ be the value of the vector at an

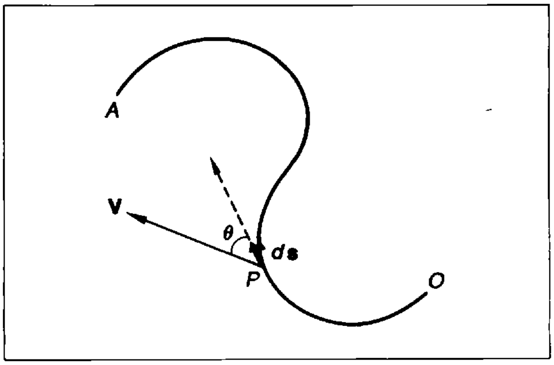


Figure 7 Tangential line integral of a vector.

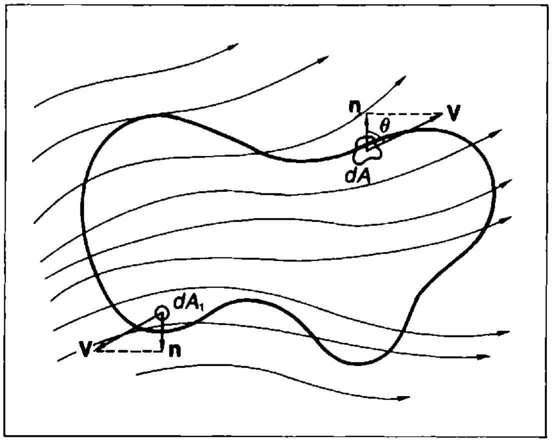


Figure 8 Normal surface integral of a vector.

element of surface dA, $\mathbf{n}$ a unit outward vector normal to dA, and θ the angle between $\mathbf{n}$ and $\mathbf{V}$. The total outward flux is then given by

$$\phi = \int_A \mathbf{V}\cdot\mathbf{n}\,dA = \int_A V\cos\theta\,dA. \qquad (30)$$

The outward flux or flow is positive if it is in the direction of the outward unit normal vector $\mathbf{n}$, and negative—if in the opposite direction. Other surface integrals that may be formed are

$$\int_A S\,dA, \quad \int_A S\mathbf{n}\,dA \quad \text{and} \quad \int_A \mathbf{V}\times\mathbf{n}\,dA, \qquad (31)$$

where S is a scalar function.

Consider a closed surface enclosing a volume v. The following volume integrals may be formed

$$\int_v \mathbf{V}\,dv \quad \text{and} \quad \int_v S\,dv, \qquad (32)$$

where $\mathbf{V}$ and S are vector and scalar fields, respectively.

7 Differentiation

(a) Differentiation of Vectors. If $\mathbf{V}$ is a vector function of a scalar variable t, then when t changes from t to $t+\Delta t$, $\mathbf{V}$ becomes $\mathbf{V}+\Delta\mathbf{V}$, and $\Delta\mathbf{V}/\Delta t$ = average rate of change of $\mathbf{V}$ with t. Also,

$$\lim_{\Delta t\to 0}\frac{\Delta\mathbf{V}}{\Delta t} = \frac{d\mathbf{V}}{dt},$$

which is the derivative of $\mathbf{V}$ with respect to t. When $\mathbf{V}$ is expressed in rectangular coordinates,

$$\mathbf{V} = \mathbf{i}V_x + \mathbf{j}V_y + \mathbf{k}V_z.$$

Since V_x, V_y, V_z are functions of t and $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ are unit vectors,

$$\frac{d\mathbf{V}}{dt} = \mathbf{i}\frac{dV_x}{dt} + \mathbf{j}\frac{dV_y}{dt} + \mathbf{k}\frac{dV_z}{dt}, \qquad (33)$$

and similarly for higher derivatives.

(b) Differentiation of Sums and Products. If $\mathbf{V} = \mathbf{A}+\mathbf{B}$, which are all functions of t,

$$\frac{d\mathbf{V}}{dt} = \frac{d}{dt}(\mathbf{A}+\mathbf{B}) = \frac{d\mathbf{A}}{dt} + \frac{d\mathbf{B}}{dt}, \qquad (34)$$

so we can say that differentiation of vector sums is distributive.

Consider the scalar product

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

and

$$\frac{dS}{dt} = \frac{d\mathbf{A}}{dt}\cdot\mathbf{B} + \mathbf{A}\cdot\frac{d\mathbf{B}}{dt}. \qquad (35)$$

Since the order of the factors can be interchanged in Eq. (35), the differentiation of scalar products gives a result that is commutative.

Consider the vector product

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

and

$$\frac{d\mathbf{V}}{dt} = \frac{d\mathbf{A}}{dt}\times\mathbf{B} + \mathbf{A}\times\frac{d\mathbf{B}}{dt}. \qquad (36)$$

Since a change in order of the factors changes the sign of the result, the differentiation of vector products gives a result that is non-commutative.

(c) Partial Differentiation. Consider a vector as a function of more than one scalar independent variable. Let a vector $\mathbf{V}$ be a function of the Cartesian coordinates x, y, and z. If y and z remain constant while x changes, the partial derivative $\partial\mathbf{V}/\partial x$ denotes the rate of change of $\mathbf{V}$ with respect to x.

If x, y, and z change simultaneously by dx, dy, and dz, the total change or total differential is

$$d\mathbf{V} = \frac{\partial\mathbf{V}}{\partial x}dx + \frac{\partial\mathbf{V}}{\partial y}dy + \frac{\partial\mathbf{V}}{\partial z}dz,$$

which may be written

$$d\mathbf{V} = \left[\frac{\partial}{\partial x}dx + \frac{\partial}{\partial y}dy + \frac{\partial}{\partial z}dz\right]\mathbf{V}. \qquad (37)$$

If $\mathbf{r} = \mathbf{i}x + \mathbf{j}y + \mathbf{k}z$ is the radius vector from the origin, then

$$d\mathbf{r} = \mathbf{i}dx + \mathbf{j}dy + \mathbf{k}dz. \qquad (38)$$

We define an operator ∇ by

$$\nabla = \mathbf{i}\frac{\partial}{\partial x} + \mathbf{j}\frac{\partial}{\partial y} + \mathbf{k}\frac{\partial}{\partial z}. \tag{39}$$

By the use of this operator, Eq. (39), and Eq. (38), Eq. (37) may be rewritten as

$$d\mathbf{V} = (\nabla \cdot d\mathbf{r})\mathbf{V}. \tag{40}$$

The vector differential operator ∇ is of great physical significance in vector analysis. It is useful in defining three important quantities—namely, the gradient, divergence, and curl, which will now be discussed.

8 Gradient, Divergence, and Curl

(a) The Gradient of a Scalar Field. In studying the field of some scalar function S we sometimes need to know the rate of change of S in going from one point to another. This information is obtained with the aid of a vector called the gradient of the field.

When $S(x, y, z)$ is a scalar point function, then the vector

$$\mathbf{i}\frac{\partial S}{\partial x} + \mathbf{j}\frac{\partial S}{\partial y} + \mathbf{k}\frac{\partial S}{\partial z}$$

is called the gradient of S, grad S, or ∇S. Symbolically, in Cartesian coordinates,

$$\nabla S = \mathbf{i}\frac{\partial S}{\partial x} + \mathbf{j}\frac{\partial S}{\partial y} + \mathbf{k}\frac{\partial S}{\partial z}. \tag{41}$$

Gradient S or ∇S is a vector perpendicular to a level or equipotential surface at the point x, y, z, a vector which is equal in magnitude to the fastest rate of increase of S with distance and points in the direction of the fastest rate of increase of S.

(b) The Divergence of a Vector Field. If $\mathbf{V}$ represents a vector field whose components, together with their partial derivatives, are continuous, a useful scalar field can be derived from it by the scalar multiplication of ∇ and $\mathbf{V}$. The scalar quantity resulting from $\nabla \cdot \mathbf{V}$ is known as the divergence of $\mathbf{V}$ or div $\mathbf{V}$. Symbolically, in Cartesian coordinates,

$$\nabla \cdot \mathbf{V} = \left(\mathbf{i}\frac{\partial}{\partial x} + \mathbf{j}\frac{\partial}{\partial y} + \mathbf{k}\frac{\partial}{\partial z}\right) \cdot (\mathbf{i}V_x + \mathbf{j}V_y + \mathbf{k}V_z)$$

or

$$\nabla \cdot \mathbf{V} = \frac{\partial V_x}{\partial x} + \frac{\partial V_y}{\partial y} + \frac{\partial V_z}{\partial z}. \tag{42}$$

The divergence of a vector field at a point gives the rate of flow per unit volume into or out of a volume element. If the divergence is positive, the flow is away from the point; if it is negative, the flow is toward the point. In the case of thermal, electric, or magnetic fields, the existence of a divergence means the presence of a source or sink of flux at the point. When the divergence is zero everywhere, there are no sources or sinks and the flux leaving equals the flux entering any volume element.

(c) The Curl of a Vector Field. The curl of a vector field $\mathbf{V}$ is the vector multiplication of ∇ and $\mathbf{V}$. The vector quantity resulting from $\nabla \times \mathbf{V}$ is known as the curl of $\mathbf{V}$ or curl $\mathbf{V}$. Symbolically, in Cartesian coordinates,

$$\nabla \times \mathbf{V} = \left(\mathbf{i}\frac{\partial}{\partial x} + \mathbf{j}\frac{\partial}{\partial y} + \mathbf{k}\frac{\partial}{\partial z}\right) \times (\mathbf{i}V_x + \mathbf{j}V_y + \mathbf{k}V_z),$$

$$= \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ V_x & V_y & V_z \end{vmatrix}$$

or

$$\nabla \times \mathbf{V} = \mathbf{i}\left(\frac{\partial V_z}{\partial y} - \frac{\partial V_y}{\partial z}\right) + \mathbf{j}\left(\frac{\partial V_x}{\partial z} - \frac{\partial V_z}{\partial x}\right) + \mathbf{k}\left(\frac{\partial V_y}{\partial x} - \frac{\partial V_x}{\partial y}\right). \tag{43}$$

In order to discuss the physical meaning of the curl of a vector, we shall first define the mathematical quantity *circulation*. Circulation is the line integral of a vector field along a given path. If the vector field is $\mathbf{V}$, its circulation about a closed path is

$$\oint \mathbf{V} \cdot d\mathbf{s},$$

where $d\mathbf{s}$ is an element of length along the path. The curl of a vector field at a point is the limit approached by the ratio

$$\frac{\text{circulation about a small closed path}}{\text{area of surface bounded by that path}}$$

as the path is allowed to shrink to a point. The curl is thus the limiting value of circulation per unit area.

9 More Vector Operators

(a) Div Grad. If S is a scalar function of position in space,

$$\text{grad } S = \nabla S = \mathbf{i}\frac{\partial S}{\partial x} + \mathbf{j}\frac{\partial S}{\partial y} + \mathbf{k}\frac{\partial S}{\partial z},$$

$$\begin{aligned}\text{div grad } S &= \nabla \cdot \nabla S \\ &= \left(\mathbf{i}\frac{\partial}{\partial x} + \mathbf{j}\frac{\partial}{\partial y} + \mathbf{k}\frac{\partial}{\partial z}\right) \\ &\quad \cdot \left(\mathbf{i}\frac{\partial S}{\partial x} + \mathbf{j}\frac{\partial S}{\partial y} + \mathbf{k}\frac{\partial S}{\partial z}\right), \\ &= \frac{\partial^2 S}{\partial x^2} + \frac{\partial^2 S}{\partial y^2} + \frac{\partial^2 S}{\partial z^2},\end{aligned}$$

and the operation

$$\text{div grad} = \nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}, \qquad (44)$$

∇^2 is known as Laplace's operator. ∇^2 can also operate on a vector, if

$$\mathbf{V} = \mathbf{i}V_x + \mathbf{j}V_y + \mathbf{k}V_z,$$

$$\nabla^2\mathbf{V} = \nabla^2\mathbf{i}V_x + \nabla^2\mathbf{j}V_y + \nabla^2\mathbf{k}V_z.$$

This result is important in mechanics and in electricity and magnetism.

(b) Curl Grad. If S is a scalar point function of position in space,

$$\text{grad } S = \nabla S = \mathbf{i}\frac{\partial S}{\partial x} + \mathbf{j}\frac{\partial S}{\partial y} + \mathbf{k}\frac{\partial S}{\partial z}$$

is a vector, and it is possible to calculate its curl.

$$\begin{aligned}\text{curl grad } S &= \nabla \times \nabla S \\ &= \left(\mathbf{i}\frac{\partial}{\partial x} + \mathbf{j}\frac{\partial}{\partial y} + \mathbf{k}\frac{\partial}{\partial z}\right) \\ &\quad \times \left(\mathbf{i}\frac{\partial S}{\partial x} + \mathbf{j}\frac{\partial S}{\partial y} + \mathbf{k}\frac{\partial S}{\partial z}\right) \\ &= \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ \frac{\partial S}{\partial x} & \frac{\partial S}{\partial y} & \frac{\partial S}{\partial z} \end{vmatrix} = 0. \qquad (45)\end{aligned}$$

The result illustrates the fact that the line integral of any vector resulting from the gradient of a scalar (for example, grad S) around a closed path is zero.

(c) Grad Div. If $\mathbf{V}$ is a vector field, div $\mathbf{V}$ is a scalar field which, therefore, has a gradient.

$$\begin{aligned}\text{grad div } \mathbf{V} &= \nabla(\nabla \cdot \mathbf{V}) \\ &= \left(\mathbf{i}\frac{\partial}{\partial x} + \mathbf{j}\frac{\partial}{\partial y} + \mathbf{k}\frac{\partial}{\partial z}\right)\left(\frac{\partial V_x}{\partial x} + \frac{\partial V_y}{\partial y} + \frac{\partial V_z}{\partial z}\right) \\ &= \mathbf{i}\left(\frac{\partial^2 V_x}{\partial x^2} + \frac{\partial^2 V_y}{\partial x \partial y} + \frac{\partial^2 V_z}{\partial x \partial z}\right) \\ &\quad + \mathbf{j}\left(\frac{\partial^2 V_x}{\partial x \partial y} + \frac{\partial^2 V_y}{\partial y^2} + \frac{\partial^2 V_z}{\partial y \partial z}\right) \\ &\quad + \mathbf{k}\left(\frac{\partial^2 V_x}{\partial x \partial z} + \frac{\partial^2 V_y}{\partial y \partial z} + \frac{\partial^2 V_z}{\partial z^2}\right). \qquad (46)\end{aligned}$$

(d) Div Curl

$$\text{curl } \mathbf{V} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ V_x & V_y & V_z \end{vmatrix} = \left(\frac{\partial V_z}{\partial y} - \frac{\partial V_y}{\partial z}\right)\mathbf{i}$$

$$+ \left(\frac{\partial V_x}{\partial z} - \frac{\partial V_z}{\partial x}\right)\mathbf{j} + \left(\frac{\partial V_y}{\partial x} - \frac{\partial V_x}{\partial y}\right)\mathbf{k},$$

and

$$\begin{aligned}\text{div curl } \mathbf{V} &= \frac{\partial}{\partial x}\left(\frac{\partial V_z}{\partial y} - \frac{\partial V_y}{\partial z}\right) + \frac{\partial}{\partial y}\left(\frac{\partial V_x}{\partial z} - \frac{\partial V_z}{\partial x}\right) \\ &\quad + \frac{\partial}{\partial z}\left(\frac{\partial V_y}{\partial x} - \frac{\partial V_x}{\partial y}\right) \\ &= \frac{\partial^2 V_z}{\partial x \partial y} - \frac{\partial^2 V_y}{\partial x \partial z} + \frac{\partial^2 V_x}{\partial y \partial x} \\ &\quad - \frac{\partial^2 V_z}{\partial y \partial x} + \frac{\partial^2 V_y}{\partial z \partial x} - \frac{\partial^2 V_x}{\partial x \partial y} = 0.\end{aligned}$$

(e) Curl Curl.

$$\text{curl } \mathbf{V} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ V_x & V_y & V_z \end{vmatrix} = \mathbf{i}\left(\frac{\partial V_z}{\partial y} - \frac{\partial V_y}{\partial z}\right)$$

$$+ \mathbf{j}\left(\frac{\partial V_x}{\partial z} - \frac{\partial V_z}{\partial x}\right) + \mathbf{k}\left(\frac{\partial V_y}{\partial x} - \frac{\partial V_x}{\partial y}\right),$$

$$\text{curl curl } \mathbf{V} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ \left(\dfrac{\partial V_z}{\partial y} - \dfrac{\partial V_y}{\partial z}\right) & \left(\dfrac{\partial V_x}{\partial z} - \dfrac{\partial V_z}{\partial x}\right) & \left(\dfrac{\partial V_y}{\partial x} - \dfrac{\partial V_x}{\partial y}\right) \end{vmatrix}$$

$$= \mathbf{i}\left(\frac{\partial^2 V_y}{\partial x \partial y} - \frac{\partial^2 V_x}{\partial y^2} + \frac{\partial^2 V_z}{\partial z \partial x} - \frac{\partial^2 V_x}{\partial z^2} + \frac{\partial^2 V_x}{\partial x^2} - \frac{\partial^2 V_x}{\partial x^2}\right)$$

$$+ \text{analogous terms for } \mathbf{j} \text{ and } \mathbf{k}$$

$$= \text{grad div} - \nabla^2,$$

so that the operation

$$\text{curl curl} = \text{grad div} - \nabla^2.$$

(f) Other Formulas Involving ∇. If A and B are differentiable vector functions, and ϕ and ψ are differentiable scalar function of position x, y, z, then

$$\nabla(\phi + \psi) = \nabla\phi + \nabla\psi$$

$$\nabla \cdot (\mathbf{A} + \mathbf{B}) = \nabla \cdot \mathbf{A} + \nabla \cdot \mathbf{B}$$

$$\nabla \times (\mathbf{A} + \mathbf{B}) = \nabla \times \mathbf{A} + \nabla \times \mathbf{B}$$

$$\nabla \cdot (\phi \mathbf{A}) = (\nabla\phi) \cdot \mathbf{A} + \phi(\nabla \cdot \mathbf{A})$$

$$\nabla \times (\phi \mathbf{A}) = (\nabla\phi) \times \mathbf{A} + \phi(\nabla \times \mathbf{A})$$

$$\nabla(\phi\psi) = (\nabla\phi)\psi + \phi(\nabla\psi)$$

$$\nabla \cdot (\mathbf{A} \times \mathbf{B}) = \mathbf{B} \cdot (\nabla \times \mathbf{A}) - \mathbf{A} \cdot (\nabla \times \mathbf{B})$$

$$\nabla \times (\mathbf{A} \times \mathbf{B}) = (\mathbf{B} \cdot \nabla)\mathbf{A} - \mathbf{B}(\nabla \cdot \mathbf{A}) - (\mathbf{A} \cdot \nabla)\mathbf{B} + \mathbf{A}(\nabla \cdot \mathbf{B})$$

$$\nabla(\mathbf{A} \cdot \mathbf{B}) = (\mathbf{B} \cdot \nabla)\mathbf{A} + (\mathbf{A} \cdot \nabla)\mathbf{B} + \mathbf{B} \times (\nabla \times \mathbf{A}) + \mathbf{A} \times (\nabla \times \mathbf{B}).$$

10 Gauss's Divergence Theorem and Stokes' Theorem

The total flux through a closed surface A is given by

$$\phi = \oint_A \mathbf{V} \cdot \mathbf{n}\, dA,$$

where V is a vector field and n is a unit outward vector normal to dA. Also, the total flux diverging from the volume v enclosed by the surface A is given by

$$\phi = \oint_V \nabla \cdot \mathbf{V} dv.$$

Therefore,

$$\oint_A \mathbf{V} \cdot \mathbf{n}\, dA = \oint_V \nabla \cdot \mathbf{V} dv, \tag{47}$$

which is Gauss's divergence theorem.

Stokes' theorem states that the tangential line integral of a vector function V around a closed path is equal to the normal surface integral of curl V over the enclosed surface or cap enclosed by the path.

In symbols,

$$\oint \mathbf{V} \cdot d\mathbf{s} = \int_A \mathbf{n} \cdot \text{curl } \mathbf{V} dA, \tag{48}$$

where **n** is a unit vector normal to the element of surface.

Proof:

Let the surface (Figure 9) be divided into infinitesimal surface elements. Consider the shaded area of the vector area $\mathbf{n}dA$. At the middle of dA, the vector field has a curl. From the definition of the curl (namely, that it is the limiting value of the circulation per unit area), we may write

$$\oint \mathbf{V} \cdot d\mathbf{s} = \mathbf{n} \cdot \text{curl } \mathbf{V} dA$$

for this element of area. A similar process is applied to all the elements. The line integral along the common sides of the elements cancel, thus leaving only the contributions along the surface boundary. Therefore,

$$\oint \mathbf{V} \cdot d\mathbf{s} = \oint_A \mathbf{n} \cdot \text{curl } \mathbf{V} dA,$$

which is Stokes' theorem.

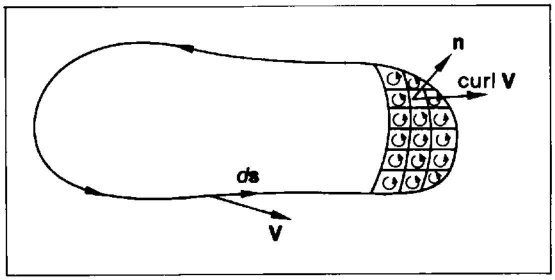


Figure 9 Stokes's theorem.

APPENDIX II UNIT CONVERSION FACTORS

1 Length

1 meter = 39.37 inches
1 kilometer = 0.6214 miles
1 angstrom = 10^{-10} meters
1 micron = 10^{-6} meters
1 foot = 0.3048 meters
1 inch = 0.0254 meters
1 mile = 1.609 kilometers

2 Force

1 newton = 10^5 dynes = 0.2248 pounds
1 kilogram-force = 9.807 newtons = 2.205 pounds
1 pound = 0.4536 kilogram-force

3 Mass

1 kilogram = 0.0685 slugs
1 slug = 14.59 kilograms

4 Pressure

atmosphere = 1.013×10^5 nt/m^2
= 1.013×10^6 dynes/cm^2
= 14.7 lb/in^2

5 Energy

1 joule = 10^7 ergs = 0·7376 foot-pounds
1 kilocalorie = 4185 joules = 3.968 B.T.U.
= 3086 foot-pounds
1 electron volt = 1.60×10^{-12} ergs
= 1.60×10^{-19} joules
1 kilowatt-hour = 2.655×10^6 foot-pounds
1 foot-pound = 1.356 joules
1 B.T.U. = 778 foot-pound
= 0.252 kilocalories

6 Power

1 kilowatt = 737.6 foot-pounds/sec
= 1.341 horsepower
1 horsepower = 550 foot-pounds/sec
= 0.7457 kilowatts

APPENDIX III FUNDAMENTAL CONSTANTS

Name of quantity	Symbol	Value
Velocity of light in vacuum	c	2.9979×10^8 m/sec
Charge of electron	q_e	-1.602×10^{-19} coul
Rest mass of electron	m_e	9.109×10^{-31} kg
Planck's constant	h	6.625×10^{-34} joules-sec
Boltzmann's constant	k	1.380×10^{-23} joules/°K
Avogadro's number	N_0	6.025×10^{23} molecules/mole
Universal gas constant (chemical scale)	R	8.314 joules/mole-°K
Mechanical equivalent of heat	J	4.185 joules/cal
Standard atmospheric pressure	1 atm	1.013×10^5 nt/m^2
Volume of ideal gas at 0°C and 1 atm		22.415 liters/mole
Absolute zero of temperature	0°K	−273.15°C
Acceleration due to gravity (sea level, at equator)	g	9.78049 m/sec^2
Universal gravitational constant	G	6.673×10^{-11} nt-m^2/kg^2
Mass of earth	m_E	5.975×10^{24} kg
Mean radius of earth		6.371×10^6 m = 3959 miles
Coulomb's law constant	K	8.98×10^9 nt-m^2/coul^2
Permittivity of free space	ϵ_0	$8.85 \times 10^{-12} \frac{\text{nt-m}^2}{\text{coul}^2}$
Permeability of free space	μ_0	$4\pi \times 10^{-7}$ nt/amp^2

APPENDIX IV MATHEMATICAL TABLES

1 Natural Trigonometric Functions

Angle					Angle					Angle				
Degree	Radian	Sine	Cosine	Tangent	Degree	Radian	Sine	Cosine	Tangent	Degree	Radian	Sine	Cosine	Tangent
0°	.000	0.000	1.000	0.000										
1°	.017	.018	1.000	.018	31°	.541	.515	.857	.601	61°	1.065	.875	.485	1.804
2°	.035	.035	0.999	.035	32°	.559	.530	.848	.625	62°	1.082	.883	.470	1.881
3°	.052	.052	.999	.052	33°	.576	.545	.839	.649	63°	1.100	.891	.454	1.963
4°	.070	.070	.998	.070	34°	.593	.559	.829	.675	64°	1.117	.899	.438	2.050
5°	.087	.087	.996	.088	35°	.611	.574	.819	.700	65°	1.134	.906	.423	2.145
6°	.105	.105	.995	.105	36°	.628	.588	.809	.727	66°	1.152	.914	.407	2.246
7°	.122	.122	.993	.123	37°	.646	.602	.799	.754	67°	1.169	.921	.391	2.356
8°	.140	.139	.990	.141	38°	.663	.616	.788	.781	68°	1.187	.927	.375	2.475
9°	.157	.156	.988	.158	39°	.681	.629	.777	.810	69°	1.204	.934	.358	2.605
10°	.175	.174	.985	.176	40°	.698	.643	.766	.839	70°	1.222	.940	.342	2.747
11°	.192	.191	.982	.194	41°	.716	.658	.755	.869	71°	1.239	.946	.326	2.904
12°	.209	.208	.978	.213	42°	.733	.669	.743	.900	72°	1.257	.951	.309	3.078
13°	.227	.225	.974	.231	43°	.751	.682	.731	.933	73°	1.274	.956	.292	3.271
14°	.244	.242	.970	.249	44°	.768	.695	.719	.966	74°	1.292	.961	.276	3.487
15°	.262	.259	.966	.268	45°	.785	.707	.707	1.000	75°	1.309	.966	.259	3.732
16°	.279	.276	.961	.287	46°	0.803	.719	.695	1.036	76°	1.326	.970	.242	4.011
17°	.297	.292	.956	.306	47°	.820	.731	.682	1.072	77°	1.344	.974	.225	4.331
18°	.314	.309	.951	.325	48°	.838	.743	.669	1.111	78°	1.361	.978	.208	4.705
19°	.332	.326	.946	.344	49°	.855	.755	.656	1.150	79°	1.379	.982	.191	5.145
20°	.349	.342	.940	.364	50°	.873	.766	.643	1.192	80°	1.396	.985	.174	5.671
21°	.367	.358	.934	.384	51°	.890	.777	.629	1.235	81°	1.414	.988	.156	6.314
22°	.384	.375	.927	.404	52°	.908	.788	.616	1.280	82°	1.431	.990	.139	7.115
23°	.401	.391	.921	.425	53°	.925	.799	.602	1.327	83°	1.449	.993	.122	8.144
24°	.419	.407	.914	.445	54°	.942	.809	.588	1.376	84°	1.466	.995	.105	9.514
25°	.436	.423	.906	.466	55°	.960	.819	.574	1.428	85°	1.484	.996	.087	11.43
26°	.454	.438	.899	.488	56°	.977	.829	.559	1.483	86°	1.501	.998	.070	14.30
27°	.471	.454	.891	.510	57°	.995	.839	.545	1.540	87°	1.518	.999	.052	19.08
28°	.489	.470	.883	.532	58°	1.012	.848	.530	1.600	88°	1.536	.999	.035	28.64
29°	.506	.485	.875	.554	59°	1.030	.857	.515	1.664	89°	1.553	1.000	.018	57.29
30°	.524	.500	.866	.577	60°	1.047	.866	.500	1.732	90°	1.571	1.000	.000	∞

2 Table of Logarithms to Base 10

													P.P.		
N	0	1	2	3	4	5	6	7	8	9	1	2	3	4	5
10	0000	0043	0086	0128	0170	0212	0253	0204	0334	0374	4	8	12	17	21
11	0414	0453	0492	0531	0569	0607	0645	0682	0719	0755	4	8	11	15	19
12	0792	0828	0864	0899	0934	0969	1004	1038	1072	1106	3	7	10	14	17
13	1139	1173	1206	1239	1271	1303	1335	1307	1399	1430	3	6	10	13	16
14	1461	1492	1523	1553	1584	1614	1644	1673	1703	1732	3	6	9	12	15
15	1761	1790	1818	1847	1875	1903	1931	1959	1987	2014	3	6	8	11	14
16	2041	2068	2095	2122	2148	2175	2201	2227	2253	2279	3	5	8	11	13
17	2304	2330	2355	2380	2405	2430	2455	2480	2504	2529	2	5	7	10	12
18	2553	2577	2601	2625	2648	2672	2695	2718	2742	2765	2	5	7	9	12
19	2788	2810	2833	2856	2878	2900	2923	2945	2967	2989	2	4	7	9	11
20	3010	3032	3054	3075	3096	3118	3139	3160	3181	3201	2	4	6	8	11
21	3222	3243	3263	3284	3304	3324	3345	3365	3385	3404	2	4	6	8	10
22	3424	3444	3464	3483	3502	3522	3541	3560	3579	3598	2	4	6	8	10
23	3617	3636	3655	3674	3692	3711	3729	3747	3766	3784	2	4	5	7	9
24	3802	3820	3838	3856	3874	3892	3909	3927	3945	3962	2	4	5	7	9
25	3979	3997	4014	4031	4048	4065	4082	4099	4116	4133	2	3	5	7	9
26	4150	4166	4183	4200	4216	4232	4249	4265	4281	4298	2	3	5	7	8
27	4314	4330	4346	4362	4378	4393	4409	4425	4440	4456	2	3	5	6	8
28	4472	4487	4502	4518	4533	4548	4564	4579	4594	4609	2	3	5	6	8
29	4624	4639	4654	4669	4683	4698	4713	4728	4742	4757	1	3	4	6	7
30	4771	4786	4800	4814	4829	4843	4857	4871	4886	4900	1	3	4	6	7
31	4914	4928	4942	4955	4969	4983	4997	5011	5024	5038	1	3	4	6	7
32	5051	5065	5079	5092	5105	5119	5132	5145	5159	5172	1	3	4	5	7
33	5185	5198	5211	5224	5237	5250	5263	5276	5289	5302	1	3	4	5	6
34	5315	5328	5340	5353	5366	5378	5391	5403	5416	5428	1	3	4	5	6
35	5441	5453	5465	5478	5490	5502	5514	5527	5539	5551	1	2	4	5	6
36	5563	5575	5587	5599	5611	5623	5635	5647	5658	5670	1	2	4	5	6
37	5682	5694	5705	5717	5729	5740	5752	5763	5775	5786	1	2	3	5	6
38	5798	5809	5821	5832	5843	5855	5866	5877	5888	5899	1	2	3	5	6
39	5911	5922	5933	5944	5955	5966	5977	5988	5999	6010	1	2	3	4	6
40	6021	6031	6042	6053	6064	6075	6085	6096	6107	6117	1	2	3	4	5
41	6128	6138	6149	6160	6170	6180	6191	6201	6212	6222	1	2	3	4	5
42	6232	6243	6253	6263	6274	6284	6294	6304	6314	6325	1	2	3	4	5
43	6335	6345	6355	6365	6375	6385	6395	6405	6415	6425	1	2	3	4	5
44	6435	6444	6454	6464	6474	6484	6493	6503	6513	6522	1	2	3	4	5
45	6532	6542	6551	6561	6571	6580	6590	6599	6609	6618	1	2	3	4	5
46	6628	6637	6646	6656	6665	6675	6684	6693	6702	6712	1	2	3	4	5
47	6721	6730	6739	6749	6758	6767	6776	6785	6794	6803	1	2	3	4	5
48	6812	6821	6830	6839	6848	6857	6866	6875	6884	6803	1	2	3	4	4
49	6902	6911	6920	6928	6937	6946	6955	6964	6972	6981	1	2	3	4	4
50	6990	6998	7007	7016	7024	7033	7042	7050	7059	7067	1	2	3	3	4
51	7076	7084	7093	7101	7110	7118	7126	7135	7143	7152	1	2	3	3	4
52	7160	7168	7177	7185	7193	7202	7210	7218	7226	7235	1	2	2	3	4
53	7243	7251	7259	7267	7275	7284	7292	7300	7308	7316	1	2	2	3	4
54	7324	7332	7340	7348	7356	7364	7372	7380	7388	7396	1	2	2	3	4

Table of Logarithms to Base 10 (cont'd)

													P.P.		
N	0	1	2	3	4	5	6	7	8	9	1	2	3	4	5
55	7404	7412	7419	7427	7435	7443	7451	7459	7466	7474	1	2	2	3	4
56	7482	7490	7497	7505	7513	7520	7528	7536	7543	7551	1	2	2	3	4
57	7559	7566	7574	7582	7589	7597	7604	7612	7619	7627	1	2	2	3	4
58	7634	7642	7649	7657	7664	7672	7679	7686	7694	7701	1	1	2	3	4
59	7709	7716	7723	7731	7738	7745	7752	7760	7767	7774	1	1	2	3	4
60	7782	7789	7796	7803	7810	7818	7825	7832	7839	7846	1	1	2	3	4
61	7853	7860	7868	7875	7882	7889	7896	7903	7910	7917	1	1	2	3	4
62	7924	7931	7938	7945	7952	7959	7966	7973	7980	7987	1	1	2	3	3
63	7993	8000	8007	8014	8021	8028	8035	8041	8048	8055	1	1	2	3	3
64	8062	8069	8075	8082	8089	8096	8102	8109	8116	8122	1	1	2	3	3
65	8129	8136	8142	8149	8156	8162	8169	8176	8182	8189	1	1	2	3	3
66	8195	8202	8209	8215	8222	8228	8235	8241	8248	8254	1	1	2	3	3
67	8261	8267	8274	8280	8287	8293	8299	8306	8312	8319	1	1	2	3	3
68	8325	8331	8338	8344	8351	8357	8363	8370	8376	8382	1	1	2	3	3
69	8388	8395	8401	8407	8414	8420	8426	8432	8439	8445	1	1	2	3	3
70	8451	8457	8463	8470	8476	8482	8488	8494	8500	8506	1	1	2	2	3
71	8513	8519	8525	8531	8537	8543	8549	8555	8561	8567	1	1	2	2	3
72	8573	8579	8585	8591	8597	8603	8609	8615	8621	8627	1	1	2	2	3
73	8633	8639	8645	8651	8657	8663	8669	8675	8681	8686	1	1	2	2	3
74	8692	8698	8704	8710	8716	8722	8727	8733	8739	8745	1	1	2	2	3
75	8751	8756	8762	8768	8774	8779	8785	8791	8797	8802	1	1	2	2	3
76	8808	8814	8820	8825	8831	8837	8842	8848	8854	8859	1	1	2	2	3
77	8865	8871	8876	8882	8887	8893	8899	8904	8910	8915	1	1	2	2	3
78	8921	8927	8932	8938	8943	8949	8954	8960	8965	8971	1	1	2	2	3
79	8976	8982	8987	8993	8998	9004	9009	9015	9020	9025	1	1	2	2	3
80	9031	9036	9042	9047	9053	9058	9063	9069	9074	9079	1	1	2	2	3
81	9085	9090	9096	9101	9106	9112	9117	9122	9128	9133	1	1	2	2	3
82	9138	9143	9149	9154	9159	9165	9170	9175	9180	9186	1	1	2	2	3
83	9191	9196	9201	9206	9212	9217	9222	9227	9232	9238	1	1	2	2	3
84	9243	9248	9253	9258	9263	9269	9274	9279	9284	9289	1	1	2	2	3
85	9294	9299	9304	9309	9315	9320	9325	9330	9335	9340	1	1	2	2	3
86	9345	9350	9355	9360	9365	9370	9375	9380	9385	9390	1	1	2	2	3
87	9395	9400	9405	9410	9415	9420	9425	9430	9435	9440	0	1	1	2	2
88	9445	9450	9455	9460	9465	9469	9474	9479	9484	9489	0	1	1	2	2
89	9494	9499	9504	9509	9513	9518	9523	9528	9533	9538	0	1	1	2	2
90	9542	9547	9552	9557	9562	9566	9571	9576	9581	9586	0	1	1	2	2
91	9590	9595	9600	9605	9609	9614	9619	9624	9628	9633	0	1	1	2	2
92	9638	9643	9647	9652	9657	9661	9666	9671	9675	9680	0	1	1	2	2
93	9685	9689	9694	9699	9703	9708	9713	9717	9722	9727	0	1	1	2	2
94	9731	9736	9741	9745	9750	9754	9759	9763	9768	9773	0	1	1	2	2
95	9777	9782	9786	9791	9795	9800	9805	9809	9814	9818	0	1	1	2	2
96	9823	9827	9832	9836	9841	9845	9850	9854	9859	9863	0	1	1	2	2
97	9868	9872	9877	9881	9886	9890	9894	9899	9903	9908	0	1	1	2	2
98	9912	9917	9921	9926	9930	9934	9939	9943	9948	9952	0	1	1	2	2
99	9956	9961	9965	9969	9974	9978	9983	9987	9991	9996	0	1	1	2	2

3 Table of Exponentials

e^x

x	0	1	2	3	4	5	6	7	8	9
0.0	1.000	1.010	1.020	1.031	1.041	1.051	1.062	1.073	1.083	1.094
0.1	1.105	1.116	1.127	1.139	1.150	1.162	1.174	1.185	1.197	1.209
0.2	1.221	1.234	1.246	1.259	1.271	1.284	1.297	1.310	1.323	1.336
0.3	1.350	1.363	1.377	1.391	1.405	1.419	1.433	1.448	1.462	1.477
0.4	1.492	1.507	1.522	1.537	1.553	1.568	1.584	1.600	1.616	1.632
0.5	1.649	1.665	1.682	1.699	1.716	1.733	1.751	1.768	1.786	1.804
0.6	1.822	1.840	1.859	1.878	1.896	1.916	1.935	1.954	1.974	1.994
0.7	2.014	2.034	2.054	2.075	2.096	2.117	2.138	2.160	2.181	2.203
0.8	2.226	2.248	2.270	2.293	2.316	2.340	2.363	2.387	2.411	2.435
0.9	2.460	2.484	2.509	2.535	2.560	2.586	2.612	2.638	2.664	2.691
1.0	2.718	2.746	2.773	2.801	2.829	2.858	2.886	2.915	2.945	2.974
1.1	3.004	3.034	3.065	3.096	3.127	3.158	3.190	3.222	3.254	3.287
1.2	3.320	3.353	3.387	3.421	3.456	3.490	3.525	3.561	3.597	3.633
1.3	3.669	3.706	3.743	3.781	3.819	3.857	3.896	3.935	3.975	4.015
1.4	4.055	4.096	4.137	4.179	4.221	4.263	4.306	4.349	4.393	4.437
1.5	4.482	4.527	4.572	4.618	4.665	4.712	4.759	4.807	4.855	4.904
1.6	4.953	5.003	5.053	5.104	5.155	5.207	5.259	5.312	5.366	5.419
1.7	5.474	5.529	5.585	5.641	5.697	5.755	5.812	5.871	5.930	5.989
1.8	6.050	6.110	6.172	6.234	6.297	6.360	6.424	6.488	6.554	6.619
1.9	6.686	6.753	6.821	6.890	6.959	7.029	7.099	7.171	7.243	7.316
2.0	7.389	7.463	7.538	7.614	7.691	7.768	7.846	7.925	8.004	8.085
2.1	8.166	8.248	8.331	8.415	8.499	8.585	8.671	8.758	8.846	8.935
2.2	9.025	9.116	9.207	9.300	9.393	9.488	9.583	9.679	9.777	9.875
2.3	9.974	10.07	10.18	10.28	10.38	10.49	10.59	10.70	10.80	10.91
2.4	11.02	11.13	11.25	11.36	11.47	11.59	11.70	11.82	11.94	12.06
2.5	12.18	12.30	12.43	12.55	12.68	12.81	12.94	13.07	13.20	13.33
2.6	13.46	13.60	13.74	13.87	14.01	14.15	14.30	14.44	14.59	14.73
2.7	14.88	15.03	15.18	15.33	15.49	15.64	15.80	15.96	16.12	16.28
2.8	16.44	16.61	16.78	16.95	17.12	17.29	17.46	17.64	17.81	17.99
2.9	18.17	18.36	18.54	18.73	18.92	19.11	19.30	19.49	19.69	19.89
3.0	20.09	20.29	20.49	20.70	20.91	21.12	21.33	21.54	21.76	21.98
3.1	22.20	22.42	22.65	22.87	23.10	23.34	23.57	23.81	24.05	24.29
3.2	24.53	24.78	25.03	25.28	25.53	25.79	26.05	26.31	26.58	26.84
3.3	27.11	27.39	27.66	27.94	28.22	28.50	28.79	29.08	29.37	29.67
3.4	29.96	30.27	30.57	30.88	31.19	31.50	31.82	32.14	32.46	32.79

x	.0	.1	.2	.3	.4	.5	.6	.7	.8	.9
3	20.09	22.20	24.53	27.11	29.96	33.12	36.60	40.45	44.70	49.40
4	54.60	60.34	66.69	73.70	81.45	90.02	99.48	109.9	121.5	134.3
5	148.4	164.0	181.3	200.3	221.4	244.7	270.4	298.9	330.3	365.0
6	403.4	445.9	492.7	544.6	601.8	665.1	735.1	812.4	897.8	992.3
7	1097	1212	1339	1480	1636	1808	1998	2208	2441	2697
8	2981	3295	3641	4024	4447	4915	5432	6003	6634	7332
9	8103	8955	9897	10938	12088	13360	14765	16318	18034	19930

$$\log_{10} e^x = x \log_{10} e = 0.43429\, x$$

Table of Exponentials (contd)

e^{-x}

x		0	1	2	3	4	5	6	7	8	9
0.0		1.000	.9900	.9802	.9704	.9608	.9512	.9418	.9324	.9231	.9139
0.1		.9048	.8958	.8869	.8781	.8694	.8607	.8521	.8437	.8353	.8270
0.2		.8187	.8106	.8025	.7945	.7866	.7788	.7711	.7634	.7558	.7483
0.3		.7408	.7334	.7261	.7189	.7118	.7047	.6977	.6907	.6839	.6771
0.4		.6703	.6637	.6570	.6505	.6440	.6376	.6313	.6250	.6188	.6126
0.5		.6065	.6005	.5945	.5886	.5827	.5769	.5712	.5655	.5599	.5543
0.6		.5488	.5434	.5379	.5326	.5273	.5220	.5169	.5117	.5066	.5016
0.7		.4966	.4916	.4868	.4819	.4771	.4724	.4677	.4630	.4584	.4538
0.8		.4493	.4449	.4404	.4360	.4317	.4274	.4232	.4190	.4148	.4107
0.9		.4066	.4025	.3985	.3946	.3906	.3867	.3829	.3791	.3753	.3716
1.0		.3679	.3642	.3606	.3570	.3535	.3499	.3465	.3430	.3396	.3362
1.1		.3329	.3296	.3263	.3230	.3198	.3166	.3135	.3104	.3073	.3042
1.2		.3012	.2982	.2952	.2923	.2894	.2865	.2837	.2808	.2780	.2753
1.3		.2725	.2698	.2671	.2645	.2618	.2592	.2567	.2541	.2516	.2491
1.4		.2466	.2441	.2417	.2393	.2369	.2346	.2322	.2299	.2276	.2254
1.5		.2231	.2209	.2187	.2165	.2144	.2122	.2101	.2080	.2060	.2039
1.6		.2019	.1999	.1979	.1959	.1940	.1920	.1901	.1882	.1864	.1845
1.7		.1827	.1809	.1791	.1773	.1755	.1738	.1720	.1703	.1686	.1670
1.8		.1653	.1637	.1620	.1604	.1588	.1572	.1557	.1541	.1526	.1511
1.9		.1496	.1481	.1466	.1451	.1437	.1423	.1409	.1395	.1381	.1367
2.0		.1353	.1340	.1327	.1313	.1300	.1287	.1275	.1262	.1249	.1237
2.1		.1225	.1212	.1200	.1188	.1177	.1165	.1153	.1142	.1130	.1119
2.2		.1108	.1097	.1086	.1075	.1065	1054	.1043	.1033	.1023	.1013
2.3		.1003	*9926	*9827	*9730	*9633	*9537	*9442	*9348	*9255	*9163
2.4	0.0	9072	8982	8892	8804	8716	8629	8544	8458	8374	8291
2.5	0.0	8208	8127	8046	7966	7887	7808	7730	7654	7577	7502
2.6	0.0	7427	7353	7280	7208	7136	7065	6995	6925	6856	6788
2.7	0.0	6721	6654	6587	6522	6457	6393	6329	6266	6204	6142
2.8	0.0	6081	6020	5961	5901	5843	5784	5727	5670	5613	5558
2.9	0.0	5502	5448	5393	5340	5287	5234	5182	5130	5079	5029
3.0	0.0	4979	4929	4880	4832	4783	4736	4689	4642	4596	4550
3.1	0.0	4505	4460	4416	4372	4328	4285	4243	4200	4159	4117
3.2	0.0	4076	4036	3996	3956	3916	3877	3839	3801	3763	3725
3.3	0.0	3688	3652	3615	3579	3544	3508	3474	3439	3405	3371
3.4	0.0	3337	3304	3271	3239	3206	3175	3143	3112	3081	3050

x		.0	.1	.2	.3	.4	.5	.6	.7	.8	.9
3	0.0	4979	4505	4076	3688	3337	3020	2732	2472	2237	2024
4	0.0	1832	1657	1500	1357	1228	1111	1005	*9095	*8230	*7447
5	0.00	6738	6097	5517	4992	4517	4087	3698	3346	3028	2739
6	0.00	2479	2243	2029	1836	1662	1503	1360	1231	1114	1008
7	0.000	9119	8251	7466	6755	6112	5531	5004	4528	4097	3707
8	0.000	3355	3035	2747	2485	2240	2035	1841	1666	1507	1364
9	0.000	1234	1117	1010	*9142	*8272	*7485	*6773	*6128	*5545	*5017
10	0.0000	4540	4108	3717	3363	3043	2754	2492	2254	2040	1846

$$\log_{10} e^{-x} = -x \log_{10} e = -0.43429\,x$$

Answer Section

ANSWERS TO ODD-NUMBERED PROBLEMS

Chapter 1

1. 25°C 3. 25°C 5. 80°F 7. 122°F, 50°C

Chapter 2

1. 23.3°C 3. (a) lost in the form of vapor or steam (b) 100°C 5. 0.57 cal/gm °C 7. 7200 cal
9. 543 cal/gm 11. 0.10 cal/gm °C 13. (a) 5.9°C (b) 0°C 15. 10 kcal/gm

Chapter 3

1. 340°C 5. 8.64 cm^3 7. 83.5°C 9. 0.291 cm^3 11. $d = \frac{L\alpha_A}{\alpha_B - \alpha_A}$
13. l(steel) = 21.7 cm, l(brass) = 13.7 cm 15. 369 cm^3 17. 9.65×10^{-4} °C^{-1} 19. 27.8 cm^3
21. (a) 2.67×10^{-5} °C^{-1} (b) 18×10^{-5} °C^{-1}

Chapter 4

1. 190°C 3. 31.6 kcal/sec-m^2 5. 0.12 7. 3700 9. 5920°K 11. (a) 640 kcal
(b) 19,500 seconds 13. 100.5°C 15. (a) −2.5°C (b) −6.7°C

Chapter 5

1. (a) yes (b) yes (c) positive 3. 13.5 mm 5. 5°C, 51% 7. 10^{-2} atm 9. 0.01°C

Chapter 6

5. $D = 3.23$ cm 11. $\delta = 52.4°$

Chapter 7

5. $p = 7.5$ cm 7. $h = 15.1$ m 9. 1.6R from the front of the bowl
11. Image is erect, real, twice original size and located $\frac{19}{3}R$ to right of spherical surface

Chapter 8

1. Correct: final image is 3.4R behind front surface; Incorrect: final image is 1.5R behind front surface 3. $f = -60$ cm 7. ±12 cm, ±60 cm 9. 60 cm behind the diverging lens
11. (a) $f = -75$ cm (b) 100 cm 13. 2.5, 16.7 cm 17. $f/4$

Chapter 9

1. $t = 2 \times 10^{-6}$ sec 3. $c = 3.01 \times 10^8$ m/sec 5. $c = 2.998 \times 10^8$ m/sec

Chapter 10

3. 4.3 m **5.** $y = 2\sin[2\pi(0.2x - 4t)]$; $f = 4$ cycles/sec **9.** (a) π radians (b) 0 (c) 0 (d) $\pi/2$ radians **11.** $f_1 = 523$ or 501 cycles/sec **13.** $n\lambda = 2\pi r$

13. $n\lambda = 2\pi r$

Chapter 11

1. $\lambda = 6800$ Å **3.** $t = \frac{3}{4} \times 10^{-7}$ m **7.** (a) $y = 8.1$ mm (b) $n = 3$ for 4500 Å; $n = 8$ for 6000 Å
9. $n = 182$ **11.** $(N_a - N_{vac}) \approx 94$

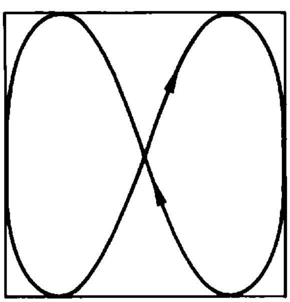
Chapter 12

5. (a) $r_8 = 4.01$ mm (b) $r_8 = 1.26$ cm **9.** (a) $\theta_1 = 11.5°$; $\theta_2 = 20.5°$ (b) $\Delta y = 8.5$ cm
17. (c) $\lambda_x = 4920$ Å (d) $\theta_x = 22°$

Chapter 13

1. (a) $x_{total} = \sqrt{3}\,A \sin\left[\frac{2\pi}{\lambda}(z - vt) + 60°\right]$ (b) $x_{total} = \sqrt{7}\,A \sin[\theta_z + 70.9°]$

3. (a)



9. (a) $|\mathbf{A}+\mathbf{B}| = 23.2$ (b) $\alpha = 72.9°$
$|\mathbf{A}-\mathbf{B}| = 30.5$ $\beta = 126°$
$|2\mathbf{A}+\mathbf{B}| = 30.1$ $\gamma = 61.9°$

11. (a) $\theta_{max} = 45°$ (b) $I/I_0 = \frac{1}{8}$ **15.** (a) 53° (b) vertical **17.** $t = 8.56 \times 10^{-7}$ m
19. (b) All light transmitted by the first polaroid film will be transmitted

Chapter 14

1. 11.5 ft/sec^2 **3.** 142 ft **5.** -1.33 ft/sec^2 **7.** (a) 56 ft (b) -36 ft/sec, 36 ft from ground
9. (a) 3000 ft west (b) 157 ft/sec **11.** (a) 2 m/sec (b) 15 m, 5 m/sec south **13.** (a) 36 ft (b) 0.5 sec (c) 48 ft/sec (d) -32 ft/sec^2

Chapter 15

5. $L/L_0 \approx 1 - \frac{4}{9} \times 10^{-8}$ **7.** $\Delta t_2 = 5.77 \times 10^{-9}$ sec **9.** $v = 0.6 \times 10^8$ m/sec toward S_1

Chapter 16

1. $F = 1$ nt **5.** $s \approx 2.1 \times 10^{-10}$ m **7.** $x = \frac{9}{10} R_{EM}$ **11.** $g = g_0/(1 + h/R_E)^2$
13. $t < 12 \times 10^3$ sec or $\approx \frac{1}{7}$ day **15.** (a) $F_{21} = \frac{4}{27} \times 10^{-10}$ nt toward m_1 (b) $F_{23} = \frac{1}{4} \times 10^{-10}$ nt toward m_3 (c) $a_y = \frac{2}{27} \times 10^{-10}$ m/sec^2 up; $a_x = \frac{1}{8} \times 10^{-10}$ m/sec^2 to right (d) $a = 1.45 \times 10^{-11}$ m/sec^2; $\theta \approx 29°$

Chapter 17

1. (b) $\omega/(2\sin\varphi)$ (c) 30° **3.** (a) $\frac{8}{3}$ m/sec^2 (b) 2.1 m/sec^2; 6.3 nt **5.** 1610 lb
7. (a) 20 cm/sec^2 (b) 0.384 nt $= 3.84 \times 10^4$ dynes **11.** (a) 19.6 nt (b) 2.1 kg
13. (a) 170 lb (b) 150 lb (c) 0 ("free fall")

Chapter 18

1. (a) 4 ft/sec^2 (b) 1.4 lb **3.** 4.25 m/sec^2 **5.** (a) 3.92 m/sec^2 (b) 16.2 nt (c) 2.75 kg
7. (a) 5.20 lb (b) 20.5 lb (c) 25.5 lb **9.** (a) 42 ft, 1.5 seconds (b) 259 ft **11.** (a) 86,900 ft
(b) $x = 23{,}000$ ft, $y = 53{,}400$ ft **13.** (a) 30 seconds (b) 14,400 ft (c) 96,000 ft
15. (a) 15.8 seconds (b) 500 ft **17.** 0.318 rev/sec

Chapter 19

1. 49,000 cm/sec **3.** 14 cm/sec **5.** 3.1 ft/sec **7.** 1000 m **9.** (a) 1 ft/sec
(b) 40,000 lb sec **11.** (a) 20 ft/sec (b) 2 ft/sec^2, 0.5 lb (c) 0.062 **13.** 40 m/sec, 0.2 seconds
15. (a) 188 slug ft/sec (b) 376 lb

Chapter 20

1. (a) 1000 ft lb (b) 211.5 lb (c) 45.3 ft lb (d) 743.2 ft lb **3.** (a) 100 ft lb (b) 0.50
5. (a) 10^6 joules; 1.96×10^3 joules (b) 3.6×10^5 joules; 6.42×10^5 joules (c) 1×10^6 joules, 0
7. (a) 1.06 hp (b) 300 watts **9.** (a) 2700 ft lb (b) 1800 ft lb (c) 624 ft lb (d) 276 ft/lb
11. 0.30 hp, 224 watts **13.** (a) 150 lb (b) 1.14 **15.** (a) 41.7 lb (b) 834 ft lb (c) 0.152

Chapter 21

1. (a) 50 rad/sec (b) 10 rad/sec^2 (c) 100 rad/sec (d) 100 cm/sec^2 (e) 1000 cm/sec
(f) 10^5 cm/sec^2 **3.** (a) 2 lb ft (b) 4.5 rad/sec^2 (c) 27 rad/sec (d) 162 ft-lb (e) 54 ft
5. 16 rad/sec **7.** (a) 15.8 seconds (b) 500 ft **9.** (a) 1.33 rad/sec (b) 0.89 rad/sec^2
11. (a) 1220 ft (b) 12.5 lb **13.** 14.3 rad/sec **15.** 32 ft/sec

Chapter 22

1. (a) 0.314 m/sec (b) -0.49 m/sec^2 (c) 0.99 nt/m **3.** (a) 1.15 m/sec (b) 0.82 m
5. 0.10 **7.** (a) 1.75 sec^{-1} (b) 0.061 joules (c) 1.32 m/sec (d) 14.5 m/sec^2
9. (a) 0.40 nt (b) $y = A \sin 2\pi\left(\dfrac{t}{0.05\text{ sec}} - \dfrac{x}{2\text{ cm}}\right)$ **11.** 80.4 cm/sec; 64,700 cm/sec^2

Chapter 23

1. 1.5 m **3.** 0.026 cm **5.** (a) 0.112 in. (b) 2.08×10^3 ft-lb **7.** 0.24 joules **9.** 0.50 joules
11. 10^5 nt **13.** 0.46 in^2 **15.** 17.3 lb/in^2

Chapter 24

1. 7.45×10^3 nt/m^2 **3.** 1000 nt **5.** 4 m/sec **7.** 2.5 gm/cm^2; 1.2 gm/cm^3
9. (a) 10.0166 gm (b) 0.16% **11.** 28 lb/in^2 **13.** (a) 30 ft/sec and 40 ft/sec (b) 0 ft and 10.8 ft
15. 3.96 lb/in^2

Chapter 25

1. 60.6 joules; 191 joules **5.** (a) 10^5 joules (b) 6.93 joules (c) 0 **9.** (a) 2.50×10^5 joules
(b) 6.93×10^4 joules (c) 1.50×10^5 joules **11.** 298°C **13.** (a) -9.05 kcal (b) 7.56 kcal
(c) 0 (adiabatic) **15.** (a) 2833 cm^3 (b) 287 (c) 140 **17.** 200 m **19.** -78.5 kcal/kg

Chapter 26

1. (a) 320°K (b) 20% **3.** (a) 33.4 joules (b) 57% (c) 2.33 **5.** 540 cal/gm
7. 600 joules **9.** (a) 11.4 kg (b) 3.68×10^5 joules **11.** 1.25 **13.** 0.62 kcal/°K

Chapter 27

1. (a) true (b) false **5.** (a) 1170 joules (b) 2600 m/sec **7.** 16.4 kcal; 11.7 kcal; 4.7 kcal;
11.7 kcal **11.** 94,400 cm/sec

Chapter 28

1. 1.95×10^{11} nt/m^2 **3.** (a) 5030 watts (b) 1060 cycles/sec **5.** 20 db **7.** 310 cps **9.** 29 seconds **11.** 0.20 watts/m^2 **13.** 19.4 m/sec

Chapter 29

1. (a) 1.73×10^{-10} nt toward A (b) 13.8×10^{-4} nt/coul (c) 0 **3.** (a) 9×10^4 volts (b) 7.07×10^5 volts/m at an angle of 38.2° with the $+x$ direction (c) 0.45 joules **5.** (b) 20 volts (c) 3.2×10^{-6} nt downward (d) 3.5×10^{14} m/sec^2 downward **7.** (b) 100 volts (c) 8×10^{-16} nt downward (d) 8.8×10^{14} m/sec^2 downward **9.** (a) 1.45×10^6 m/sec (b) 2.76×10^{-9} sec (c) 9.60×10^{-19} joules **11.** 8×10^3 volts; 0 **13.** -4 volts/m; 4 volts/m **15.** (a) $E = -c$ (b) $E = -gc(d_1 - d_2)$ (c) volts/m

Chapter 30

1. $E = \frac{a^2\rho}{2r\epsilon_0}$; $V = \frac{a^2\rho}{2\epsilon_0} \ln \frac{r_0}{r}$ **5.** $E = \frac{\rho r}{3\epsilon_0}$; $V = \frac{\rho}{6\epsilon_0}[r_0^2 - r^2]$ **7.** (a) 0 (b) $K\frac{Q_1}{r^2}$ (c) $K\frac{Q_1 + Q_2}{r^2}$ **9.** (a) 8×10^{-4} coul (b) 0.16 joules **11.** (a) 133 volts (b) 0.053 joules (c) polarized oil dielectric **15.** (a) 2 micromicrofrads (b) 2×10^{-9} coul (c) 667 volts and 333 volts (d) 10^{-6} joules **17.** (a) 13.3 microcoul/m^2 (b) 4.5 microcoul/m^2 (c) 4.5 microcoul/m^2 (d) 13.3 microcoul/m^2 (e) 6.6 joules/m^3 **19.** $E = \frac{Q}{4\pi\epsilon r^2}$; $P = \frac{\chi_e Q}{4\pi K r^2}$; $D = \frac{Q}{4\pi r^2}$

Chapter 31

1. 6.05 ohms **3.** 0.0025/°C **5.** 6 ohms **7.** (a) 0.30 ohms (b) 12 volts **9.** 2.1 volts **11.** (a) 25 watts (b) 4500 joules **13.** 600 ohms **15.** (a) 27 volts (b) 27 watts (c) 4 ohms (d) 12 volts (e) 2 amp **17.** 0.5 seconds

Chapter 32

1. 300 nt **3.** 12 nt-m **5.** (a) 0 (b) 0.0125 amp-m^2 **7.** (a) 0.80 amp-m^2/sec along negative z-axis (b) 0.173 nt-m parallel to y-axis **9.** (a) 25 amp (b) 1.6 amp **11.** 2.3 amp **13.** 2450 ohms **15.** 1.69 ohms **17.** (a) 0.303 ohms (b) 14,970 ohms **19.** 5×10^{-5} nt/amp-m vertically downward

Chapter 33

1. 0.04 webers **3.** 3.14×10^{-3} webers/m^2 **5.** 0.050 volts **7.** 80 joules **9.** 16.7 millisec **15.** 3 volts **17.** (a) 2 amp (b) 240 joules (c) $\frac{5}{6}$ amp; $\frac{7}{12}$ amp/sec **19.** (a) 3 amp/sec (b) 2 amp/sec (c) 1 amp (d) 3 amp

Chapter 34

1. 0.08 amp-m^2 **3.** (a) 175 amp-turns/m (b) 0.264 webers/m^2 **7.** (a) 1 (b) 6×10^{-4} **9.** *a* **11.** 12 amp **13.** 300°C

Chapter 35

1. 100 ohms **3.** (a) 11.31 ohms (b) lag (c) 45 deg **5.** (a) 221 microfarads (b) 0 deg **7.** (a) 200 rad/sec (b) 0.707 amp **9.** (a) 100 ohms; 40 ohms (b) 100 volts; 141 volts **11.** (a) 56.2 ohms (b) 2.14 amp (c) 0.09 (d) 23 watts **13.** (a) 300 ohms (b) 500 ohms (c) 0.20 amp and 0.283 amp (d) 80 volts (e) 50 volts (f) 16 watts **15.** (a) 2 amp (b) $V_c = 60$ volts; $V_R = 88$ volts; $V_{coil} = 194$ volts **17.** (a) 110 volts (b) 1 amp (c) 2200 watts

Chapter 36

1. (a) $\partial \mathbf{D}/\partial t = \text{curl}\,\mathbf{H}$ (b) Displacement current $= A$ curl $\mathbf{H}$ **3.** (a) 1.41 (b) 2.13 **7.** (a) 1×10^8 m/sec (b) 3 (c) 8×10^{-7} joules/m^2 **11.** 6.65×10^{-12} watts/m^2

Index†

†11*t* refers to a table.